Journal of International Money and Finance 127 (2022) 102654

Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf

Debt is not free ☆

Marialuz Moreno Badia *, Paulo Medas, Pranav Gupta, Yuan Xiang [1]

International Monetary Fund, Washington, DC 20431, USA

A R T I C L E I N F O

Article history:
Available online 4 May 2022

JEL Classification:
E43
E62
F34
F37
H63

Keywords:
Crisis
Debt
Sovereign default
Fiscal
Machine learning

A B S T R A C T

Many countries face record public debt levels but have until recently benefited from favorable interest-growth differentials. The empirical evidence on the link between high debt and crises is also inconclusive. This paper re-examines the importance of public debt as a leading indicator of fiscal crises using machine learning techniques to account for complex interactions ignored in the literature. We find that public debt is among the most important predictors: beyond certain debt levels, the likelihood of crises increases regardless of the interest-growth differential. Excessive current account deficits and private credit and their interaction with public debt are also important indicators of distress. These results underscore the risks from high debt levels and the role of broader economic imbalances in the emergence of fiscal crises.

1. Introduction

As the COVID-19 pandemic pushed the world into a deep recession, governments took sizable and unprecedented fiscal action in a low interest rate environment (IMF, 2020). But this came with a cost, as public debt has now risen to record highs across the globe. A critical policy question is whether high and rapidly increasing debt levels could signal future fiscal crises, which tend to be associated with lower real economic growth and permanent output losses (Medas et al., 2018; Asonuma et al., 2019).

Despite Reinhart and Rogoff's (2009, 2011a) seminal work on the perils of excessive debt, the empirical literature on the relationship between public debt and fiscal (or sovereign debt) crises is still inconclusive. Some argue that public debt may have no fiscal cost, or it is not a relevant indicator of fiscal sustainability for advanced economies facing historically low interest rates (Blanchard, 2019, Furman and Summers, 2020). If interest rates are lower than the economic growth rate, there may be no reason to maintain a primary surplus as it would be feasible to issue debt without later increasing taxes.

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

* Corresponding author.

E-mail addresses: MMorenobadia@imf.org (M. Moreno Badia), PMedas@imf.org (P. Medas), PGupta@imf.org (P. Gupta), yuanxiang@discover.com (Y. Xiang).

[1] Since the completion of this paper, the author has moved to Discover Financial Services, Riverwoods, IL 60015, USA.

https://doi.org/10.1016/j.jimonfin.2022.102654
0261-5606/

The lack of strong evidence in the literature on the importance of public debt stems from various methodological challenges. First, not only are crises rare events but also limitations with debt data make robust modeling difficult.[2] More importantly, complex nonlinearities and interactions may be at play. For example, a government may respond to a period of a low interest-growth differential by increasing the deficit—which may in turn lead to higher debt and ultimately a crisis. Classic econometric models typical of the early warning literature are not well suited to distill these effects. To help shed some light on the policy debate and to resolve various methodological challenges, we use machine learning models to identify robust predictors of fiscal crises and ask whether public debt is a reliable leading indicator by itself or when interacting with other variables.

We bring evidence to bear on the issue by studying fiscal crises in a broader sample (188 countries) than in the previous literature going back to the 1980s. One of the main novelties of our empirical strategy is to take an agnostic approach to the selection of predictors following a two-step procedure. As a starting point, we consider a wide range of predictors that the literature commonly associates with the onset of crises. Also, we do not take a view as to what the relevant moments of the variables are but instead consider many permutations yielding a total of 748 indicators. By leveraging machine learning models, we can fit complex and flexible functional forms to our data without overfitting. In a second step, we reduce the large set of predictors to the ones that contain more information than noise by using what is referred to in the machine learning literature as "feature selection algorithms", a workhorse in genomics research. We then use a battery of statistical measures to go beyond the black box, allowing us to uncover the relative importance of variables and their interactions.

Our results show public debt in its various forms is among the most important group of predictors. However, some forms of debt are more important than others—in particular, public external debt—and there is strong evidence of non-linearities. Remarkably, the interest-growth differential has low predictive value. Beyond certain debt levels the likelihood of fiscal crises increases significantly irrespective of whether the interest-growth differential is highly positive or negative. An event study gives some insights as to the possible reasons: it is only at the onset of the crisis that the interest-growth differential tends to spike, making it immaterial for signaling purposes.

The empirical analysis also reveals that it is not only public debt that matters. Among advanced economies, other macroeconomic imbalances such as excessive current account deficits and private debt are among the top predictors of fiscal crises. Interactions are also important. The probability of a crisis rises steeply not only at high public debt levels but also at relatively moderate ones when accompanied by other economic factors, such as high inflation, large current account deficits or credit gaps. These findings highlight the importance of monitoring the broader health of the economy to prevent the occurrence of a crisis.

Our work is related to the extensive early warning literature on sovereign debt crises (Detragiache and Spilimbergo, 2001; Manasse et al., 2003; Manasse and Roubini, 2009; Chakrabarti and Zeaiter, 2014). Relative to that literature, our contribution is twofold. First, we analyze the predictive importance of public debt relying on a novel dataset with a larger and more comprehensive coverage of debt and macroeconomic variables taking an agnostic approach about the relevant indicators while using feature selection machine learning algorithms to reduce dimensionality. We also examine the predictive power of the interest-growth differential while accounting for its interactions with debt. Our second contribution is to leverage machine learning to analyze complex non-linearities and interactions that had been ignored in the literature for the most part. Our paper sheds light on the nature of the complex dynamics at play, showcasing the potential of machine learning in macroeconomics, a field where the use of these techniques is still in its infancy.

The rest of the paper is organized as follows. Section 2 provides a survey of the literature on the determinants of fiscal crises over the last five decades to frame our empirical strategy. Section 3 discusses the definition of fiscal crises and the salient features of the dataset. Section 4 describes our methodology. The following section presents our results, exploring at length the predictor importance and identifying the extent to which non-linearities and interaction effects play a role. The concluding section offers some policy implications.

2. Does debt matter? Lessons from the literature

The literature on fiscal crises and their determinants has evolved significantly over time reflecting the changing nature of sovereign defaults and other forms of fiscal distress. Initially, the focus was on developing countries. The 1950s–1960s was a period when greater indebtedness was seen as a means to promote economic growth among less developed nations. By the 1970s, borrowing started to be associated also with periods of external imbalances. Nonetheless, debt was generally seen in a positive light (Solberg, 1988). By the early 1980s, the number of fiscal crises surged and so did the research on the drivers of debt distress. We present an overview of the literature over the last 50 years on the determinants of crises (Fig. 1 and Table 1). We did a comprehensive survey of 42 papers chosen out of a pool of 63 references based on their empirical relevance and whether they clearly identify key predictors of crises (Supplement 1.1).

1970s–1990s. In this period, the literature focused on assessing the capacity of the sovereign to manage its debt service and avoid defaults on external debt (Feder et al., 1981; Taffler and Abassi, 1984; Hajivassiliou, 1987). The definition of crisis

[2] Historical datasets on debt covering decades (if not centuries) of data have only become available recently, and they include only a few countries or use a narrow and changing definition of debt limiting the scope of research—e.g., Reinhart and Rogoff (2009) and (Abbas et al., 2011) on public debt; and Jordà et al. (2016) on private debt.

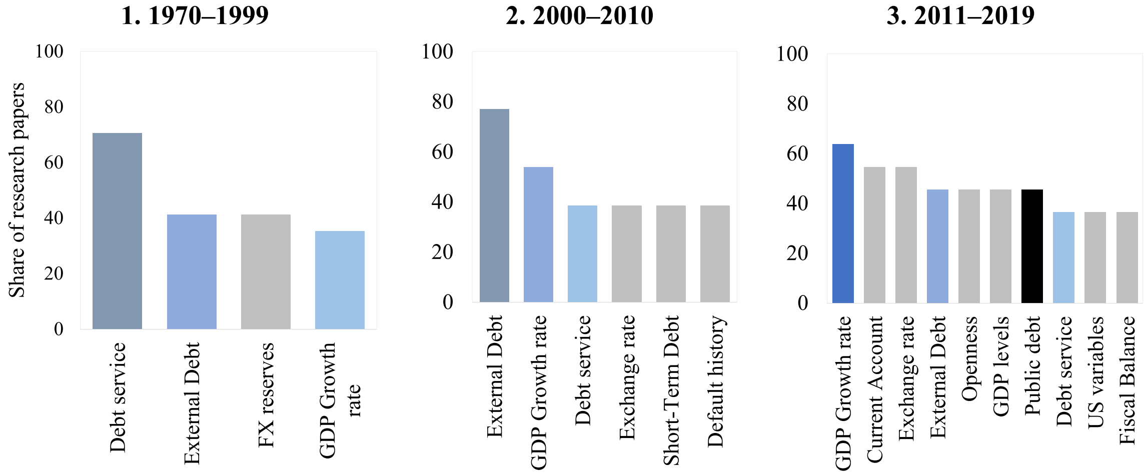

Fig. 1. Predictors of Fiscal Crises in the Literature. The charts are based on a literature review of 42 empirical papers (for more details, see Supplement 1.1). Variables plotted are those that are statistically significant in at least a third of the papers published during the period of reference.

Table 1
Most Common Predictors in the Literature. This table is based on a literature review of 42 empirical papers for the period 1970–2019 and shows the percentage of surveyed papers in which the corresponding variable is important/significant (for more details, see Supplement 1.1).

Variables	Share (%)
External Debt	72
Debt Service	68
GDP	63
FX reserves	50
Current Account	31
Exchange Rate	28
Openness	27
Political Variables	25
Public Debt	25
GDP per Capita	21
Short-Term Debt	21
Inflation	21
Default History	21
US Variables	21
Exports	17
Fiscal Balance	17
Imports	14
Private Debt	12
Commodity Prices	9
Government Expenditures	9
Savings	7
Cost of Rescheduling	4
Trade Balance	4
IMF Credit	4
Banking Crisis History	4

was limited to debt rescheduling or arrears on external debt. In some studies, the objective was to identify the determinants of the capacity to repay (Rivoli and Brewer, 1997; Lee, 1991; Berg and Sachs, 1988). The empirical strategy was often based on a logit model (with a small number of predictors) or linear regressions. The selection of predictors was often dictated by the financing mix of countries covered (mainly developing countries). Also, the attention was on external imbalances and not on public debt. Overall, external variables related to the capacity of a country to repay its obligations (external debt service; size of external debt[3], foreign exchange reserves to imports) were found to be among the most important predictors of crises (Fig. 1, panel 1).

[3] The papers over this period are often not explicit about the definition of external debt and whether it includes the external liabilities of both the private and public sectors.

2000–10. The definition of crisis expands to include not only debt defaults but also access to IMF programs above a certain quota. The previous definition was considered too restrictive as countries in distress might have been able to avoid defaults by getting official credit. The logit model remained the main tool to predict crises and identify drivers (Ciarlone and Trebeschi, 2005; Detragiache and Spilimbergo, 2001). However, a few papers tried different approaches, including early attempts at using machine learning techniques based on classification and regression trees (Manasse and Roubini, 2009) or neural networks (Fioramanti, 2008).[4]

Fiscal crises continued to be seen from the perspective of sovereign default on external creditors. As such, a large share of papers identifies external debt as a predictor of crises (Fig. 1, panel 2). Other common predictors included real GDP growth, debt service and the maturity of debt, exchange rate, and default history. There was little emphasis on public debt.

2011 onwards. In the aftermath of the global financial crisis, the attention is no longer only on external defaults. There is an acknowledgement that fiscal crises may reflect other types of distress and affect both external and domestic creditors. The definition of crisis now includes debt defaults, IMF programs, implicit debt defaults (high inflation, domestic arrears), and loss of market access (Sumner and Berti, 2017; Medas et al., 2018; Bruns and Poghosyan, 2018). There is more attention to the robustness of results—especially the out-of-sample predictive power—but it remains a weakness in the literature of early warning systems (Cerovic et al. 2018, Berg et al., 2005).

An important difference with our approach is that the empirical research remained constrained by the use of traditional econometric techniques. In particular, the preference was for parsimonious approaches relying on a limited set of indicators partly reflecting the priors of the researcher and difficulties addressing overfitting and data constraints. Among the most common predictors are the level of GDP or economic growth and external variables (current account, exchange rate, external debt and degree of openness). Public debt and fiscal-related variables are also identified but less frequently (Fig. 1, panel 3).

