



WP/20/1

# IMF Working Paper

---

## Debt Is Not Free

by Marialuz Moreno Badia, Paulo Medas, Pranav Gupta, Yuan Xiang

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

## IMF Working Paper

Fiscal Affairs Department

### Debt Is Not Free

Prepared by Marialuz Moreno Badia, Paulo Medas, Pranav Gupta, Yuan Xiang<sup>1</sup>

Authorized for distribution by Catherine Pattillo

January 2020

**IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

#### Abstract

With public debt soaring across the world, a growing concern is whether current debt levels are a harbinger of fiscal crises, thereby restricting the policy space in a downturn. The empirical evidence to date is however inconclusive, and the true cost of debt may be overstated if interest rates remain low. To shed light into this debate, 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 previously ignored in the literature. We find that public debt is the most important predictor of crises, showing strong non-linearities. Moreover, beyond certain debt levels, the likelihood of crises increases sharply regardless of the interest-growth differential. Our analysis also reveals that the interactions of public debt with inflation and external imbalances can be as important as debt levels. These results, while not necessarily implying causality, show governments should be wary of high public debt even when borrowing costs seem low.

JEL Classification Numbers: E43, E62, F34, F37, H63

Keywords: crisis, debt, default, fiscal, machine learning

Author's E-Mail Address: [MMorenobadia@imf.org](mailto:MMorenobadia@imf.org), [PMedas@imf.org](mailto:PMedas@imf.org), [PGupta@imf.org](mailto:PGupta@imf.org), [YXiang@imf.org](mailto:YXiang@imf.org)

---

<sup>1</sup> We are indebted to James Feigenbaum, Catherine Pattillo, Shuyi Liu, Nobuo Yoshida, Kazusa Yoshimura, participants in the 8<sup>th</sup> annual workshop on Fiscal Policies and Institutions organized by the IMF in Brussels, and colleagues at the IMF for useful comments. Juliana Gamboa Arbelaez, Ade Adeyemi, and Kadir Tanyeri provided excellent research assistance and support with coding.

Contents	Page
I. Introduction .....	5
II. Does Debt Matter? Lessons from the Literature .....	7
III. Data .....	9
A. Measuring Fiscal Crises .....	9
B. Predictors.....	11
IV. Methodology .....	13
A. Variable selection.....	14
B. Assessing variable importance .....	16
C. Studying interactions and nonlinearities .....	16
V. Results .....	17
A. Variable selection.....	17
B. Variable importance .....	19
C. An analysis of selected predictors .....	20
VI. Conclusion .....	23
References .....	25
Tables	
Table 1. Fiscal Crises Episodes (1980–2016).....	11
Table 2. Out-of-Sample Performance of Alternative Feature Selection Algorithms.....	18
Figures	
Figure 1. Predictors of Fiscal Crises in the Literature .....	34
Figure 2. Most Common Predictors in the Literature .....	34
Figure 3. Countries with Fiscal Crises, 1980–2016.....	35
Figure 4. Overlap with Other Crises, 1980–2016.....	35
Figure 5. Debt Statistics: Country Coverage, 1980–2016 .....	35
Figure 6. Public and Public External Debt, 1980–2016.....	36
Figure 7. Interest-Growth Differential, 1980–2016.....	36
Figure 8. Feature Selection Algorithms .....	37
Figure 9. Robustness in Variable Selection .....	37
Figure 10. Variable Importance by Group of Predictors .....	37
Figure 11. Contribution to Probability of a Crisis .....	38
Figure 12. Partial Dependence Plots <sup>1/</sup> and Event Studies.....	39
Figure 13. Overall Interaction Strength .....	40
Figure 14 Top-10 Interactions with Public .....	40
Figure 15. Bivariate Partial Dependent Plots: Public External Debt and r-g.....	40
Figure 16. Inflation: Univariate Partial Dependence Plots .....	41

Figure 17. Inflation and Public External Debt: Bivariate Partial Dependence Plots .....	41
Figure 18. Current Account: Univariate Partial Dependence Plots .....	42
Figure 19. Current Account and Public External Debt: Bivariate Partial Dependence Plots ..	42
Figure 20. Credit Gap: Univariate Partial Dependence Plots .....	43
Figure 21. Credit Gap and Public External Debt: Bivariate Partial Dependence Plots .....	43

## Appendixes

1. Literature Review.....	44
2. Fiscal Crisis: Definitions and Data Sources.....	51
3. Sample of Countries.....	52
4. Data: Definition, Sources, and Predictor Groupings .....	53
5. Methodological Details .....	64
Appendix Table 5.1. Out-of-Sample Performance .....	67

*“Economic ruin admits of varied interpretations, but most of them apply at present to the greater part of Europe, and most of the ruin is to be ascribed to the piling up of debt”*

J. S. Nicholson, “Adam Smith on Public Debts,” *The Economic Journal*, 1920

## I. INTRODUCTION

Should governments worry about high public debt? The answer from standard debt sustainability frameworks is yes, for not only does real economic growth dip in the immediate aftermath of a fiscal crisis, but the loss of output is often permanent (Medas et al. 2018; Asonuma et al. 2019). However, despite Reinhart and Rogoff’s (2009 and 2011a) seminal work on the perils of excessive debt, the empirical literature on the relationship between public debt and fiscal crises is still surprisingly inconclusive to this day. The case for more public debt is being reinforced by weak economic activity across the globe, large investment needs, and increasing concerns that monetary policy may be reaching its limits particularly in advanced economies. And yet, the risk of fiscal crises still casts a long shadow. Therefore, as many countries remain riddled with mounting debt, one of the most pressing questions facing policymakers is whether current high debt levels are a bellwether of future crises with large economic costs.

The argument that “public debt may have no fiscal cost” (Blanchard 2019) is also gaining traction as many countries face historically low interest rates and the global stock of negative-yielding debt is hovering around \$12 trillion by the end of 2019. The underlying rationale is that if interest rates are lower than the economic growth rate—that is, the interest-growth differential is negative—there is no reason to maintain a primary surplus as it would be feasible to issue debt without later increasing taxes.

To help shed more light on this debate, 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. The lack of strong evidence 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 modelling difficult.<sup>2</sup> More importantly, it is very hard to distill complex nonlinearities and interactions from the classic econometric techniques typically employed in the early warning literature. Namely, the debt dynamics depend on many factors (e.g. interest rate, growth, deficit, shocks) and some of those interact with each other: 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 increase the risk of a crisis. Therefore, with few settled facts, the debate on whether

---

<sup>2</sup> In the aftermath of the global financial crisis, several scholars compiled new panel datasets on debt covering many decades (if not centuries) of data—e.g. Reinhart and Rogoff (2009) and Abbas et al. (2011) on public debt; and Jordà, Schularick, and Taylor (2016) on private debt. Unfortunately, these datasets tend to include only a few countries or use a narrow and changing definition of debt limiting the scope of research.

public debt matters for predicting fiscal crises remains unresolved. Our paper takes a step toward filling that gap.

We bring evidence to bear on the issue by studying fiscal crises in a broader sample (188 countries) than previously used in the 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, including covariates 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. Our objective, however, goes beyond obtaining a well-performing prediction model as we also want to identify those variables enabling that good prediction. Thus, 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”. 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 the most important group of predictors. This finding is not an exclusive feature of the period immediately predating the global financial crisis but applies more broadly. The narrative, however, is far from simple as some forms of debt are more important than others—in particular, public external debt—and there is strong evidence of non-linearities across all income groups. Remarkably, the interest-growth differential has low predictive value. What is more, beyond certain debt levels the likelihood of fiscal crises increases significantly irrespective of whether the interest-growth differential is highly positive or negative. Event studies give some insights as to why this may be the case: it is only at the onset of the crisis that the interest-growth differential spikes, making it immaterial for signaling purposes. Our empirical analysis also reveals that it is not solely public debt that matters as its interactions with other variables are as important. Hence, the probability of a crisis rises steeply at high public debt levels but also at relatively moderate ones when accompanied by high inflation or large current account deficits. Notably, these results hold across all income groups.

Our work is related to the extensive early warning literature on sovereign debt crises (see, for example, Detragiache and Spilimbergo 2001; Manasse, Roubini, and Schimmelpfennig 2003; and Chakrabarti and Zeaiter 2014). Relative to that literature, our contribution is twofold. First, we analyze systematically the predictive importance of public debt thanks to a more comprehensive dataset on debt and its characteristics (such as the creditor structure) than previous studies and an unparalleled number of macro indicators. To the best of our knowledge, this is also the first study to examine explicitly the predictive power of the interest-growth differential. The use of a much broader sample enables us to explore the sensitivity of our results along the time and country grouping dimensions. As an additional layer ensuring robustness, we assess the performance of alternative algorithms not only in

terms of the out-of-sample predictive accuracy but also the stability of the variable selection with respect to sampling variation. Our second contribution is to leverage machine learning to analyze complex non-linearities and interactions previously ignored in the literature. Our results suggest that this omission may go a long way in explaining the lack of conclusive empirical evidence on the importance of public debt of past studies. Our paper is the first attempt to shed 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 II provides a survey of the literature on the determinants of fiscal crises covering the last five decades. We not only describe past approaches to crisis prediction but also discuss the changing nature of the predictors identified as important since the 1970s. Section III discusses the definition of fiscal crises and the main characteristics of the data with an emphasis on debt. In Section IV, we describe the main features of our empirical strategy. Section V presents the results on selection of variables and predictor importance. We then conclude by recapping our main findings and their implications for the current policy debate.

## II. DOES DEBT MATTER? LESSONS FROM THE LITERATURE

The literature on fiscal crises and their determinants has evolved significantly over time. This has been in part a reflection of the changing nature of sovereign debt defaults and other forms of fiscal distress. As new data and econometric techniques became available, economists tried to uncover the empirical regularities surrounding fiscal crises with a special emphasis on the the second financial era (i.e. the period after World War II).

Initially, the reasearch on sovereign crises focused mainly on developing countries. All in all, the 1950s–1960s was a period when greater indebttness was seen as a way to help promote economic growth among less developed nations. By the 1970s, borrowing started to be associated not only with development needs but also with periods of external current account imbalances. Nonetheless, debt was generally seen in a positive light up to that point (Solberg 1988). By the early 1980s, however, the number of fiscal crises surged and so did the work on understanding the drivers of debt distress. We next present an overview of the literature over the last 50 years and some of the challenges that remain especially on identifying the key determinants of crises (Figures 1 and 2). We survey 42 papers chosen out of a pool of 63 references based on their empirical relevance and whether they clearly identify key predictors of crises (for details, see Appendix 1). We organize the discussion into three different periods referring to the years when the papers in question were published.

### *1970s–1990s: capacity to repay*

The literature during this period was mainly 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 selection of predictors was often dictated

by the financing mix of countries covered (many of them developing countries). Not surprisingly, these papers found external variables related to the capacity of a country to repay its obligations (external debt service; size of external debt<sup>3</sup>, foreign exchange reserves to imports) to be among most important predictors of crises although economic growth was also identified as a key indicator (Figure 1, panel 1).

The definition of crisis in these studies was in general limited to debt rescheduling or arrears on external debt. The empirical strategy was often based on a logit model with a small number of predictors or, in a few cases, a linear regression model. In some instances, the studies were trying to identify determinants (sometimes, institutional factors) of the capacity to repay and did not test whether the models were useful in predicting future crises (Rivoli and Brewer 1997; Lee 1991; Berg and Sachs 1988). There was also little attention paid to studying the role of public debt in helping predict crises as the focus was on external imbalances.

### ***2000–10: the importance of external debt***

At the turn of the century, the definition of crisis starts to expand to include not only debt defaults but also access to IMF programs above a certain quota. The concern was that the previous definition was too restrictive as countries in distress might have been able to avoid defaults by getting official credit from the IMF. The logit model remained one of the most popular tools to predict crises and identify drivers (Ciarlone and Trebeschi 2005; Detragiache and Spilimbergo 2001). However, there were a few papers trying different approaches, including Manasse and Roubini (2009), with a classification and regression tree, and Fioramanti (2008) with neural networks.

Nonetheless, fiscal crises continued to be seen from the perspective of external vulnerabilities, that is whether the sovereign would default on external creditors. As such, there was significant attention to external debt, with a large share of papers identifying it as a predictor of crises (Figure 1, panel 2). Other common predictors included real GDP growth, debt service and the maturity of debt, exchange rate, and default history. In this context, there was still limited focus on public debt, especially domestic debt and arrears.