A growing number of papers examines the role of public debt, but the evidence is mixed. Savona and Vezzoli (2015) and Bruns and Poghosyan (2018) do not find evidence that public debt matters for predicting crises, while Cerovic et al. (2018) Sumner and Berti, 2017 find some evidence that it does but it is not robust across specifications. Changes in public debt are a significant predictor of debt crises in Reinhart and Rogoff (2011a) although the result does not hold for the post-World War II period.

The literature in general has been unable to explore complex dynamics. Even recent research using machine learning for predicting sovereign debt crises (Savona and Vezzoli, 2015; Alaminos et al., 2021), only considers a small set of indicators, in some cases ignores public debt, and does not analyze nonlinear interactions among specific predictors and how they affect the probability of a crisis—which we do in this paper.[5]

3. Data

3.1. Measuring fiscal crises

There is no common definition of fiscal crises in the literature, but most studies focus on sovereign debt crises triggered by external defaults (Detragiache and Spilimbergo, 2001; Chakrabarti and Zeaiter, 2014). In some instances, however, heightened budgetary distress may be associated with domestic arrears or inflation (Reinhart and Rogoff, 2011b), or a default is avoided thanks to official creditor assistance (Manasse et al., 2003). To capture these different facets, we follow Medas et al. (2018) and identify fiscal crises if any of the following four criteria is met (Appendix A):

1. *Credit events*. A crisis is triggered when the debt service is not paid on the due date, or the creditor incurs any other type of losses including through debt restructuring.
2. *Exceptionally large official financing*. Episodes where the country receives large financial support from the IMF or the European Union.
3. *Implicit domestic public debt default*. Two criteria are considered: (1) periods of high inflation (usually associated with monetary financing of the budget); or (2) accumulation of domestic arrears.
4. *Loss of market confidence*. Episodes associated with extreme market pressures as proxied by: (1) loss of market access, capturing sovereign defaults or bond issuance coming to a halt; or (2) very large borrowing costs or sovereign yield spikes.

[4] Manasse and Roubini (2009) is closer in spirit to our paper though there are important differences on methodology and data. In particular, they mainly rely on the classification and regression tree (CART) methodology, and therefore, the importance of predictors is highly dependent on the sample used. By contrast, we start with a much broader set of variables, using alternative feature selections algorithms to ensure the choice of predictors is robust to the sample selection. Our analysis also assesses how changes in predictors and their interactions affect the probability of a crisis.

[5] Our work is also part of a recent effort among academics and policymakers to develop early warning systems to predict different types of crises using machine learning. See for example, Holopainen and Sarlin (2017), Jarmulska (2020), Bluwstein et al. (2021), Hellwig (2021), International Monetary Fund (IMF) (2021), and Wang et al. (2021). Most of this literature, however, focuses on financial crises or does not look at non-linearities and interactions among predictors. A related strand of the literature has used machine learning to analyze sovereign risks with a narrower focus (CDS spreads) based on a small sample of European countries (see, Arakelian et al., 2019; Belly et al., 2021).

To construct the fiscal crisis variable, we update the Medas et al. (2018) dataset with the latest available information and data revisions for the period 1980–2018 and make several methodological improvements. First, we expand the coverage on sovereign debt yields. Second, we improve the information on domestic arrears where possible. Finally, the quality of the data is checked through contacts with IMF country teams. Although the number of crisis episodes is not materially different relative to previous data vintages (for example, International Monetary Fund (IMF), 2021), the start date changes for a significant number of them underscoring the importance of using as comprehensive and up-to-date information as possible.

Overall, we identify 384 crisis episodes for a sample of 188 countries over the period 1980–2018, making ours one of the most comprehensive studies of fiscal crises to date (Supplement 1.2). [6]On average, countries have experienced two fiscal crises since 1980 with more than three quarters of countries having at least one crisis (Table 2). Low-income developing countries (LIDCs) is the group with the highest frequency of crises—about two-thirds are in fiscal distress at any point in time—followed by emerging market economies (EMs)—on average, 40 percent. On the other hand, fiscal crises are rare events among advanced economies (AEs): less than 15 percent of them are in fiscal distress in any given year.

A cursory look at the data suggests that the 1990s was the decade with the highest concentration of crises. At the peak, about half of the countries (EMs for the most part) were in fiscal distress (Fig. 2). There was also some bunching in the early 1980s, reflecting falling commodity prices and rising global interest rates, and in 2010, following the global financial crisis—pointing to a potential role for global factors as precursors of fiscal crises. Overall, credit events are the most frequent type of crises accounting for close to two-thirds of episodes. Nonetheless, AEs are outliers relative to other income groups with most episodes associated with loss of market confidence and/or exceptional large official financing.

Although fiscal crises are usually not accompanied by other types of distress, in about a third of cases there is overlap with currency crises. Consistent with Reinhart (2002), we find that most of these cases relate to EMs and LIDCs, underscoring the importance of external financing among these countries. The synchronicity with financial crises is relatively low even though banking crises may put fiscal accounts under pressure through a contingent liability channel (Reinhart and Rogoff, 2011a). Triple crises are even less common.

3.2. Predictors

As discussed in Section 2, there is no consensus over the relative importance of alternative predictors of fiscal crises partly reflecting the diversity of methodologies and samples used in the literature. In addition, theoretical priors, and data constraints at the time of previous studies may have biased the selection of indicators. To address these shortcomings, we canvass the empirical and theoretical literature to identify potential predictors of crises.[7] As a result, our dataset covers a broad array of economic indicators and institutional country characteristics totaling 140 distinct variables.[8] Furthermore, the analysis uses several permutations of each variable—such as levels, differences, and lags (e.g., first and second lags, first and second differences, 5 or 10-year trailing differences)—and cross-sectional averages—allowing to capture dependencies arising from global factors or spillover effects. Overall, this yields 748 indicators encompassing among others: different measures of debt, economic activity, level of development, prices, fiscal aggregates, external indicators, global factors, demographics, and institutions (Supplement 1.3).

A contribution of this paper is to assemble a more comprehensive range of debt metrics to make a more robust assessment of their relevance as predictors of crises. We include both public debt and various indicators of private indebtedness. Accounting for these different forms of debt is important given their interaction at times of crises (Mbaye et al., 2018a). To construct consistent time series of debt, we leverage the Global Debt Database which includes private, public, and total debt for 190 countries going as far back as 1950 (Mbaye, Moreno Badia, and Chae, 2018b). We also capture some of the characteristics of public debt that have been identified as important in the sovereign debt crises literature. The data on external public debt involved extensive comparison of alternative sources of data to ensure consistency for a wide coverage of countries.

In our analysis we also construct a measure of the interest-growth differential variable (henceforth, "*r-g*"). The starting point is the recursive equation behind changes in the debt-to-GDP ratio (Escolano, 2010) with *r-g* defined as:

$$\left(\frac{r-g}{1+g}\right)$$

where r is the nominal effective interest rate and g the nominal GDP growth rate. We calculate the effective rates using consistent time series for the stock of public debt, which has been a challenge in the literature.[9] In some cases, the interest bill

[6] For a list of countries included in each income group, see Appendix B.

[7] A variable is included if 70 percent of the data exists. To take advantage of all available information, we impute missing values for any given variable with the training sample median.

[8] We do not use real time data as historical vintages are not readily available for most of our sample. Thus, our models are estimated based on actual data as of 2020.

[9] The effective interest rate is calculated as the ratio of the interest bill in period t and the stock of public debt (average of debt stocks in t and t-1). Accounting for the average debt stocks is important given large fluctuations within a year, particular at times of fiscal distress.

Table 2
Fiscal Crisis Episodes, 1980–2018. This table present some basic statistics on fiscal crisis episodes. Crisis starts can be associated with more than one criterion. Therefore, the breakdown does not need to add up to 100. A year is considered to be a fiscal crisis year when at least one of the four criteria is met. To separate between crisis events, we require at least two years of no fiscal crisis between the distinct events. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries.

	Total	AEs	EMs	LIDCs
Number of crisis starts	384	19	190	175
of which 1/(percent)				
Credit event	66.9	0.0	60.0	81.7
Exceptionally large official financing	21.4	42.1	21.6	18.9
Implicit domestic default	5.7	10.5	7.4	3.4
Loss of market confidence	17.7	84.2	24.2	3.4
Average per country	2.0	0.5	2.0	3.0
Number of countries with no fiscal crisis	45.0	24.0	18.0	3.0
Average duration	5.1	2.7	4.9	5.5

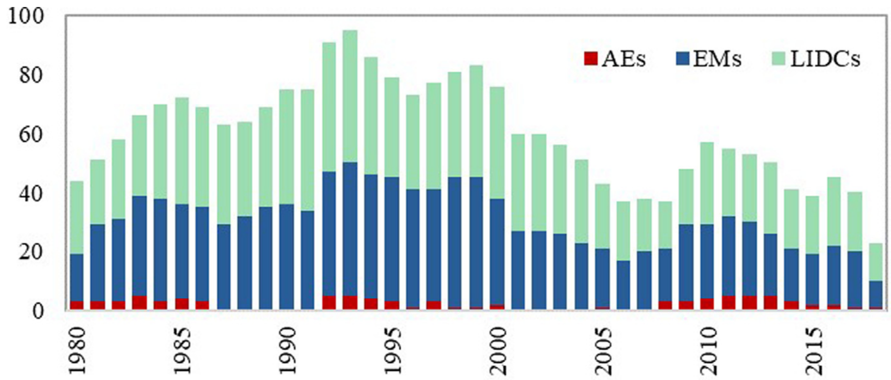

Fig. 2. Countries with Fiscal Crises, 1980–2018 (Number of Countries). AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries.

may refer to a broader perimeter of government than the stock of debt, which implies that the interest-growth differential may be overestimated (similar to other studies that look at long time series). However, LIDCs—the group for which debt stocks usually refers to the narrower perimeter of government—typically report the interest bill for the same level of government. Consistent with other studies (Mauro et al., 2015; Barrett, 2018; Escolano et al., 2017), our data suggests that, on average, *r-g* has been close to zero or negative since the 1980s (Fig. 3). However, there is a wide dispersion within each income group and positive *r-g* are not an anomaly.

Finally, we also compile data for various fiscal indicators (fiscal deficit, revenues, and expenditures).

4. Methodology

Our main objective is to identify a stable and robust set of predictors of fiscal crises from a large number of variables. As some indicators may be useful in predicting crises but only when interacting with other variables or in a non-linear way, it is important that the estimation strategy captures complex dynamics. Our model of choice is a random forest (Breiman, 2001)—henceforth RF—as it can deal with complexity and deliver significant improvements in crisis prediction problems relative to standard econometric approaches typically used in the early warning literature (International Monetary Fund (IMF), 2021; Bluwstein et al., 2021). Fernandez-Delgado et al. (2014) also find that, on average, RF is the best performer in a large scale empirical evaluation of 179 classification algorithms.

RF aggregates many decision trees, each run in a random sample of variables and country-years. By averaging the predictions of many trees, RF cancels out the noisy components of each tree, increasing the ability to predict on new data. The advantage is that RF can potentially incorporate a very large number of predictors without running into overfitting problems. The downside is that it makes it more difficult to distinguish relevant from irrelevant variables and to understand how each indicator affects the probability of a crisis (Degenhardt et al., 2019).

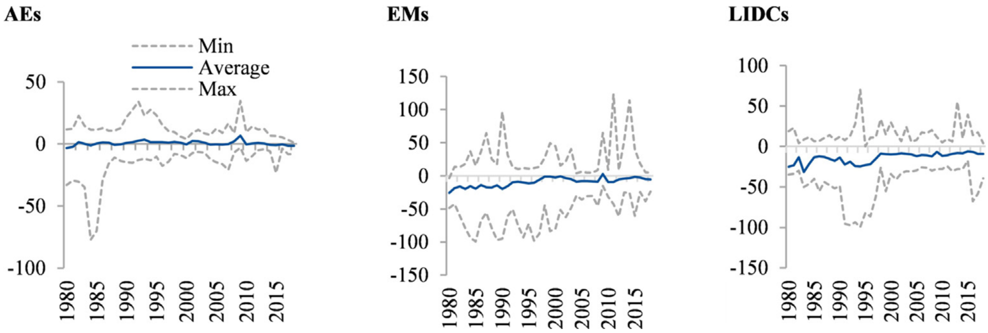

Fig. 3. Interest-Growth Differential, 1980–2018. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries.

Our empirical approach is to reduce the initial large set of predictors (748 variables) to the ones that contain more information than noise. To that end, we use various selection procedures to identify a stable set of predictors and diverse statistical techniques to analyze the importance of variables and ensure robustness.

4.1. Variable selection

Variable selection is a crucial issue in many applied classification and regression problems (e.g., Hastie et al., 2001). In our benchmark RF, we start with a large set of variables (748) to assign a probability of a crisis in the next two years. More specifically, let f be a predictive model:

$$\hat{y} = f(\boldsymbol{X})$$

where $\mathbf{X}$ is a matrix with n (annual) observations and m variables and $\hat{y} \in [0, 1]$ is the predictive probability of a fiscal crisis over the next two years. y takes value 1 if there is a crisis in the next two years and 0 otherwise.

To reduce the number of variables to those that are most relevant, we use "feature selection" algorithms—typically used as a workhorse in genomics research (Saeys et al., 2007; Ma and Huang, 2008; Hilario and Kalousis, 2008; Duval and Hao, 2010; Degenhardt et al., 2019). These algorithms identify the relative importance of features eliminating those that are unimportant. We focus on three algorithms built around RFs that have been widely used in the literature (see Appendix C for details): (i) *P-values computed with permutation importance (PIMP)* which selects relevant predictors based on repeated permutations of the outcome vector (likelihood of a crisis), leaving correlation patterns between predictors unchanged (see Altmann et al., 2010); (ii) *Recursive Feature Elimination (RFE)* which, by removing step-by-step the least important variables, aims to find a minimal set of variables for a good prediction model (Díaz Uriarte and Álvarez de Andres, 2006); and (iii) *Boruta* (Kursa and Rudnicki, 2010), which removes variables by comparing performance with shuffled copies of all variables (shadow features).

In choosing among these algorithms, we consider two criteria. First, the *predictive power*. Ideally, we want the empirical power of the smaller variable set to be close or as good as the full set. We compare the out-of-sample predictive performance of three RF models estimated using the variables selected by individual feature selection algorithms (PIMP/RFE/Boruta) against the RF estimated with the full set of variables. The main performance measure used is the area under the receiver-operator curve (AUROC), although we also report results using the log likelihood and mean squared errors (MSE). The AUROC also allows to benchmark our results against other studies. Intuitively, the AUROC assesses the accuracy of binary models against the alternative of a coin toss. A perfectly accurate model would display an AUROC of 1, while one with no predictive power over a coin toss would show a value of 0.5.

Secondly, it is also important to assess the *stability of variables selected* by each algorithm as these can vary due to small changes in the data (Degenhardt et al., 2019). To assess stability, we construct two separate samples by randomly dropping 5 percent of observations, comparing the overlap of the variables selected by each algorithm in each sample. We use the Pearson Correlation Coefficient to measure the overlap as it allows us to make comparisons between two sets of arbitrary cardinality (Nogueira and Brown, 2016). The Pearson coefficient takes values between -1 and 1, with 1 meaning perfect overlap between the sets.

4.2. Assessing variable importance

There are several methods in the literature to rank explanatory variables by their relative importance. In this paper, we use two approaches:

- *Out of bag permuted predictor importance*. The relative variable importance is estimated by measuring the increase in the prediction error after permuting a feature. A variable is "important" if shuffling its values increases the model error. At the opposite end, a variable is "unimportant" if shuffling its values leaves the model error unchanged. The model errors (with and without shuffling) are calculated on the oob sample.[10]
- *Shapley values*. Variables can be ranked by their contribution to the probability of a crisis using *Shapley values* (Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017). Similarly to cooperative game theory, Shapley values in the machine learning context measure each variable's contribution (payoff) to an individual predictions' deviation from the historical mean. They are constructed as the mean of each variable's marginal contribution to the forecast for every possible combination of other variables. To assess the discriminating value of a particular variable, we calculate the differences in Shapley values between crisis and non-crisis events.

4.3. Studying interactions and nonlinearities

To analyze non-linearities and heterogeneous interactions between various variables, helping open the black box and better understand the results, we rely on partial dependence plots (Greenwell, 2017; Friedman, 2001; Friedman and Popescu, 2008). A partial dependence plot (PDP) shows the marginal effect of one or several variables on the predicted outcome and can identify if the relationship between the predictor and the outcome is linear, monotonic or more complex. PDPs can either be a line plot (univariate) or a surface plot (bivariate). Univariate PDP shows the relationship between a variable and the predicted outcome (probability of a crisis), whereas a bivariate PDP helps to visualize the predicted outcome for a pair of variables (Appendix C). Intuitively, the partial dependence function at a particular feature value represents the average prediction if we force all data points to assume that feature value.

5. Results

5.1. Variable selection

There are wide-ranging differences in the variable selected by each feature selection algorithm—from 101 variables in the RFE, 157 in the PIMP, to 270 in Boruta (including permutations)—underscoring their different objectives and making it difficult to determine which one is the best a priori. Therefore, we start by comparing their out-of-sample predictive performance against the RF estimated using the full set of predictors (Table 3). Although we pool all countries for estimation purposes, results are disaggregated by income groups. Overall, we find that the full model performs better for AEs and EMs than for LIDCs, but the predictive power is higher than previous studies across the board. By way of comparison, the AUC is 0.82 for AEs and EMs and 0.73 for LIDCs while Cerovic et al. (2018) report a maximum AUC of 0.69 and 0.68 respectively.[11] As a robustness, we also run two alternative specifications of the RF using a smaller set of variables taken from Cerovic et al. (2018) and from Manasse and Roubini (2009) subject to data availability. Overall, the AUC is significantly lower than the RF using a full set of variables, ranging between 0.52 and 0.55 depending on country groupings, pointing to a potential loss of valuable information if we were to restrict the variable set.

The performance across the different feature selection algorithms is similar and close to the RF estimated with all variables (Table 3). Therefore, we can significantly reduce the number of predictors without reducing predictive power. Boruta tends to be somewhat superior when comparing the AUROC, especially for AEs and EMs. The marginally lower performance of PIMP and the RFE suggests some relevant predictors may have been dropped.

In choosing among alternative feature selection algorithms, we also want to ensure the stability in the choice of predictors as results can be sensitive to small perturbations in the sample (Nogueira and Brown 2016). The Pearson index is 0.94 for Boruta indicating that despite changing the sample, there is high overlap of the selected variables between replicates. Both PIMP (0.81) and RFE (0.59) have a lower index suggesting results are more sensitive to small changes in variables selected, consistent with findings in the literature. Given its predictive power and stability, we choose Boruta as the benchmark algorithm to study the relative importance of predictors.

5.2. Variable importance

Despite reducing the initial set by almost two thirds, the number of variables selected by Boruta is still large—consistent with the idea that crises are a complex phenomenon. As the predictive power of any individual indicator is likely to be small, it may be difficult to fully distinguish the impact of closely related variables. We group them into 22 categories to make it easier to interpret the results (see Supplement 1.3). Based on the out-of-bag permuted predictor importance[12], we find:

---

[10] In the RF, each tree is constructed using a different bootstrap sample from the original data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the k^{th} tree.

[11] Cerovic et al. (2018) is a natural benchmark as it is one of the few studies that has a large sample of countries and examines LIDCs in detail covering the period 1970–2015.

[12] While our analysis on variable importance (and interactions) relies on Boruta, we also show the variable importance from the other two feature selection algorithms in Supplement 1.4. Public debt and public debt service are always among the top three predictors irrespective of the algorithm.

Table 3
Out-of-Sample Performance of Alternative Feature Selection Algorithms. Models predict probability of crisis start occurring in year t+1 or t+2. For more details on methodology see Appendix C. RF-748=random forest using all (748) potential predictors. RFE= Recursive Feature Elimination; PIMP= P-values computed with permutation importance. AUROC=Area under the receiver operator curve; MSE= Mean squared error.

	Boruta	RFE	PIMP	RF-748
Advanced and emerging market economies				
AUROC	0.800	0.787	0.784	0.821
log(likelihood)	−0.283	−0.278	−0.283	−0.280
MSE	0.083	0.081	0.083	0.080
Low income developing countries				
AUROC	0.727	0.722	0.722	0.731
log(likelihood)	−0.514	−0.516	−0.517	−0.514
MSE	0.171	0.173	0.173	0.170

- Public debt service and public debt are among the most important groups of predictors (Fig. 4). This should not be surprising as fiscal crises by and large involve some degree of debt distress. But, as discussed in Section 2, the previous literature has only found weak evidence that public debt matters. By including a much broader set of debt measures and characteristics (for example, whether creditors are foreign) and accounting explicitly for nonlinearities and interactions among variables, we are able to capture the complex dynamics at play in the run up to a crisis.
- In line with the literature, external variables—in particular, external capital flows, current account, and to a lesser extent the exchange rate—are also important. This is consistent with the overlap of fiscal crises with currency crises which reflects periods when external investors perception on the sovereign ability to fulfill its debt obligations changes or when the inflow of capital or trade gets reversed.
- Institutional slow-moving variables are also ranked among the most important predictors. These include the level of development (GDP per capita), demographics, and to a lesser degree the quality of institutions. Stronger institutional frameworks likely help developed countries to better manage shocks and avoid crises. Nonetheless, it is likely that these variables are helping discriminate countries more prone to crises rather than the exact timing.
- Indicators of economic activity and inflation also have some predictive power. This may be explained by the fact that changes in economic activity often lead to budgetary difficulties or change in market perceptions, triggering a confidence crisis. However, the impact of these variables on the probability of a crisis could be higher if associated with other factors (see Section 5.3).
- Surprisingly, fiscal flow variables (deficits, revenues, spending) appear to be less important predictors of fiscal crises. A possible explanation is that those flows mainly affect the likelihood of a crisis through the buildup of public debt.
- Similarly, our results suggest that total external debt may be relatively less important than previously thought. That is, after taking into account the role of public external debt, which appears to be the main channel through which foreign liabilities may have an impact on fiscal crises, total external debt has limited predictive value. In addition, the interest-growth differential is among the less important predictors.

As a robustness test, we also look at Shapley differences as an alternative to measure the discriminating power of predictors between crisis and non-crisis observations, and differences across country groupings. Overall, we find that the ranking of variables for the full sample of countries is similar between the two approaches with only a few differences (e.g., crisis history and external debt service seems to be more important using Shapley differences). We also compare the Shapley differences for the pre-and post-2000 period. Overall, public and public debt service are among the top categories in both time periods.[13]

The importance of the predictors across income groups are broadly similar but with a few significant differences (Table 4). Public debt and public debt service are the most important group of predictors for EMs and LIDCs, together with the crisis history of a country, with external variables (current account and capital flows) being somewhat less important. The reverse order is true for AEs (public debt remains among the top categories). In addition, private debt is also among the top predictors for AEs suggesting that imbalances in other sectors of the economy eventually impact public finances. Namely, loss of confidence of external investors on the economy could affect the government's ability to fund itself. In addition, what may have started as a debt crisis in the private sector may end up on the balance sheet of the government (e.g., directly via bailout of banks or indirectly through the ensuing recession). These factors are also present among EMs and LIDCs, although to a lesser degree.

5.3. An analysis of selected predictors

In what follows, we undertake a more in-depth analysis of key leading predictors. We pay special attention to public debt, and specifically public external debt which has one of the highest predictive values and accounts for a large share of public

[13] Detailed results are available upon request.

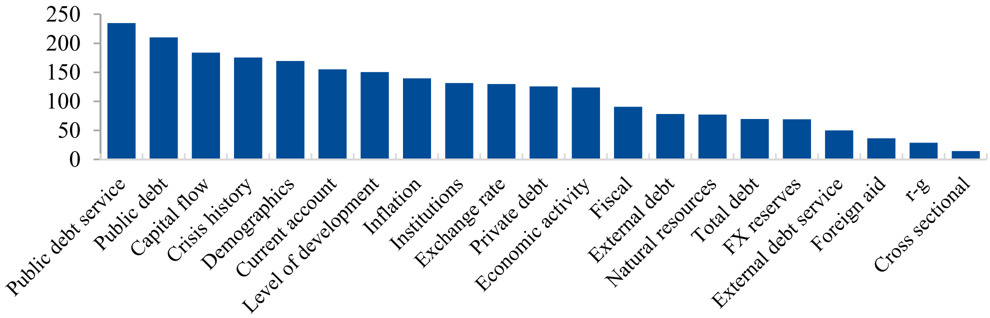

Fig. 4. Variable Importance by Group of Predictors. Variable importance is calculated using an in-built out of bag permuted predictor importance function in R based on the Random Forest estimated with the variables selected by Boruta.

Table 4
Contribution to Probability of Crisis: Shapley Values. This table shows the mean Shapley value difference (crisis versus non-crisis observations) by income group.