### ***2011 onwards: a growing debate on the role of public debt***

In the aftermath of the global financial crisis, several papers start to use a more comprehensive definition of fiscal crises. The attention is no longer on external defaults alone, but papers also acknowledge that fiscal crises may reflect other types of distress and affect both external and domestic creditors. Hence, the crisis definitions now include debt

---

<sup>3</sup> It is important to note that the papers over this period are often not explicit about the definition of external debt and whether it includes the external liabilities of the private and public sectors.



defaults (mainly external), IMF programs, implicit debt defaults (high inflation, domestic arrears), and loss of market access (Medas et al. 2018; Bruns and Poghosyan 2018; and Sumner and Berti 2017). A broader set of methodologies is also used, although the logit model remains a popular approach. Several papers also pay closer attention to the robustness of results, especially the out-of-sample predictive power, but this remains a weakness in the literature of early warning systems more generally (Berg, Borensztein, and Pattillo 2005; and Cerovic et al. 2018).

Although there is an effort to examine a broader set of potential predictors, the empirical research remains constrained by the use of traditional econometric techniques. In many cases, preference has been given to parsimonious approaches relying on a limited set of indicators partly reflecting the priors of the researcher but also 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, and to a less degree external debt and degree of openness). Public debt and fiscal-related variables are also identified but less frequently so (Figure 1, panel 3).

While there are a growing number of papers examining the role of public debt, the evidence so far 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) and Sumner and Berti (2017) find some evidence that it does but it is not robust across specifications; changes in public debt are also a significant predictor of debt crises in Reinhart and Rogoff (2011a) although the result does not hold for the post-World War II period. In addition, there have been no studies explicitly analyzing the interest-growth differential as a predictor of fiscal crisis. The lack of conclusive evidence on the importance of public debt in predicting fiscal crises has likely fuel the recent policy debate on whether governments should worry about public debt levels especially when interest rates are low.

By in large, the literature has also been unable to explore complex dynamics. Despite some recent research using machine learning techniques for predicting sovereign debt crises (Savona and Vezzoli 2015), these studies only consider a small set of indicators and stop short of analyzing non-linearities or interactions among predictors.

### **III. DATA**

#### **A. 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 (see, for example, 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, Roubini, and Schimmelpfennig 2003). To capture these different facets, we follow Medas et al (2018) and

identify fiscal crises in any given year if any of the following four criteria is met (for details, see Appendix 2):

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.

Based on this definition, we identify 418 crisis episodes for a sample of 188 countries over the period 1980–2016, making ours one of the most comprehensive study of fiscal crises to date (for details on the sample, see Appendix 3). On average, countries have experienced two fiscal crises since 1980 with more than three quarters of countries having at least one crisis (Table 1). Low-income countries (LICs) 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 (Figure 3). To a lesser extent, there was also some bunching in the early 1980s—reflecting the collapse of commodity prices and a surge in global interest rates—and in 2010—following the onset of the global financial crisis—pointing to the potential importance of 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 (Figure 4). Consistent with the empirical evidence in Reinhart (2002), we find that most of these cases relate to EMs and LICs, underscoring the importance of external financing among these countries, an issue we explore later in this paper. The synchronicity with financial crises is, however, relatively low suggesting that, although banking crises may precede or coincide with sovereign debt crises through a contingent liability channel (Reinhart and Rogoff 2011a), the root cause of a fiscal

crisis may often lie elsewhere. Triple crises are even rarer, accounting for only 3 percent of events.

Table 1. Fiscal Crises Episodes (1980–2016)

	<b>Total</b>	<b>AEs</b>	<b>EMs</b>	<b>LICs</b>
Number of crises starts	<b>418</b>	25	202	191
of which <sup>1/</sup> (percent):				
Credit event	<b>72.7</b>	20.0	68.3	84.3
Exceptionally large official financing	<b>33.7</b>	56.0	34.7	29.8
Implicit domestic default	<b>9.8</b>	24.0	11.4	6.3
Loss of market confidence	<b>25.1</b>	84.0	32.2	9.9
Average per country	<b>2.2</b>	0.7	2.1	3.2
Number of countries with no fiscal crisis	<b>36</b>	19	15	2
Average duration (years)	<b>5.2</b>	3.2	5.3	5.3

Sources: Bloomberg; Datastream; Eurostat; Gelos, Sahay, and Sandleris (2004); Guscina, Sheheryar, and Papaioannou (2017); IMF, International Financial Statistics; OECD; Reuters; and authors' calculations.

1/ 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.

## B. Predictors

As discussed in Section II, 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, priors prevailing at the time of previous studies (as well as data availability) may have biased the selection of indicators to be tested, narrowing the set to a few variables of interest to the researcher in question and leaving out important interactions (Chakrabarti and Zeaiter 2014). To address these shortcomings, we canvass the empirical and theoretical literature to identify potential predictors of crises. Our aim is to cast as wide net as possible, the only constraint being data availability.<sup>4</sup> As a result, our dataset covers an unusually rich array of economic indicators and institutional country characteristics. Furthermore, the analysis uses several permutations of each variable—such as levels, lags, and differences at various horizons—and cross-sectional averages—which allow us 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. Appendix 4 gives a detailed description of the variables and sources.

<sup>4</sup> A variable is included if 70 percent of the data exists. To take advantage of much information as possible we impute missing values for a given variable with the training sample median of the non-missing values.

An important contribution of this paper is to assemble a more comprehensive range of debt metrics than previous studies, allowing us to make a more robust assessment of their relevance as predictors of crises. We do not restrict ourselves to public debt but also include various indicators of private indebtedness as well as the total stock of debt in the economy. Accounting for these different forms of debt is important as the line between them may become blurry at times of crises (Mbaye, Moreno Badia, and Chae 2018a) and what may matter is not any single measure but their interactions. To construct consistent time series of debt, we leverage the Global Debt Database—an unmatched account of private, public, and total debt for 190 countries going as far back as 1950 (for details, see Mbaye, Moreno Badia, and Chae 2018b). We scale debt not just by GDP but also use other indicators that can proxy for available liquidity such as reserves or revenues, complementing the information provided by debt service ratios. We also capture some of the characteristics of public debt that have been identified as important in the sovereign debt crises literature. Significant effort was placed on building metrics of external public debt comparing alternative sources of data to ensure the consistency of the series. Figure 5 shows the debt data coverage in our sample. A salient feature is that there are large differences in public debt profiles across country groups not just in terms of levels but also the composition—external debt being the main component among LICs but not in other income groups (Figure 6). In the remainder of the paper we will exploit this heterogeneity to identify what debt characteristics may have higher discriminating value.

Given the current policy debate, we also pay special attention to the construction of the interest-growth differential variable (henceforth, “ $r-g$ ”). Our starting point is the recursive equation behind changes in the debt-to-GDP ratio—see Escolano (2010) for a detailed discussion—and thus,  $r-g$  is defined as:

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

where  $r$  is the effective interest rate and  $g$  the 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.<sup>5</sup> The trade-off is that, to the extent that the interest bill refers to a broader perimeter of government than the stock of debt, the interest-growth differential may be over-estimated. This shortcoming also applies to other studies that look at long histories. However, LICs—the group for which debt stocks usually refers to the narrower perimeter of

---

<sup>5</sup> 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  $t-1$ ). This interest rate is different from the measure used to define fiscal crises which is based on marginal yields and spreads (see Appendix 2). For countries that issue foreign-currency denominated debt, it may be important to account for the depreciation-adjusted interest-growth differentials (Escolano, Shabunina and Woo 2017). However, as in other studies, limited data availability on the currency composition of debt prevents us from making this adjustment. Nonetheless, among the set of predictors we also include various measures of exchange rate depreciation.

government—typically report the interest bill for the same level of government so this is less of a problem. Consistent with other studies (see, Mauro et al. 2015; Barrett 2018; Escolano, Shabunina, and Woo 2017), our data suggests that, on average, the interest-growth differential has been close to zero or negative since the 1980s across all income groups (Figure 7). However, there is a wide dispersion within each group and positive interest-growth differentials are not an anomaly.

Finally, we also compile data for various fiscal indicators that have often been overlooked in the literature. We include not only the fiscal deficit but also revenues and expenditures (and the disaggregation into its primary component). To capture potential valuation effects and contingent liabilities that may lead to fiscal distress, we calculate stock-flow adjustments ( $SFA_t$ ) as follows:

$$SFA_t = d_t - \left( \frac{1+r}{1+g} \right) d_{t-1} + p_t$$

where  $d_t$  and  $d_{t-1}$  are the stock of public debt in periods  $t$  and  $t-1$  respectively, and  $p_t$  is the primary balance in period  $t$ .

#### IV. 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 our estimation strategy captures complex dynamics. Our model of choice is a random forest (Breiman 2001) for two reasons. First, it can deal with complexity and deliver significant improvements in predicting fiscal crises relative to standard econometric approaches typically used in the early warning literature and other machine learning algorithms (see Appendix 5 for an empirical comparison of the out-of-sample performance of the random forest (RF) against other econometric approaches). Second, in a large scale empirical evaluation of 179 classification algorithms tested across 121 real world datasets, Fernandez-Delgado et al. (2014) find that, on average, RF is the best performer.

RF aggregates many decision trees, each run in a random sample of variables and country-years. Decision trees are very flexible models but the bigger a tree grows, the less likely is that it will generalize well to out-of-sample data. This is usually referred to as overfitting. 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 to this approach 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, Seifert and Szymczak 2019). Our empirical approach is to reduce our initially large set of variables to the ones that contain more information than noise. In the remainder

of this section, we describe the various selection procedures used in this paper to identify a stable set of predictors and the statistical techniques to analyze the importance of variables and their interactions. Appendix 5 gives more technical details on methodology.

### A. Variable selection

Variable selection is a crucial issue in many applied classification and regression problems (e.g. Hastie, Tibshirani, and Friedman 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:

$$\hat{y} = f(\mathbf{X}),$$

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 and 0 if there is no crisis in the next two years.

To reduce the number of variables to those that are most relevant, we use what is referred to in the machine learning literature as “feature selection” algorithms. These techniques have been used as the workhorse in genomics research (for a review, see Saeys, Inza and Larranaga 2007; Ma and Huang 2008; Hilario and Kalousis 2008; Duval and Hao 2010; and Degenhardt, Seifert and Szymczak 2017). The basic principle of these algorithms is to identify the relative importance of features (each of them using different criteria to determine the relative ranking) and eliminate those features that are unimportant according to some pre-defined metric. We focus on the four algorithms built around RFs that have been more widely used in the literature:

- ***P-values computed with permutation importance (PIMP)***. Altmann et al. (2010) developed a method for selecting relevant predictors based on repeated permutations of the outcome vector (i.e. the likelihood of a crisis), leaving correlation patterns between predictors unchanged. For each permutation of the outcome, the relevance for all predictor variables is assessed. This leads to a vector of importance measures for every variable, called the “null importances”. The PIMP algorithm fits a probability distribution to the population of null importances (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 (with respect to the PIMP scores) are kept.
- ***Recursive Feature Elimination (RFE)***. RFE aims to find a minimal set of variables which leads to a good prediction model (Diaz Uriarte and 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.
- ***VSURF***. Developed by Genuer et al. (2015), this algorithm returns two subsets of variables. The first is a subset of important variables including some redundancy which can be relevant for interpretation, and the second one is a smaller subset corresponding to a model trying to avoid redundancy focusing more closely on the prediction objective. The two-stage strategy is based on a preliminary ranking of the explanatory variables using the RF permutation-based score of importance and proceeds using a stepwise forward strategy for variable introduction.