Advanced Economies		Emerging Market Economies		Low Income Developing Countries	
Variables	Shapley Value	Variables	Shapley Value	Variables	Shapley Value
Current account	0.094	Public debt service	0.094	Public debt service	0.083
Inflation	0.070	Crisis history	0.071	Public debt	0.078
Capital flows	0.066	Public debt	0.061	Crisis history	0.078
Private debt	0.061	External debt service	0.059	External debt service	0.054
Public debt	0.056	Current account	0.053	Demographics	0.050
Crisis history	0.053	Inflation	0.052	Level of development	0.040
Public debt service	0.043	Capital flows	0.043	Current account	0.039
Level of development	0.043	Demographics	0.037	Private debt	0.038
Institutions	0.041	Level of development	0.033	Capital flows	0.037
Demographics	0.033	Exchange rate	0.033	Inflation	0.037
Exchange rate	0.026	Institutions	0.032	Institutions	0.032
Total debt	0.025	Private debt	0.026	Exchange rate	0.028
External debt	0.024	Economic activity	0.024	Economic activity	0.027
Economic activity	0.021	Natural resources	0.021	Natural resources	0.022
Natural resources	0.020	External debt	0.018	External debt	0.019
Fiscal	0.014	FX reserves	0.017	Total debt	0.018
External debt service	0.013	Total debt	0.017	FX reserves	0.017
FX reserves	0.013	Fiscal	0.016	Fiscal	0.016
r-g	0.005	Foreign aid	0.007	Foreign aid	0.012
Cross sectional	0.004	r-g	0.006	r-g	0.006
Foreign aid	0.002	Cross sectional	0.005	Cross sectional	0.003

debt in EMs and LIDCs. We also study the interactions with other indicators; in particular, interest rates and inflation, as well as the potential role of external and financial imbalances.

Public debt

Fig. 5 (panels 1 and 2) shows the univariate PDP for public external debt, depicting how the average predicted probability of entering a crisis varies when public external debt changes. Overall, there is a positive non-linear relationship between the two that holds across all income groups although with some differences.[14] For AEs, the probability of a crisis is relatively flat until debt levels reach around 70 percent of GDP and rises steeply afterwards. For EMs, the estimated probability starts rising gradually when debt is above 40 percent of GDP and there is also a steeper rise around 70 percent. For LIDCs, predicted probabilities are much higher from the start, but there is also a steepening of the curve around 70 percent of GDP, although less than other income groups.

To get further insights as to when a probability is high enough to be concerned, we calculate the probability at which the model identifies a crisis. This is done by computing the threshold that minimizes the sum of type I and type II errors (missed crises and false alarms), an approach typically used in the literature (Berg et al., 2005).[15] The threshold for AEs is 9.9 percent, which results in capturing 85 percent of the crises while false alarms are kept at 27 percent. For EMs, the probability threshold is around 18 percent and for LIDCs 33 percent.

[14] One caveat of the PDP is that it relies on the assumption of independence among features. Therefore, the results may be biased if variables are highly correlated. The Accumulated Local Effects (ALE) solve this problem by calculating differences in predictions instead of averages (Apley, 2016). As a robustness check, we use the ALE approach and confirm our findings on the non-linearities of public external debt still hold.

[15] We also require that we miss at most one third of the crises.

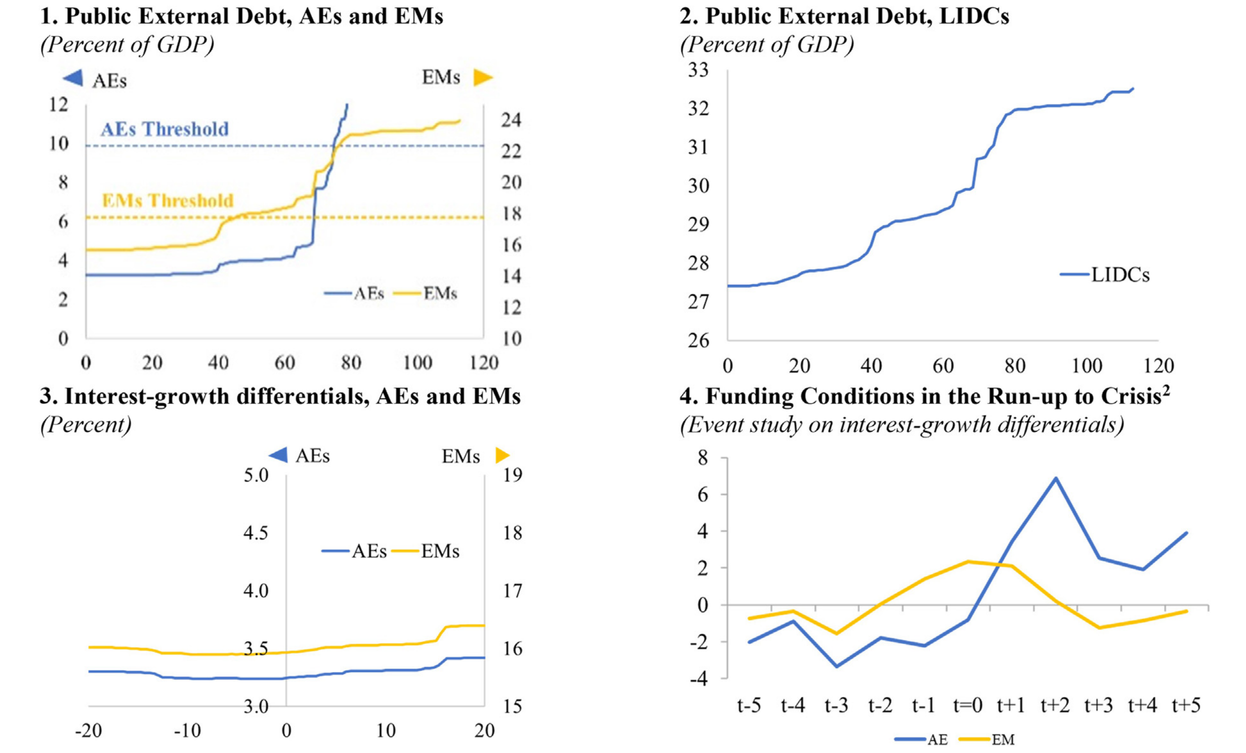

Fig. 5. Partial Dependence Plots and Event Studies. Charts (1)–(3) display Partial Dependence Plots based on the Boruta random forest. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt (charts 1 and 2) and the interest-growth differential (chart 3). Dotted lines show probability thresholds based on minimizing the sum of type I and type II errors. Chart 4 display an event study based on the framework developed by Gourinchas and Obstfeld (2012) where t = 0 is the start of the fiscal crisis. We estimate the equation$y_{i,t} = \alpha_i + \sum_{j=-5}^{5} \beta_{t+j} D_{i,t+j} + \varepsilon_{i,t}$where y is the interest-growth differential, and Di,t a dummy equal to 1 when the country is j periods away from the start of a crisis in period t and zero otherwise. Each data point is the interest-growth differential at time t + k, relative to "non-crisis" times benchmark.

Overall, public external debt is one of the few variables for which at high enough values (around 75 percent of GDP for AEs and 46 percent for EMs), the estimated probability breaches the crisis threshold regardless of other factors. Important to note that the thresholds can vary significantly depending on the method used. The approach in this paper (and in the literature) implicitly gives greater weight to not missing a crisis, given the high cost of these events. Using an alternative approach, to ensure that we only miss 20 percent of the crises, the debt threshold for AEs would remain similar, but would be around 70 percent of GDP for EMs (as we would miss more crises than other previous approach) and around 40 percent for LIDCs. These results point to the importance of public debt as a leading indicator of fiscal crises. But its importance is also related to interaction effects, which we explore next.

The interest-growth differential

A question commonly raised is what are the implications for debt sustainability, and the risk of a crisis, when interest rates are low (in some cases at the zero lower bound) especially for AEs (Mehrotra, 2017; Blanchard, 2019; Garín et al., 2019, Furman and Summers, 2020). The question is specifically related to the interest-growth differential (r-g). As growth is associated with tax buoyancy, it ultimately determines the ability to manage debt for a given interest rate, so the two are important to explain debt dynamics. The variable importance analysis suggests that r-g provides very limited information. This is confirmed by the PDPs: even for large variations of r-g, the estimated probability of a crisis barely changes (Fig. 5, panel 3). To explain this apparent puzzle, we conduct an event study to analyze the dynamics of *r-g* in the run-up to a crisis following Gourinchas and Obstfeld (2012). Overall, the event study shows that r-g can remain low for long and only surge at the onset of the crisis (especially for AEs) limiting its role as a leading indicator (Fig. 5, panel 4).[16]

The bivariate PDPs show that a low interest-growth differential does not dampen the risks of high debt (Fig. 6). Cells highlighted in red depict combinations of public external debt and r-g for which the estimated probability of a crisis is above the

[16] As showcased during the European sovereign debt crisis, these dynamics may partly reflect the spike in interest rates at the start of the distress episode (Beirne and Fratzscher, 2013; Bocola et al., 2019). Mauro and Zhou (2021) also argue that sovereign defaults may not necessarily be preceded by high (positive) interest-growth differentials.

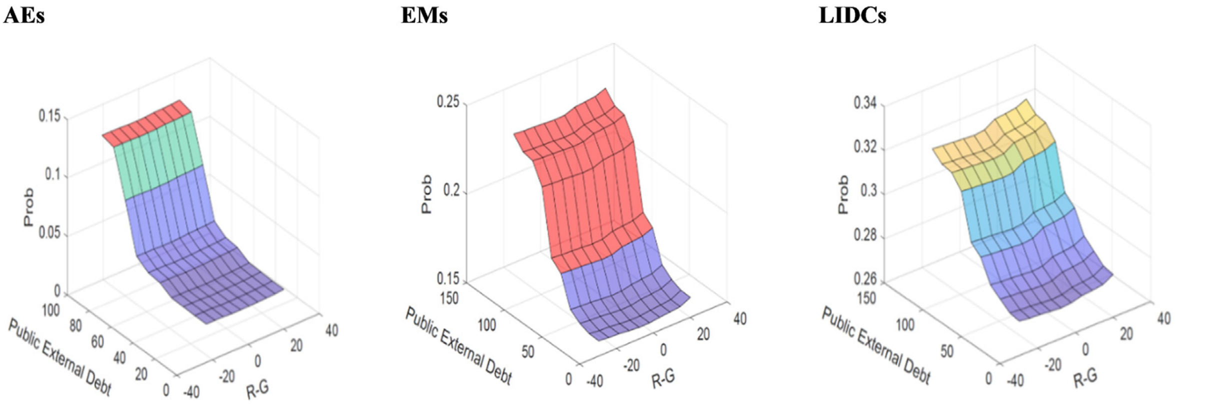

Fig. 6. Bivariate Partial Dependence Plots: Public External Debt and r-g. These charts display bivariate Partial Dependence Plots for different country groupings. Cells highlighted in red depict combinations of public external debt and the interest-growth differential for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing the sum of type I and type II errors (missed crises and false alarms). The darker the blue color, the lower the probability of a crisis. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

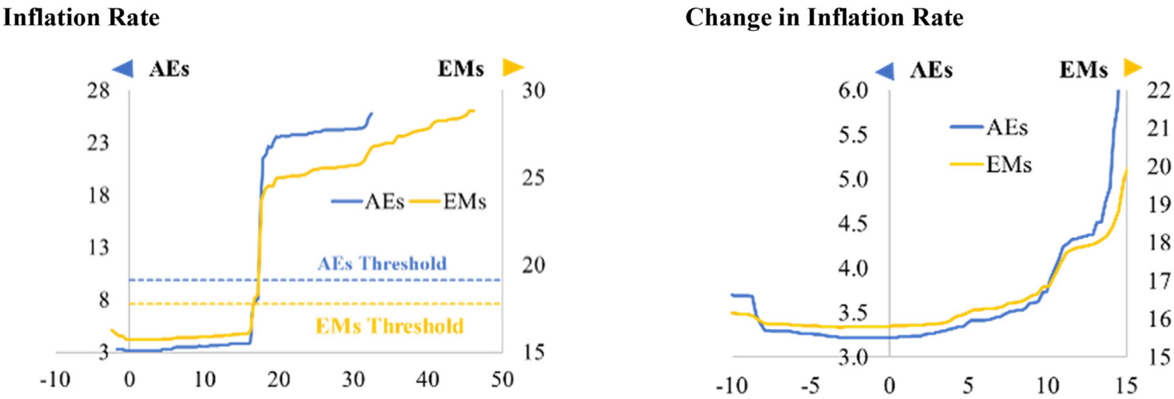

Fig. 7. Inflation: Univariate Partial Dependence Plots. These charts display univariate Partial Dependence Plots (PDP) based on the Boruta random forest. Estimated probabilities are plotted in the vertical axis and inflation in the horizontal axis. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt. AEs = Advanced economies; EMs = Emerging market economies.