In choosing among these algorithms, we consider two criteria:

- ***High predictive power***. Ideally, we want the empirical power of the smaller variable set to be at least as good as the the full set. To make that evaluation, we compare the out-of sample predictive performance of four RF estimated using the features selected by each algorithm against the RF estimated with the full set of variables. We check the statistical significance of the difference between the performance of each algorithm and the full RF model by calculating t-tests based on standard errors adjusted for two-way clustering (see Cameron, Gelbach, and Miller 2011). The main performance measure for these comparisons is the area under the receiver-operator curve (AUROC), although other measures such as of the log likelihood and mean squared errors (MSE) are also reported. Using the AUROC, one of the most common metrics in the early warning literature, 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.
- ***Stability of feature selection***. As noted by Degenhardt, Seifert and Szymczak (2019), it is important to verify the stability of variables selected by each algorithm as these can vary due to small changes in the data and therefore results may not be reliable. To assess stability, we construct two separate samples by randomly dropping 5 percent of

observations, comparing the overlap of the features 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 (see, Nogueira and Brown 2016). The Pearson coefficient takes values between -1 and 1, with 1 meaning perfect overlap between the two sets.

## B. 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 feature is “important” if shuffling its values increases the model error, because in this case the model relied on that feature for the prediction. At the opposite end, a feature is “unimportant” if shuffling its values leaves the model error unchanged, because in that case the model ignored that feature. The model errors (with and without shuffling) are calculated on the oob sample.<sup>6</sup>
- *Shapley values.* Variables are 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 also calculate the differences in Shapley values between crisis and non-crisis events. Note that Shapley values do not give the difference of the predicted value after removing the feature from the trained model, but rather give the contribution of a feature value to the difference between the actual prediction and the mean prediction of the sample.

## C. Studying interactions and nonlinearities

### *Partial dependence plots*

As outlined above, machine learning techniques allow us to capture non-linearities and heterogenous interactions between various predictors. To analyze these complex relationships, we rely on partial dependence plots (Greenwell 2017; and Friedman and Popescu 2008; Friedman 2001). A partial dependence plot (PDP) shows the marginal effect of one or several features on the predicted outcome and can identify if the relationship

---

<sup>6</sup> 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. This is the so-called oob sample.



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 relationship between a feature and the predicted outcome (in our case the probability of a crisis), whereas, a bivariate PDP helps to visualize predicted outcome for a pair of features by marginalizing over other variables (see Appendix 5 for details). Intuitively, the partial dependence function at a particular feature value (for example, 40 percent of public external debt) represents the average prediction (in our case, the probability of a fiscal crisis) if we force all data points to assume that feature value. *H-statistics*

We also use a measure of relative strengths of two-way interaction, that is to what extent two features interact with each other in a given model, following Friedman and Popescu (2008). The measure of relative strength of interaction (termed H-statistic) uses the concept of PDPs and is defined as the variance of the difference between observed bivariate PDP and the two individual PDPs. This captures the fraction of variance of  $PD_{xj}$  not captured by variance of  $PD_x + PD_j$  over the data distribution. This helps to rank all (n-1) pairs of  $n^{\text{th}}$  variables based on their relative strength of interactions between each other.

Mathematically, the H-statistic for the interaction between feature j and k is:

$$H_{jk}^2 = \sum_{i=1}^n [PD_{jk}(x_j^{(i)}, x_k^{(i)}) - PD_j(x_j^{(i)}) - PD_k(x_k^{(i)})]^2 / \sum_{i=1}^n PD_{jk}(x_j^{(i)}, x_k^{(i)})^2$$

The amount of the variance explained by the interaction is used as interaction strength of H-statistic. The statistic is 0 if there is no interaction at all and 1 if the variance of the  $PD_{jk}$  is explained by the sum of the partial dependence functions. An interaction statistic of 1 between two features means that each single  $PD$  function is constant and the effect on the prediction only comes through the interaction.

## V. RESULTS

### A. Variable selection

There are wide-ranging differences in the variable sets selected by each algorithm—from less than 10 variables in the VSURF to more than 300 in Boruta (Figure 8)—underscoring the different objectives of each of them 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 2). Although we pool all countries for estimation purposes, results are disaggregated by income groups for comparison with previous studies. We find that the full model performs better for AEs and EMs than for LICs but, overall, the predictive power is higher than previous studies. By way of comparison, the AUC in the full model is 0.81 for AEs and EMEs and 0.71 for LICs while Cerovic et. al (2018) report a maximum AUC of 0.69 and 0.68 respectively (see also Appendix 5 for the

performance of alternative estimation methods).<sup>7</sup> Among the feature selection algorithms, Boruta is always at least as good as the full model for both income groups irrespective of the performance metric considered. Moreover, it is statistically significantly better than the full model across the board when assessed by the loglikelihood and for AEs and EMEs when assessed by the MSE. At the other extreme, the VSURF has the worst performance throughout, while the PIMP is usually worse than either the Boruta or the RFE. The RFE, one of the most popular methods for variable selection, is somewhere in between, underperforming relative to Boruta across some metrics.

Table 2. Out-of-Sample Performance of Alternative Feature Selection Algorithms

	<b>Boruta</b>	<b>RFE</b>	<b>VSURF</b>	<b>PIMP</b>	<b>RF</b>
<b><i>Advanced and emerging market economies</i></b>					
AUROC	0.805 (0.025)	0.793* (0.026)	0.734*** (0.031)	0.791** (0.026)	0.806 (0.026)
Log(likelihood)	-0.289*** (0.027)	-0.286** (0.029)	-0.34 (0.043)	-0.29** (0.028)	-0.299 (0.026)
MSE	0.084** (0.011)	0.085 (0.011)	0.096*** (0.011)	0.085 (0.011)	0.087 (0.01)
<b><i>Low-income countries</i></b>					
AUROC	0.708 (0.042)	0.712 (0.042)	0.688 (0.039)	0.707 (0.042)	0.706 (0.043)
Log(likelihood)	-0.547* (0.043)	-0.544** (0.044)	-0.601* (0.052)	-0.548* (0.043)	-0.556 (0.039)
MSE	0.186 (0.018)	0.185 (0.018)	0.207** (0.019)	0.187 (0.018)	0.189 (0.017)

Note: Bootstrapped standard deviations (based on 100 random resamples of the test sample with replacement) in parentheses. Stars indicate degree of confidence that a model outperforms the Random Forest estimated with the full set of 748 variables: \* 90%, \*\*=95%, and \*\*\*=99%. Models predict probability of crisis start occurring in year t+1 or t+2. Out-of-sample performance obtained from 15 rolling regressions (see Appendix 5).

In choosing among alternative feature selection algorithms, we also want to ensure the stability in the choice of predictors as results can be highly sensitive to small perturbations in the sample (Nogueira and Brown 2016). Given computational time, we restrict our comparison of stability to the best performing algorithms: Boruta, RFE, and PIMP. The Pearson index is 0.92 for Boruta indicating that despite changing the sample, there is high overlap of the selected variables between replicates (Figure 9). In contrast, the Pearson index is 0.63 for RFE and 0.81 for PIMP suggesting lower stability in the selection of variables, consistent with the findings of other studies for these algorithms (see, Degenhardt, Seifert, and Szymczak 2019).

<sup>7</sup> 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 low-income countries in detail covering the period 1970–2015.

## B. Variable importance

Given its predictive power and stability, we choose Boruta as the benchmark algorithm to study the relative importance of predictors. Despite reducing the initial set by half, the number of variables selected by Boruta is still very large (336 indicators, including permutations). Hence, we expect the predictive power of any individual indicator to be small. Moreover, it may be difficult to fully distinguish the impact of closely related (and correlated) variables. Therefore, we group them into 23 categories to make it easier the interpretation of results (see Appendix 4 for the groupings). Based on the out-of-bag permuted predictor importance, we find that:

- Public debt is the most important group of predictors followed closely by public debt service (Figure 10). This should not be surprising as fiscal crises by and large involve some degree of debt distress (e.g. inability to repay or borrowing difficulties). But as discussed in Section II, the previous literature has only found weak evidence that public debt matters. We conjecture that, by including 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. We will explore some of these below.
- Institutional slow-moving variables also rank among the most important predictors. These include the level of development (GDP per capita), demographics, and to a lesser degree the quality of institutions. Given the very different frequency of crises between AEs, EMs and LICs, it is likely that these variables are helping discriminate countries more prone to crises rather than the exact timing. In essence, this may reflect the lower vulnerability of more developed countries as stronger institutional frameworks may better prepare them to manage the exposure to shocks and avoid crises.
- In line with the literature, external variables—in particular, external capital flows and, to a less extent, external debt, current account, and the exchange rate—are also important. As discussed in Section III, fiscal crises overlap with currency crises in a third of cases and, thus, we can expect some association between fiscal crises and periods when external borrowing conditions change, or external investors become concerned with the ability of the sovereign to fulfill its debt obligations.
- Fiscal flow variables (deficits, revenues, spending) also appear as relevant, but considerably less than debt or some of the external variables. This explains why the past literature had difficulty in finding fiscal variables as robust predictors.
- At the other end, the interest-growth differential and global conditions are among the least relevant variables.

As an alternative to measure the discriminating power of predictors between crisis and non-crisis observations, we look at Shapley differences. Overall, we find that the ranking of variables across income groups are broadly similar but with some differences (Figure 11). Public debt and public debt service are the most important group of predictors for EMs and LICs but somewhat less for AEs (although public debt remains among the top 3 categories). On the other hand, private debt appears as a more important predictor than public debt for AEs possibly suggesting that 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 bail out of banks or indirectly through the ensuing recession). We also look at Shapley differences for the pre- and post-2000 period. Overall, public and public debt service are the top categories in both time periods although in reverse order (i.e. public debt has the highest ranking in the post-2000), suggesting that public debt may have been an important red flag not only in the run up to the global financial crisis.

### C. An analysis of selected predictors

In what follows, we undertake a more in-depth analysis of key leading indicators. Given the importance of public debt, we zero in on this group of predictors and, specifically, on public external debt which is the individual debt measure with the highest predictive value. We are also interested in the interactions with other indicators to help inform the ongoing policy debate on the risks associated with debt in the face of low interest rates and inflation as well as the potential role of external and financial imbalances. Throughout this analysis, we will make extensive use of univariate and bivariate PDPs. We present results disaggregated by income groups (AEs, EMs, and LICs) as the nature of crises may differ widely depending on the level of development.

#### *Public debt*

Figure 12 (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 relationship between the two that is non-linear in nature and holds across all income groups although with some differences in profiles.<sup>8</sup> For AEs, the probability of a crisis increases substantially once debt is around 70 percent of GDP. For EMs, the estimated probability is relatively flat for debt levels below 30 percent of GDP but rises steeply above those levels. For LICs, predicted probabilities are much higher from the start and the steepening of the curve takes place at lower debt levels than in other income groups.

---

<sup>8</sup> One caveat of the PDP is that it relies on the assumption of independence among features. Therefore, the results may be biased if features 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.

To get further insights of when is a probability high enough to be concerned, we calculate the probability thresholds at which the model identifies a crisis. This is done by calculating the threshold that minimize the sum of type I and type II errors (missed crises and false alarms) following the literature (see, for example, Berg, Borensztein, and Pattillo 2005). The threshold for AEs is 8.5 percent, which results in capturing 80 percent of the crises while false alarms are kept at 20 percent. For EMs, the probability threshold is 22 percent and for LICs 28 percent. Overall, public external debt is one of the few variables for which at high enough values (70 percent of GDP for AEs and EMs, and 80 percent for LICs), the estimated probability breaches the crisis threshold regardless of other factors. The somewhat higher debt level for LICs, which may seem counterintuitive given the findings of Reinhart, Rogoff, and Savastano (2003) on debt intolerance, is likely to reflect several factors. First, the minimization of the sum of errors results is picking a higher debt level for LICs likely to avoid too many false alarms. Second, it may partly reflect the higher share of concessional borrowing among LICs making the debt burden (for a given level of debt) lower than those cases with commercial debt.

Taken together, these results tell a consistent story on the high relevance of public debt as a leading indicator. But its importance is also related to interaction effects. By calculating H-statistics scores for pairwise interactions between all variables, we can disentangle to what extent a feature interacts cumulatively with all other features in the model. Estimates H-statistics scores show that the public external debt has the strongest cumulative interactions with all other variables (Figure 13). Also, the H-statistic ranks as top variables other debt and debt service indicators consistent with the analysis on variable importance. Looking into the 2-way discrete interactions, we find that the features that most interact with public external debt are external assets, the inflation rate, GDP relative to US, and public external amortization (Figure 14), suggesting that the estimated probability of a crises is higher than the individual contributions of these variables. Therefore, it is entirely possible that the probability of a crises may be high even for moderate levels of public external debt if other factors are present. We will look at some of these interactions below.