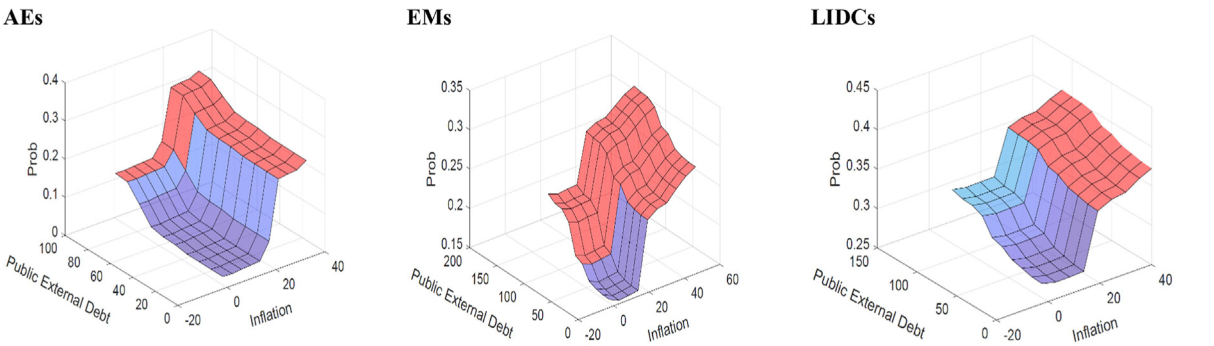

Fig. 8. Inflation and Public External Debt: Bivariate Partial Dependence Plot. These charts display bivariate Partial Dependence Plots for different country groupings. Cells highlighted in red depicts combinations of public external debt and inflation for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing the sum of type I and type II errors (missed crises and false alarms). The darker the blue color, the lower the probability of a crisis. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

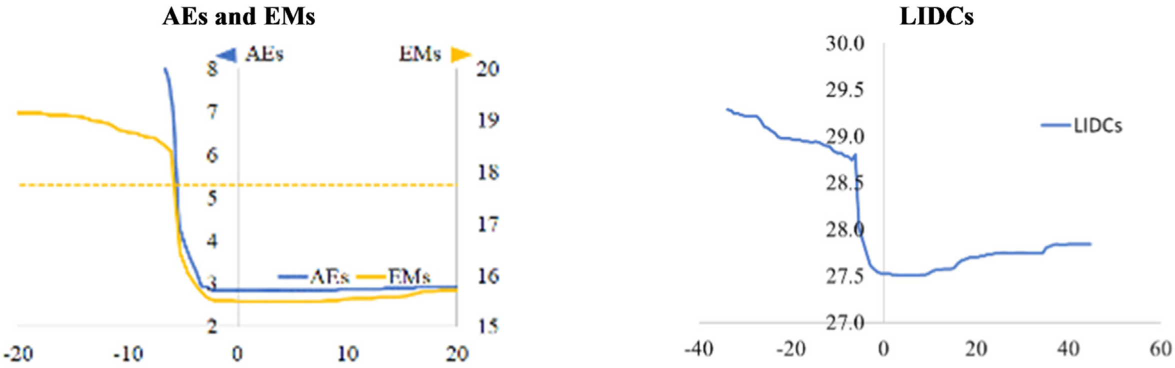

Fig. 9. Current Account: Univariate Partial Dependence Plots (Percent of GDP). These charts display univariate Partial Dependence Plots (PDP) based on the Boruta random forest. Estimated probabilities are plotted in the vertical axis and the current account in the horizontal axis. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries.

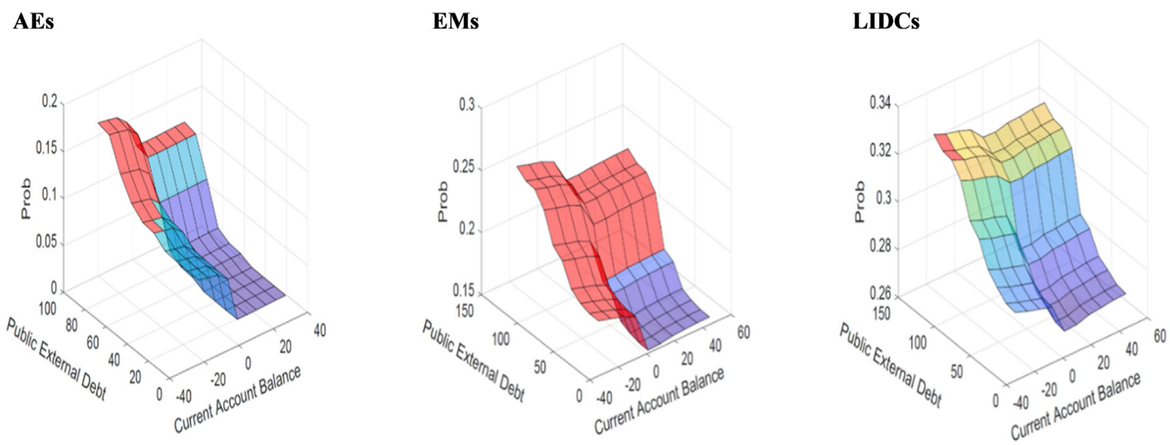

Fig. 10. Current Account and Public External Debt: Bivariate Partial Dependence Plot. Charts display bivariate PDPs for different country groupings. Cells highlighted in red depict combinations of public external debt and current account balance for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing the sum of type I and type II errors (missed crises and false alarms). The darker the blue color, the lower the probability of a crisis. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

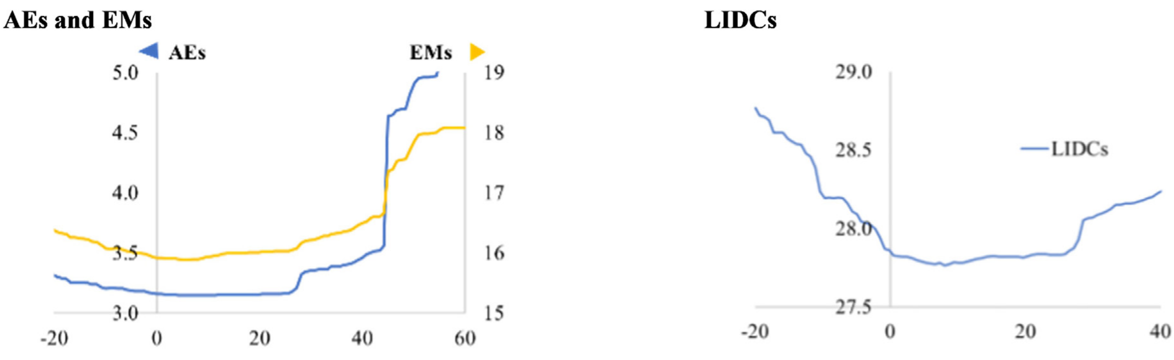

Fig. 11. Credit Gap: Univariate Partial Dependence Plots (Percent of GDP).These charts display the Partial Dependence Plots (PDPs) based on the Boruta random forest. Estimated probabilities are plotted in the vertical axis and the credit gap in the horizontal axis. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries.

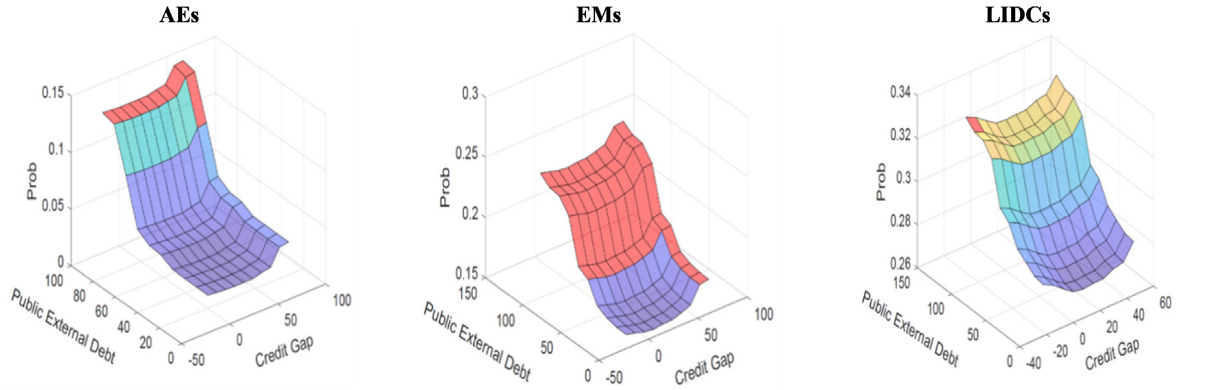

Fig. 12. Credit Gap and Public External Debt: Bivariate Partial Dependence Plots. Charts display bivariate Partial Dependence Plots for different country groupings. Cells highlighted in red depict combinations of public external debt and the credit gap for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing the sum of type I and type II errors (missed crises and false alarms). The darker the blue color, the lower the probability of a crisis. AEs = Advanced economies; EMs = Emerging market economies; LIDCs = Low income developing countries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

threshold calculated for the individual income group. In both AEs and EMs, we find that if public external debt is sufficiently high, the estimated probability breaches the crisis threshold irrespective of the interest-growth differential. One possible explanation is that governments may respond to periods of low *r-g* by increasing deficits, negating the potential benefits of low borrowing costs. That is, r-g is just one factor determining debt dynamics.

Inflation

For a long time, high inflation was associated with crises as countries resorted to the printing press to monetize fiscal deficits. But, as the years after the global financial crisis demonstrated, very low inflation can also be detrimental. The analysis suggests that there is a strong relationship between inflation and the estimated probability of crises, consistent with Reinhart and Rogoff (2011a). As with public debt, we find evidence of strong non-linearities. For AEs and EMs, the probability of crises increases significantly when inflation is above 17 percent (Fig. 7). Although the literature has established that countries with very high inflation tend to suffer from debt distress, we present evidence of the non-linear relationship.

Our results also suggest that both an increase and a decline in inflation can be associated with a higher probability of crises. As such, it provides some support for the potential risks of deflation and the snowball effect it can have on debt dynamics (Crafts, 2016). The levels of public external debt for which the estimated probabilities breach the crisis thresholds decrease with inflation (Fig. 8). This means that even for relatively low levels of debt, the probability of a crisis surges when inflation is high.

External and financial imbalances

To further explore the importance of external imbalances as driver of crises, we look at the current account balance. As with other indicators, we find a recurrent non-linear pattern. Once external deficits are between 3 and 5 percent of GDP, the probability of a crisis increases substantially (Fig. 9). For EMs, in particular, large deficits are associated with high likelihood of a crisis. There is also evidence of interactions between public external debt and the current account particularly for AEs (Fig. 10). Even for relatively moderate levels of debt, the probability of crises rises steeply when current account deficits are high. The opposite is not true. That is, current account surpluses do not appear to shield countries from crises if debt levels are high.

We also find some evidence that fiscal crises are associated with high leverage in the private sector although results are mixed depending on the country group. To capture financial imbalances, we focus on the credit gap (i.e. private debt as a share of GDP relative to the 10-year average). Our results suggest that the probability of a crisis increases significantly in AEs and EMs when the gap is above 40 percent (Fig. 11). We also find evidence of interactions with public external debt for EMs and AEs, with the estimated probability breaching the crisis thresholds for lower levels of debt if the credit gap is large (Fig. 12).

6. Conclusion

This paper contributes to the debate on the costs of public debt by revisiting its importance in predicting fiscal crises. Our results show that public debt is among the most important predictors while the interest-growth differential does not have much signaling value. Moreover, beyond certain debt levels, the likelihood of a crisis surges regardless. The *r-g* is only one factor determining debt dynamics and it tends to rise at the onset of the crisis when it is too late to help predict the occurrence of debt distress. As such, low levels of *r-g* should not be seen as a signal that debt levels are safe independently of the size.

Our analysis illustrates the value of using machine learning to analyze complex phenomena like fiscal crises. One of the main advantages is the ability to use a large set of predictors while avoiding overfitting problems. More importantly, this class of models enable the user to account for non-linearities and interaction effects that traditional econometric techniques are ill-equipped to deal with. In addition, we are able to look into the black box by analyzing the contributions of individual variables and interactions between predictors to the probability of a crisis. Our results show that there could be large information losses in early warning systems where these complex dynamics are ignored. Our findings underscore how broader imbalances in the economy eventually affect public finances increasing the risk of a fiscal crisis. For example, external imbalances and private sector indebtedness are important predictors for advanced economies. Other variables that matter include institutional factors and the history of crises, likely reflecting the government's credibility. Our results are also a reminder that relying on high inflation to address high debt is self-defeating as crises across all country groups often get triggered when inflation spikes. Finally, the interaction between public and private debt also deserves close scrutiny as the realization of contingent liabilities can be a source of fiscal stress. While this paper has taken a first step to explore these interactions, more research is needed to understand their nature (particularly in what refers to deficits, growth, and the domestic and external imbalances) and how they can lead to fiscal distress.

The machine learning techniques used in this paper do not allow us to establish causality. This is an area where computational science is still trying to make inroads. What we can confidently say is that there is a high correlation between public debt and crises and that this association is very robust.

Finally, these findings do not mean that bringing debt down is always the right policy prescription. At times of health and economic crises, increasing debt to fight the pandemic and for countercyclical purposes may have been desirable. However, the evidence presented in this paper points to the risks for those already highly indebted, suggesting that public debt might not be free after all.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are indebted to James Feigenbaum, Catherine Pattillo, Nobuo Yoshida, Kazusa Yoshimura, IMF colleagues, two anonymous referees, and participants at various seminars for helpful discussions and comments. Our paper builds on the project to revise the IMF's Vulnerability Exercise, especially the initial work (codes, data) done by Hellwig et al. (2019) and Medas et al. (2018). We gratefully acknowledge the many interactions with colleagues involved in the Vulnerability Exercise, including Chuqiao Bi, Klaus Hellwig, Andrew Hodge, Andre Botelho, and Daniela Marchettini. We would also like to thank Ade Adeyemi, Shuyi Liu, and Kadir Tanyeri for their support with coding. Juliana Gamboa Arbelaez provided excellent research assistance.

Appendix A. Fiscal crisis: Definitions and data sources

Event	Criterion	Thresholds			Literature	Sources	Notes
	Minimum two years gap between crises	AEs	EMEs	LIDCs			
(1) Credit Event	**Default, restructuring, or rescheduling** (i) of substantial size (in percent of GDP p.a.); AND	>0.5			Detragiache and Spilimbergo (2001); Chakrabarti and Zeaiter (2014); Reinhart and Rogoff (2011b)	BoC-BoE Sovereign Default Database complemented with information from IMF desks; Cruces and Trebesch (2013); World Bank	Episodes are mainly external defaults on sovereign debt denominated in foreign currency. Minimum requirements in terms of the size and accumulation of defaulted amounts are imposed to exclude small-technical defaults and avoid the perpetuation of a crisis being classified as a string of new events.
	(ii) defaulted nominal amount grows by a substantial amount (in percent p.a)	$\geq$ 10					
(2) Exceptionally large official financing	(i) **High-access IMF financial arrangement** with fiscal adjustment objective in place (in percent of quota); OR	$\geq$ 100			Baldacci et al. (2011)	IMF	Precautionary agreements are only considered when they become active with access above the threshold. IMF program data show that all high-access financial arrangements had fiscal adjustment as an overarching program objective (see Medas et al., 2018).
	(ii) **EU program**						
(3) Implicit domestic public default	(i) **High inflation rate** (in pct. of growth of annual average CPI p.a.) OR	$\geq$ 35	$\geq$ 100		Baldacci et al. (2011); Sturzenegger and Zettelmeyer (2006); and Fischer et al. (2002)	IMF (World Economic Outlook)	
	(ii) **Steep increase in domestic arrears** (in first difference of the ratio of 'other account payables (OAP)' to GDP in percentage points)	$\geq$ 1			Checherita-Westphal et al. (2015); Reinhart and Rogoff (2011a)	Eurostat; OECD (data on other accounts payable)	

Appendix A (*continued*)

Event	Criterion	Thresholds			Literature	Sources	Notes
	Minimum two years gap between crises	AEs	EMEs	LIDCs			
(4) Loss of market confidence	(i) **High price of market access** (in basis points of sovereign spreads or CDS spreads) OR						
	(a)Level of spreads (bps)	$\geq$ 1,000 bps			Sy (2004); Baldacci et al. (2011)	Reuters Datastream; Bloomberg	
	(b) Annual change in spreads (bps)	$\geq$ 300	$\geq$ 650	na			
	(ii) **Loss of market access**	when market access is lost (after maintaining market access for a 1/4 of the sample time and 2 consecutive years before the loss year)			IMF (2015); Kose et al. (2017)	Guscina et al. (2017); Gelos et al. (2004);	

Appendix B. Country classification

Advanced Economies	Emerging Markets		Low Income Developing Countries	
Australia	Albania	Lebanon	Afghanistan	Myanmar
Austria	Algeria	Libya	Bangladesh	Nepal
Belgium	Angola	Malaysia	Benin	Nicaragua
Canada	Antigua and Barbuda	Maldives	Bhutan	Niger
Cyprus	Argentina	Marshall Islands, Rep.	Burkina Faso	Nigeria
Czech Republic	Armenia	Mauritius	Burundi	Papua New Guinea
Denmark	Azerbaijan	Mexico	Cambodia	Rwanda
Estonia	Bahamas, The	Micronesia	Cameroon	São Tomé and Príncipe
Finland	Bahrain	Mongolia	Central African Republic	Senegal
France	Barbados	Montenegro	Chad	Sierra Leone
Germany	Belarus	Morocco	Comoros	Solomon Islands
Greece	Belize	Namibia	Congo, Dem. Rep. of	Somalia
Iceland	Bolivia	Oman	Congo, Republic of	South Sudan
Ireland	Bosnia and Herzegovina	Pakistan	Côte D'Ivoire	Sudan
Israel	Botswana	Palau	Djibouti	Tajikistan
Italy	Brazil	Panama	Eritrea	Tanzania
Japan	Brunei Darussalam	Paraguay	Ethiopia	Timor Leste
Korea	Bulgaria	Peru	Gambia, The	Togo
Latvia	Cape Verde	Philippines	Ghana	Uganda
Lithuania	Chile	Poland	Guinea	Uzbekistan
Luxembourg	China	Qatar	Guinea-Bissau	Vietnam
Malta	Colombia	Romania	Haiti	Yemen, Republic of
Netherlands	Costa Rica	Russia	Honduras	Zambia
New Zealand	Croatia	Samoa	Kenya	Zimbabwe
Norway	Dominica	Saudi Arabia	Kiribati	
Portugal	Dominican Republic	Serbia	Kyrgyz Republic	
San Marino	Ecuador	Seychelles	Lao PDR	
Singapore	Egypt	South Africa	Lesotho	
Slovak Republic	El Salvador	Sri Lanka	Liberia	
Slovenia	Equatorial Guinea	St. Kitts and Nevis	Madagascar	
Spain	Fiji	St. Lucia	Malawi	
Sweden	FYR Macedonia	St. Vincent and the Grenadines	Mali	
Switzerland	Gabon	Suriname	Mauritania	
United Kingdom	Georgia	Swaziland	Moldova	
United States	Grenada	Syria	Mozambique	
	Guatemala	Thailand		
	Guyana	Tonga		
	Hungary	Trinidad and Tobago		
	India	Tunisia		
	Indonesia	Turkey		
	Iran	Turkmenistan		
	Iraq	Tuvalu		
	Jamaica	Ukraine		
	Jordan	United Arab Emirates		
	Kazakhstan	Uruguay		
	Kosovo	Vanuatu		
	Kuwait	Venezuela		

Appendix C. Methodological details

Following the literature on crises, we choose a prediction window of two years (see, for example, Berg and Pattillo, 1999). Since we are interested in the transition from non-crisis to crisis state, we follow Bussiere and Fratzscher (2006) and only consider observations in which a country is not in a crisis in year *t* and drop all crisis years after the start of a crisis episode. We follow the standard practice in the literature and pool all countries to make use of the largest possible training samples and capture a wide variety of crises (see, for example, Fuertes and Kalotychou, 2006).

To estimate the probability of a crisis, we rely on a RF—an ensemble learning method based on decision trees. Trees are constructed through two random perturbation mechanisms: (1) each tree is trained on a bootstrap sample; (2) optimal variables at each split are identified from a random subset m_{try} of explanatory variables from the m predictors (i.e., $m_{try} < m$). The prediction for each leaf is the mean outcome for the observations on that leaf, and trees are fit to minimize mean squared errors. The overall prediction of the RF is the average prediction of all trees (see details below).

Calculation of the predictive power

Models were evaluated based on the out-of-sample predictive performance. For that purpose, we split the sample into two non-overlapping sub-samples: *training* (for model estimation) and *test* (for out-of-sample evaluation). To avoid possible information spillovers from the test to the training sample, we use a rolling cutoff year (beginning in 1999) to separate between the two. That is, we start by training a model with data for 1980–1999 and then roll forward both the estimation and the testing periods, adding one year at a time in each iteration. Therefore, the entire sample is split into 17 independent *training* and *test* sub-samples, each one based on a larger training sample than the previous one.

Next step involves running a baseline RF on the training sample. Each decision tree of the random forest is an interpretable model that successively splits the data into subsets by testing a single predictor at each node. Starting at the root node of the tree, all observations are divided into subsets, called leaves, based on variable cutoffs. Trees are constructed through two random perturbation mechanisms: (1) each tree is trained on a bootstrap sample; (2) optimal variables at each split are identified from a random subset $\mathrm{m}_{\mathrm{try}}$ of explanatory variables from the m predictors (i.e. $\mathrm{m}_{\mathrm{try}} < \mathrm{m}$). The prediction for each leaf is the mean outcome for the observations on that leaf, and trees are fitted to minimize mean squared errors. The overall prediction of the RF is the average prediction of all trees.

For each training sample, we use k-fold cross validation to choose the optimal $\mathrm{m}_{\mathrm{try}}$, where k is the number of years of that training sample. k-fold cross validation simulates out-of-sample prediction. We choose the $\mathrm{m}_{\mathrm{try}}$ to minimize out-of-sample log-likelihood loss. For the baseline model (RF with 748 variables), $\mathrm{m}_{\mathrm{try}}$ is chosen from a grid of candidate values. No other restrictions are placed on the tree growing process, so that each tree is grown exhaustively. The number of trees is set to 2000.

The optimal $\mathrm{m}_{\mathrm{try}}$ (estimated in the training sample using the baseline RF model with 748 variables) then feeds into each of the three individual feature selection algorithms (Boruta/PIMP/RFE). These algorithms will then select a reduced number of variables based on their importance (see details below). Once these variables have been identified, the individual predictive performance of each algorithm (Boruta, PIMP and RFE) is calculated by estimating a RF model in the training sample and then using it to predict probabilities in the out-of-sample test dataset and compute the model implied AUROC, log likelihood, and MSE scores. (Figure Appendix C.1).