### *The interest-growth differential*

Academics, investors, and policymakers alike are still getting to grips with negative interest rates and what they mean for the affordability of public debt (Blanchard 2019, Garin et al. 2019, and Mehrotra 2017). More specifically, the question is not about interest rates but how high they are relative to economic growth (the interest-growth differential).<sup>9</sup> Given our findings so far, a question is whether a low-interest growth differential should assuage concerns about high debt levels. At first glance, the variable importance analysis suggests that the interest-growth differential has very limited information. This is confirmed by PDPs:

---

<sup>9</sup> A country with high economic growth (and tax buoyancy) will be in a better condition to manage debt than a country with lower growth rate for the same interest rate. Nonetheless, a negative interest-growth differential does not ensure that debt will not increase as this will also be a function of the primary deficit.

even for large variations of the interest growth differential, the estimated probability of a crisis barely changes and remains relatively low and never breaches the crisis threshold (Figure 12, panel 3). To see what insights can be gleaned from the data that could explain this apparent puzzle, we follow Gourinchas and Obstfeld (2012) to analyze their dynamics in the run-up to a crisis. Overall, our event study shows that interest-growth differentials can remain low for long stretches of time and only shoot up at the onset of the crisis, thereby making it irrelevant as a leading indicator (Figure 12, panel 4).<sup>10</sup>

Bivariate PDPs also show that a low interest-growth differential does not dampen the risks of high debt (Figure 15). Cells highlighted in red depicts combination of public external debt and the interest-growth differential for which the estimated probability of a crisis is above the 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. The probability of a crisis for a given level of debt does not change with the interest growth differential for these countries. On the other hand, we observe some interactions for LICs. In particular, both highly negative and positive interest-growth differentials can imply a higher probability of crisis for the same level of debt. A possible reason is that they may both signal imbalances although of a different nature (e.g. a low interest-growth differential could be due to overheating). Another possible explanation is that governments may respond to periods of low interest-growth differential by increasing deficits and accumulating debt, negating the potential benefits of low borrowing costs for reducing risks. Therefore, our findings should not be interpreted as suggesting that the interest-growth differential does not matter for debt sustainability, but it 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 on the heels of the global financial crisis, the problem is just the opposite: inflation is too low. So what does it mean for the risk of fiscal distress? The analysis based on univariate PDPs suggest that there is a strong relationship between inflation and the estimated probability of crises, consistent with the findings in Reinhart and Rogoff (2011a). As with debt, we also find evidence of strong non-linearities: for AEs and EMs, the probability of crises increases significantly when inflation is above 20 percent (Figure 16).<sup>11</sup> Although the literature has long established that countries with hyperinflation tend to suffer from debt distress, ours is the first paper presenting evidence of the non-linear relationship.

---

<sup>10</sup> 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, Bornstein, DAVIS 2019). Mauro and Zhou (2019) also use an event study to argue that sovereign defaults may not necessarily be preceded by high (positive) interest-growth differentials.

<sup>11</sup> Note that the threshold for inflation in our crisis definition is higher at 35 percent for AEs and 100 percent for EMs and LICs.

Our results also suggest that both an increase and decline in inflation can be associated with a higher probability of crises, particularly for EMs, underscoring the risk of deflation and the snowball effect it can have on debt dynamics (Crafts 2016). Notably for LICs, the levels of public external debt for which the estimated probabilities breach the crisis thresholds decrease with inflation (Figure 17). This means that even for relatively low levels of debt, the probability of a crisis in LICs surges when inflation is high. This could reflect the fact that in some countries the ability to manage even relatively moderate levels of debt is limited and, in such cases, governments may resort to monetizing deficits.

### *External and financial imbalances*

As discussed in section III, there is some overlap between fiscal and currency crises. The analysis on variable importance also confirms the relevance of external factors. To further explore the importance of external imbalances as driver of crises, we look at the PDP of the current account balance. As with other indicators, we find what is by now a recurrent non-linear pattern: once external deficits are between 3–7 percent of GDP, the probability of a crisis increases substantially for AEs and EMs (Figure 18). Surprisingly, the current account deficits seem to be less relevant for predicting fiscal crises in LICs. There is also some evidence of interactions between public external debt and the current account particularly for AEs (Figure 19). Even for relatively moderate levels of debt, the probability of crises rises steeply when current account deficits are high. The opposite is however 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 one of the most popular metrics in the literature, 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 (Figure 20). We also see interactions with public external debt for EMs, with the estimated probability breaching the crisis thresholds for lower levels of debt if the credit gap is large (Figure 21). At the other end, the private debt dynamics are much less relevant for LICs likely reflecting low financial deepening and, therefore, relatively low risks compared to AEs or EMs.

## VI. CONCLUSION

This paper contributes to the debate on the costs of public debt by revisiting its importance in predicting fiscal crises. In a world of ultra-low interest rates, it is tempting to believe that there may be no costs. For those that subscribe to that theory, the natural conclusion is that now may be the time to rely more heavily on debt to attend to worthy causes such as fixing a crumbling infrastructure all while propping up a frail economy. The skeptics point to history, noting that those that ignore high debt do it at their peril as excessive debt may force disruptive fiscal adjustments or eventually lead to costly crises.

We use machine learning models to confront these dueling views with evidence. Our results show that public debt in its various forms is the most important predictor of fiscal crises and it does matter always and everywhere. But public debt is not the only game in town as its interactions with other predictors also make a difference. Surprisingly, however, the interest-growth differential does not have much signaling value: it does not really matter whether it is highly positive or negative; moreover, beyond certain debt levels, the likelihood of a crisis surges regardless.

It is important to acknowledge that 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 (see, Athey 2018). What we can confidently say is that there is a high correlation between public debt and crises and that this association is very robust. Therefore, at the current juncture, complacency about high debt levels would be ill-advised even if interest-growth differentials were to remain low. The underlying reason is that the dynamics of crises are highly non-linear and by the time the interest-growth differential may start flashing red, a crisis may well be underway catching policymakers off guard.

These findings do not mean that bringing debt down is always the right policy prescription. There are clearly cases where the use of debt for countercyclical purposes, to increase public investment, or to address other structural needs is desirable. However, the evidence presented in this paper points to the risks, suggesting that public debt might not be free after all.