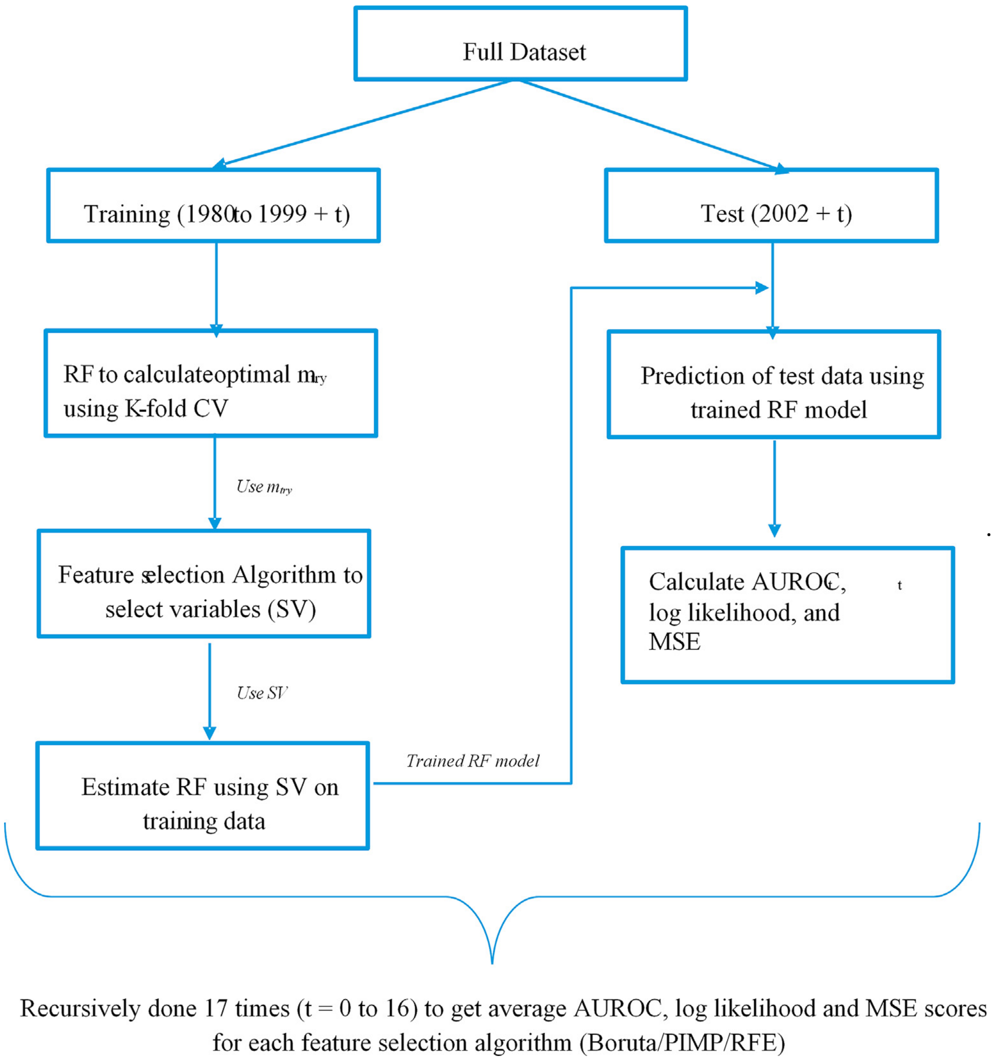

Figure Appendix C.1. Flow-chart showing the methodology to calculate the predictive power of alternative random forest models estimated with the variables selected by Boruta, PIMP, and RFE.

The above mentioned step is recursively run 17 times on independently split training and out-of-sample test datasets to calculate 17 different AUROC, log likelihood, and MSE scores for the three feature selection algorithms. We finally compare the average out-of-sample predictive performance of three RF models estimated using the features selected by each algorithm against the RF estimated with the full set of variables.

Feature selection algorithms

P-values computed with permutation importance (PIMP). Altmann et al. (2010) developed this method to correct for plausible biases when a large number of categorical variables are used and to select relevant predictors based on repeated permutations of the outcome vector (likelihood of a crisis), leaving correlation patterns between predictors unchanged. Main advantages of permuting the response vector are that the dependence between the predictor variables remains unchanged and the number of permutations can be much smaller than the number of predictor variables. For each permutation of the outcome, the importance for all predictor variables is assessed. This leads to a vector of importance measures for every variable, called the "null importance". The PIMP algorithm fits a probability distribution to the population of null importance vector (such as normal, lognormal, or gamma). Parameters of these distributions are estimated using maximum likelihood methods and P-values are calculated as the probability of observing an importance score that is larger than the original importance score under the estimated distribution. Only significant predictors are kept.

Recursive Feature Elimination (RFE). RFE aims to find a minimal set of variables which leads to a good prediction model (Díaz Uriarte and Álvarez de Andres, 2006). It starts with a RF built on all variables. A specific proportion of the least important variables is then removed, and a new RF is generated using the remaining variables. These steps are recursively applied until the out-of-bag (obb) predictive error is larger than the initial/previous oob error. At each step the prediction performance is estimated based on the out-of-bag samples that were not used for model building. The set of variables that leads to the RF with the smallest oob error or to an error within a small range of the minimum is selected.

Boruta. This algorithm was developed to identify all relevant variables within a classification framework (Kursa and Rudnicki, 2010). It compares the importance of the real predictor variables with those of random so-called shadow variables. For each real variable a statistical test is performed comparing its importance with the maximum value of all the shadow variables. Variables with significantly larger (smaller) importance values are declared as important (unimportant). All unimportant variables and shadow variables are removed and the previous steps are repeated until all variables are classified or a pre-specified number of runs has been performed.

Partial Dependence Plots

To understand how PDPs are calculated, consider a predictor set $X = \{x_1, x_2, x_3, x_n\}$. We construct a subset X^S which would either contain $\{x_1\}$ or $\{x_1, x_2\}$ depending on if we want to generate univariate PDPs or bivariate PDPs. Bivariate PDPs are generally used to study interactions between two variables. Let X^C be the complementary set of X^S in X. A PDP of a predicted response variable in X is defined by the expectation of predicted responses with respect to X^C:

$$fS(XS) = EC[f(XS, XC)] = Zf(XS, XC)pC(XC)dXC$$

where $pC(XC)$ is the marginal probability of X^C. PDP works by marginalizing the model output over the distribution of the feature in set C to show the relationship between the variable of interest and the outcome. By marginalizing over other features and assuming that each observation is equally likely, we get the following function to estimate the partial dependence using the observed predictor data:

$$fS(XS) \approx \frac{1}{N}\sum_{i=1}^{N} f\left(XS, X_i^C\right)$$

where N is the number of observations and $X_i = \left(X_i^S, X_i^C\right)$ in the ith observation. $fS(XS)$ is the partial dependence plot for X^S. If two variables say X_j and X_k do not interact with each other, then the partial dependence function can be decomposed into the sum of individual PDPs, but it would not be the case if X_j and X_k interacts. In that case, the bivariate PDPs cannot be expressed as the sum of univariate PDPs:

$$PD_{jk}\ (X_j, X_k) = PD_j\ (X_j) + PD_k\ (X_k).$$

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jimonfin.2022.102654.

References

Abbas, S.A., Belhocine, N., ElGanainy, A., Horton, M., 2011. A Historical Public Debt Database. IMF Econ. Rev. 59 (4), 717–742.
Alaminos, D., Peláez, J., Salas, M., Fernández-Gámez, M., 2021. Sovereign debt and currency crises prediction models using machine learning techniques. Symmetry 13 (4), 652.
Altmann, A., Tolosi, L., Sander, O., Lengauer, T., 2010. Permutation Importance: A Corrected Feature Importance Measure. Bioinformatics 26 (10), 1340–1347.
Apley, D.W. 2016. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. arXiv preprint arXiv:1612.08468.
Arakelian, V., Dellaportas, P., Savona, R., Vezzoli, M.M., 2019. Sovereign Risk Zones in Europe During and After the Debt Crisis. Quant. Financ. 19 (6), 961–980.
Asonuma, T., M. Chamon, A. Erce, and A. Sasahara. 2019. "Costs of Sovereign Defaults: Restructuring Strategies, Bank Distress and the Capital Inflow-Credit Channel," IMF Working Paper No. 19/69 (International Monetary Fund).
Barrett, P. 2018. Interest-Growth Differentials and Debt Limits in Advanced Economies, IMF Working Paper No. 19/82 (International Monetary Fund).
Beirne, J., Fratzscher, M., 2013. The Pricing of Sovereign Risk and Contagion during the European Sovereign Debt Crisis. J. Int. Money Financ. 34, 60–82.
Belly, G., Boeckelmann, L., Caicedo Graciano, C.M., Di Iorio, A., Istrefi, K., Siakoulis, V., Stalla-Bourdillon, A., 2021. Forecasting Sovereign Risk in the Euro Area via Machine Learning. Available at SSRN: https://ssrn.com/abstract=3974515 or https://doi.org/10.2139/ssrn.3974515.
Berg, A., Sachs, J., 1988. The Debt Crisis Structural Explanations of Country Performance. J. Dev. Econ. 29 (3), 271–306.
Berg, A., Borensztein, E., Pattillo, C., 2005. Assessing Early Warning Systems: How Have They Worked in Practice? IMF Staff Papers 52 (3), 462–502.
Berg, A., Pattillo, C., 1999. Predicting Currency Crises: The Indicators Approach and An Alternative. J. Int. Money Financ. 18, 561–586.
Breiman, L., 2001. Random Forest. Mach. Learn. 45 (1), 5–32 http://oz.berkeley.edu/users/breiman/randomforest2001.pdf.
Blanchard, O., 2019. Public Debt and Low Interest Rates. Am. Econ. Rev. 109 (4), 1197–1229.
Bluwstein, K., Buckman, M., Joseph, A., Kapadia, S., Şimşek, O., 2021. "Credit Growth, the Yield Curve, and Financial Crisis Prediction: Evidence from a Machine Learning Approach," European Central Bank, Working Paper No. 2614.
Bocola, L., Bornstein, G., Dovis, A., 2019. Quantitative Sovereign Default Models and the European Debt Crisis. J. Int. Econ. 118, 20–30.
Bruns, M., Poghosyan, T., 2018. Leading Indicators of Fiscal Distress: Evidence from the Extreme Bound Analysis. Appl. Econ. 50 (13), 1454–1478.
Bussiere, M., Fratzscher, M., 2006. Towards a New Early Warning System of Financial Crises. J. Int. Money Financ. 25 (6), 953–973.
Cerovic, S., K. Gerling, A. Hodge, and P. Medas. 2018. "Predicting Fiscal Crises", IMF Working Paper No. 18/181 (Washington: International Monetary Fund).
Chakrabarti, A., Zeaiter, H., 2014. The Determinants of Sovereign Default: A Sensitivity Analysis. Int. Rev. Econ. Financ. 33, 300–318.
Checherita-Westphal, C., Klemm, A.A., Viefers, P., 2015. "Governments' Payment Discipline: The Macroeconomic Impact of Public Payment Delays and Arrears," IMF Working Paper No. 15/13 (Washington: International Monetary Fund).
Ciarlone, A., Trebeschi, G., 2005. Designing an Early Warning System for Debt Crises. Emerg. Markets Rev. 6, 376–395.
Cruces, J.J., Trebesch, C., 2013. Sovereign Defaults: The Price of Haircuts. Am. Econ. J. Macroecon. 5 (3), 85–117.
Crafts, N., 2016. Reducing High Public Debt Ratios: Lessons from UK Experience. Fiscal Stud. 37 (2), 201–223.
Degenhardt, F., Seifert, S., Szymczak, S., 2019. Evaluation of Variable Selection Methods for Random Forests and Omics Data Sets. Briefings Bioinf. 20 (2), 492–503.

Detragiache, E., Spilimbergo, A., 2001. "Crises and Liquidity: Evidence and Interpretation," IMF Working Paper No. 01/02 (Washington: International Monetary Fund).
Díaz Uriarte, R., Álvarez de Andres, S., 2006. Gene Selection and Classification of Microarray Data using Random Forest. BMC Bioinf. 7, 3.
Duval, B., Hao, J., 2010. Advances in Metaheuristics for Gene Selection and Classification of Microarray Data. Briefings Bioinf. 11 (1), 127–141.
Escolano, J., 2010. A Practical Guide to Public Debt Dynamics, Fiscal Sustainability, and Cyclical Adjustment of Budgetary Aggregates. Technical Notes and Manuals 10/02. Washington, DC: International Monetary Fund.
Escolano, J., Shabunina, A., Woo, J., 2017. The Puzzle of Persistently Negative Interest-Rate–Growth Differentials: Financial Repression or Income Catch-Up? Fiscal Stud. 38 (2), 179–217.
Feder, G., Just, R., Ross, K., 1981. Projecting Debt Servicing Capacity of Developing Countries. J. Financ. Quant. Anal. 16 (5), 651–669.
Fernandez-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? J. Mach. Learn. Res. 15 (1), 3133–3181.
Fioramanti, M., 2008. Predicting Sovereign Debt Crises using Artificial Neural Networks: A Comparative Approach. J. Financ. Stab. 4, 149–164.
Fischer, S., Sahay, R., Vegh, C., 2002. Modern Hyper- and High Inflations. J. Econ. Lit. 40, 837–880.
Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29 (5), 1189–1232.
Friedman, J.H., Popescu, B.E., 2008. Predictive Learning Via Rule Ensembles. Ann. Appl. Stat. 2 (3), 916–954.
Fuertes, A., Kalotychou, E., 2006. Early Warning Systems for Sovereign Debt Crises: The Role of Heterogeneity. Comput. Stat. Data Anal. 51, 1420–1441.
Furman, J., Summers, L., 2020. A Reconsideration of Fiscal Policy in the Era of Low Interest Rates. Brookings Discussion Draft.
Garín, J., Lester, R., Sims, E., Wolff, J., 2019. Without Looking Closer, It May Seem Cheap: Low Interest Rates and Government Borrowing. Econ. Lett. 180, 28–32.
Gelos, R., Sahay, R., Sandleris, G., 2004, Sovereign Borrowing by Developing Countries: What Determines Market Access? IMF Working Paper No. 04/211 (Washington: International Monetary Fund).
Gourinchas, P., Obstfeld, M., 2012. Stories of the Twentieth Century for the Twenty-First. Am. Econ. J. Macroecon. 4 (1), 226–265.
Greenwell, B.M., 2017. Pdp: An R Package for Constructing Partial Dependence Plots. R J. 9 (1), 421–436 https://journal.r-project.org/archive/2017/RJ-2017-016/index.html.
Guscina, A., Sheheryar, M., Papaioannou, M., 2017. "Assessing Loss of Market Access: Conceptual and Operational Issues," IMF Working Paper No. 17/246 (Washington: International Monetary Fund).
Hajivassiliou, V.A., 1987. The External Debt Repayments Problems of LDC's: An Econometric Model Based on Panel Data. J. Econometrics 36 (1–2), 205–230.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Data Mining, Inference. and Prediction. Springer-Verlag, New York.
Hellwig, K. 2021. "Predicting Fiscal Crises: A Machine Learning Approach," IMF Working Paper No. 21/150 (International Monetary Fund).
Hellwig, K., Moreno Badia, M., Gupta, P., Medas, P., 2019. Predicting Fiscal Crises: A Machine Learning Approach. Unpublished manuscript.
Hilario, M., Kalousis, A., 2008. Approaches to Dimensionality Reduction in Proteomic Biomarker Studies. Briefings Bioinf. 9 (2), 102–118.
Holopainen, M., Sarlin, P., 2017. Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty. Quant. Financ. 17 (12), 1933–1963.
International Monetary Fund (IMF), 2015. The Fund's Lending Framework and Sovereign Debt—Further Considerations. Board Paper, Washington, DC: International Monetary Fund.
International Monetary Fund (IMF). 2020. Policies to Support People During the Covid-19 Pandemic. IMF Fiscal Monitor (April) Chapter 1, Washington, DC: International Monetary Fund.
International Monetary Fund (IMF), 2021. How to Assess Country Risk: The Vulnerability Exercise Approach Using Machine Learning. Technical Notes and Manuals 21/03. International Monetary Fund, Washington, DC.
Jarmulska, B., 2020. "Random Forest versus Logit Models: Which offers Better Early Warning of Fiscal Stress," European Central Bank, Working Paper No. 2408.
Jordà, Ò., Schularick, M., Taylor, A.M., 2016. Sovereigns versus Banks: Credit, Crises, and Consequences. J. Eur. Econ. Assoc. 14 (1), 45–79.
Kose, A., Kurlat, S., Ohnsorge, F., Sugawara, N., 2017. A Cross-Country Database of Fiscal Space. World Bank Development Prospects Group, Policy Research Working Paper No. 8157.
Kursa, M., Rudnicki, W., 2010. Feature Selection with the Boruta Package. J. Stat. Softw. 36 (11), 1–13.
Lee, S.H., 1991. Ability and Willingness to Service Debt as Explanation for Commercial and Official Rescheduling Cases. J. Bank. Finance 15 (1), 5–27.
Lundberg, S.M., Lee, S., 2017. A Unified Approach to Interpreting Model Predictions. Adv. Neural Inform. Process. Syst., 4765–4774
Ma, S., Huang, J., 2008. Penalized Feature Selection and Classification in Bioinformatics. Briefings Bioinf. 9 (5), 392–403.
Manasse, P., Roubini, N., Schimmelpfennig, A., 2003. "Predicting Sovereign Debt Crises," IMF Working Paper No. 03/221 (Washington: International Monetary Fund).
Manasse, P., Roubini, N., 2009. Rules of Thumb for Sovereign Debt Crises. J. Int. Econ. 78 (2), 192–205.
Mauro, P., Romeu, R., Binder, A., Zaman, A., 2015. A Modern History of Fiscal Prudence and Profligacy. J. Monetary Econ. 76, 55–70.
Mauro, P., Zhou, J., 2021. r-g<0: Can We Sleep More Soundly? IMF Econ. Rev. 69, 197–229.
Mbaye, S., M. Moreno-Badia, M., and K. Chae. 2018a. "Bailing Out the People: When Private Debt Becomes Public." IMF Working Paper No. 18/141 (Washington: International Monetary Fund).
Mbaye, S., Moreno-Badia, M.M., Chae, K., 2018b. The Global Debt Database: Methodology and Sources. IMF Working Paper No. 18/111 (Washington: International Monetary Fund).
Medas, P., Poghosyan, T., Xu, Y., Farah-Yacoub, J., Gerling, K., 2018. Fiscal Crises. J. Int. Money Financ. 88, 191–207.
Mehrotra, N., 2017. "Debt Sustainability in a Low Interest Rate World," Hutchins Center Working Paper No. 32.
Nogueira, S., Brown, G., 2016. Measuring the Stability of Feature Selection. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken J. (Eds.), Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016. Lecture Notes in Computer Science, vol. 9852. Springer.
Reinhart, C.M., 2002. Default, Currency Crises, and Sovereign Credit Ratings. World Bank Econ. Rev. 16 (2), 151–170.
Reinhart, C.M., Rogoff, K., 2009. This Time is Different: Eight Centuries of Financial Folly. Princeton University Press, Princeton, NJ.
Reinhart, C.M., Rogoff, K., 2011a. From Financial Crash to Debt Crisis. Am. Econ. Rev. 101 (5), 1676–1706.
Reinhart, C.M., Rogoff, K., 2011b. The Forgotten History of Domestic Debt. Econ. J. 121 (552), 319–350.
Rivoli, P., Brewer, T.L., 1997. Political Instability and Country Risk. Glob. Financ. J. 8 (2), 309–321.
Saeys, Y., Inza, I., Larranaga, P., 2007. A Review of Feature Selection Techniques in Bioinformatics. Bioinformatics 23 (19), 2507–2517.
Savona, R., Vezzoli, M., 2015. Fitting and Forecasting Sovereign Defaults using Multiple Risk Signals. Oxford Bull. Econ. Stat. 77 (1), 66–91.
Solberg, R.L., 1988. Sovereign Rescheduling: Risk and Portfolio Management. Routledge.
Strumbelj, E., Kononenko, I., 2010. An Efficient Explanation of Individual Classifications Using Game Theory. J. Mach. Learn. Res. 11, 1–18.
Sturzenegger, F., Zettelmeyer, J., 2006. Debt Defaults and Lessons from a Decade of Crises, Table 1, Chapter 1. MIT Press, Cambridge.
Sumner, S.P., Berti, K., 2017. A Complementary Tool to Monitor Fiscal Stress in European Economies (No. 049). Directorate General Economic and Financial Affairs (DG ECFIN). European Commission.
Sy, A.N.R., 2004. Rating the Rating Agencies: Anticipating Currency Crises or Debt Crises? J. Bank. Financ. 28 (11), 2845–2867.
Taffler, R.J., Abassi, B., 1984. Country Risk: A Model for Predicting Debt Servicing Problems in Developing Countries. J. Royal Statist. Soc. Ser. A (General) 147 (4), 541–561.
Wang, T., Zhao, S., Zhu, G., Zheng, H., 2021. A Machine Learning-Based Early Warning System for Systemic Banking Crises. Appl. Econ. 53 (26), 2974–2992.

Further reading

Arslanalp, S., Tsuda, T., 2014. Tracking Global Demand for Advanced Economy Sovereign Debt. IMF Econ. Rev. 62 (3), 430–464.
Baldacci, E., Petrova, I., Belhocine, N., Dobrescu, G., Mazraani, S., 2011. Assessing Fiscal Stress, IMF Working Paper No. 11/100 (International Monetary Fund).
Callier, P., 1985. Further Results on Countries' Debt-servicing Performance: The Relevance of Structural Factors. Rev. World Econ. 121 (1), 105–115.
Catão, L. and B. Sutton. 2002. "Sovereign Defaults: The Role of Volatility", IMF Working Paper No. 02/149 (Washington: International Monetary Fund).
Citron, J.T., Nickelsburg, G., 1987. Country Risk and Political Instability. J. Dev. Econ. 25 (2), 385–392.
Cline, W., 1984. Interest and Debt: Systematic Risk and Policy Response. Institute of International Finance, Washington.
Cruz, C., Keefer, P., Scartascini, C., 2018. Database of Political Institutions 2017 (DPI2017). Inter-American Development Bank. Numbers for Development.
Dawood, M., Horsewood, N., Strobel, F., 2017. Predicting Sovereign Debt Crises: An Early Warning System Approach. J. Financ. Stab. 28, 16–28.
De Cos, Pablo Hernández, Koester, G., Moral-Benito, E., Nickel, C., 2014. Signaling Fiscal Stress in the Euro Area: A Country-Specific Early Warning System. ECB Working Paper No. 1712 Frankfurt: European Central Bank.
Feder, G., Just, R.E., 1977. A Study of Debt Servicing Capacity Applying Logit Analysis. J. Develop. Econ. 4 (1), 25–38.
Frank, C.R., Cline, W.R., 1971. Measurement of Debt Servicing Capacity: An Application of Discriminant Analysis. J. Int. Econ. 1, 327–344.
Georgievska, A., Georgievska, L., Stojanovic, A., Todorovic, N., 2008. Sovereign Rescheduling Probabilities in Emerging Markets: A Comparison with Credit Rating Agencies' Ratings. J. Appl. Stat. 35 (9), 1031–1051.
Ghulam, Y., Derber, J., 2018. Determinants of Sovereign Defaults. Quart. Rev. Econ. Financ. 69, 43–55.
Hajivassiliou, V.A., 1994. A Simulation Estimation Analysis of the External Debt Crises of Developing Countries. J. Appl. Econometrics 9, 109–131.
Hilscher, J., Nosbusch, Y., 2010. Determinants of Sovereign Risk: Macroeconomic Fundamentals and the Pricing of Sovereign Debt. Rev. Financ. 14 (2), 235–262.
International Monetary Fund (IMF), 2017. Review of the Debt Sustainability Framework in Low-income Countries: Proposed Reforms.
Kolscheen, E., 2010. Sovereign Risk: Constitutions Rule. Oxford Econ. Pap. New Ser. 62 (1), 62–85.
Kraay, A., Nehru, V., 2006. When Is External Debt Sustainable? World Bank Econ. Rev. 20 (3), 341–365.
Laeven, L. Valencia, F., 2018. Systemic Banking Crises Revisited. IMF Working Paper No. 18/206 (Washington: International Monetary Fund).
Lane, P.R., Milesi-Ferretti, G.M., 2007. The External Wealth of Nations Mark II. J. Int. Econ. 73, 223–250.
Lane, P.R., Milesi-Ferretti, G.M., 2017. "International Financial Integration in the Aftermath of the Global Financial Crisis." IMF Working Paper 17/115, International Monetary Fund, Washington, DC.
Lloyd-Ellis, H., McKenzie, G.W., Thomas, S.H., 1989. Using Country Balance Sheet Data to Predict Debt Rescheduling. Econ. Lett. 31 (2), 173–177.
Maltritz, D., Molchanov, A., 2014. Country Credit Risk Determinants with Model Uncertainty. Int. Rev. Econ. Financ. 29, 224–234.
McFadden, D., Eckaus, R., Feder, G., Hajivassiliou, V.V., O'Connell, S., 1985. Is There Life After Debt? An Econometric Analysis of the Creditworthiness of Developing Countries. Int. Debt Develop. Countr., 179–209
Messmacher, M., Kruger, M., 2004. Sovereign Debt Defaults and Financing Needs (No. 4–53). International Monetary Fund.
Peter, M. 2002. Estimating Default Probabilities of Emerging Market Sovereigns: A New Look at A Not-So-New Literature. HEI Working Paper No: 06/2002, Geneva: Graduate Institute for International Studies.
Rodriguez, A., Rodriguez, P.N., 2006. Understanding and Predicting Sovereign Debt Rescheduling: A Comparison of the Areas Under Receiver Operating Characteristic Curves. J. Forecast. 25, 459–479.
Sargen, N., 1977. Economic Indicators and Country Risk Appraisal. Economic Review, Federal Reserve Bank of San Francisco, issue Fall, pp. 19–35.
Snider, L.W., 1990. The Political Performance of Third World Governments and the Debt Crisis. Am. Polit. Sci. Rev. 84 (4), 1263–1280.