## REFERENCES

- Abbas, S. A., N. Belhocine, A. ElGanainy, and M. Horton. 2011. "A Historical Public Debt Database," *IMF Economic Review*, vol. 59, issue 4, pp. 717–42.
- Altmann A., L. Tolosi, O. Sander and T. Lengauer. 2010. "Permutation Importance: A Corrected Feature Importance Measure," *Bioinformatics Volume* 26 (10), pp. 1340–1347.
- Apley, D. W. 2016. "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models." arXiv preprint arXiv:1612.08468 (2016).
- Arslanalp, S., and T. Tsuda. 2014. "Tracking Global Demand for advanced Economy Sovereign Debt," *IMF Economic Review*, Vol. 62, Iss. 3, pp. 430–64.
- 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).
- Athey, S. 2018. "The Impact of Machine Learning on Economics," in *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, University Chicago Press.
- Baldacci, E., I. Petrova, N. Belhocine, G. Dobrescu, and S. Mazraani. 2011. "Assessing Fiscal Stress," IMF Working Paper No. 11/100 (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., and M. Fratzscher. 2013. "The Pricing of Sovereign Risk and Contagion during the European Sovereign Debt Crisis," *Journal of International Money and Finance*, Vol. 34, pp. 60–82.
- Berg, A. and J. Sachs. 1988. "The Debt Crisis Structural Explanations of Country Performance," *Journal of Development Economics*, 29(3), pp. 271–306.
- Berg, A., E. Borensztein and C. Pattillo. 2005. "Assessing Early Warning Systems: How Have They Worked in Practice?" *IMF Staff Papers*, 52(3), pp. 462–502.
- Berg A. and C. Pattillo. 1999. "Predicting Currency Crises: The Indicators Approach and An Alternative," *Journal of International Money and Finance*, 18, pp. 561–86.
- Berti, K., M. Salto, M. Lequien. 2013. "An Early-detection Index of Fiscal Stress for EU Countries," European Economy. *Economic Papers*. 475. Brussels.

- Breiman, L., 2001. Random Forest. *Machine Learning*, 45(1), 5–32.  
<http://oz.berkeley.edu/users/breiman/randomforest2001.pdf>.
- Blanchard, O. 2019. “Public Debt and Low Interest Rates,” *American Economic Review* 109(4), pp. 1197–1229.
- Bluwstein, K., M. Buckman, A. Joseph, M. Kang, S. Kapadia, and O. Şimşek. 2019. “Credit Growth, the Yield Curve, and Financial Crisis Prediction: Evidence from a Machine Learning Approach,” unpublished manuscript.
- Bocola, L., G. Bornstein, and A. Dovis. 2019. “Quantitative Sovereign Default Models and the European Debt Crisis,” *Journal of International Economics*, Vol. 118, pp. 20–30.
- Bruns, M. and T. Poghosyan. 2018. “Leading Indicators of Fiscal Distress: Evidence from the Extreme Bound Analysis,” *Applied Economics*, 50(13), pp. 1454–78.
- Bussiere, M. and M. Fratzscher. 2006. “Towards a New Early Warning System of Financial Crises,” *Journal of International Money and Finance*, 25(6), pp.953–973.
- Callier, P. 1985. “Further Results on Countries’ Debt-servicing Performance: The Relevance of Structural Factors,” *Review of World Economics*, 121(1), pp.105–115.
- Cameron, A.C., J. B. Gelbach, and D. L. Miller. 2011. “Robust Inference with Multiway Clustering,” *Journal of Business & Economic Statistics* Vol. 29, No. 2, pp. 238–49.
- Catão, L. and B. Sutton. 2002. “Sovereign Defaults: The Role of Volatility,” IMF Working Paper No. 02/149 (Washington: International Monetary Fund).
- 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., and H. Zeaiter. 2014. “The Determinants of Sovereign Default: A Sensitivity Analysis,” *International Review of Economics and Finance*, 33, pp. 300–18.
- Checherita-Westphal, C., A. Klemm, A., and P. Viefers. 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., and G. Trebeschi. 2005. “Designing an Early Warning System for Debt Crises.” *Emerging Markets Review* 6, pp. 376–95.
- Citron, J.T. and Nickelsburg, G. 1987. “Country Risk and Political Instability,” *Journal of Development Economics*, 25(2), pp. 385–92.

- Cline, W. 1984. Interest and Debt: Systematic Risk and Policy Response. *Washington: Institute of International Finance*.
- Cortes, C. and V. Vapnik. 1995. "Support-vector Networks," *Machine Learning*, Vol. 20, No. 3, pp. 273–97.
- Cruces J.J. and C. Trebesch. 2013. "Sovereign Defaults: The Price of Haircuts," *American Economic Journal: Macroeconomics*, 5(3), pp. 85–117.
- Crafts, N. 2016. "Reducing High Public Debt Ratios: Lessons from UK Experience," *Fiscal Studies*, Vol. 37, iss. 2, pp. 201–23.
- Cruz, C., P. Keefer and C. Scartascini. 2018. "Database of Political Institutions 2017 (DPI2017)." Inter-American Development Bank. Numbers for Development.
- Dawood, M., N. Horsewood, and F. Strobel. 2017. "Predicting Sovereign Debt Crises: An Early Warning System Approach," *Journal of Financial Stability*, 28, pp. 16–28.
- De Cos, Pablo Hernández, G. Koester, E. Moral-Benito, and C. Nickel. 2014. "Signaling Fiscal Stress in the Euro Area: A Country-Specific Early Warning System," ECB Working Paper No. 1712 Frankfurt: European Central Bank.
- Degenhardt, F., S. Seifert, and S. Szymczak. 2019. "Evaluation of Variable Selection Methods for Random Forests and Omics Data Sets," *Briefings in Bioinformatics* 20 (2), pp. 492–503.
- Detragiache, E. and A. Spilimbergo. 2001. "Crises and Liquidity: Evidence and Interpretation," IMF Working Paper No. 01/02 (Washington: International Monetary Fund).
- Díaz Uriarte, R. and S. Álvarez de Andres. 2006. "Gene Selection and Classification of Microarray Data using Random Forest," *BMC Bioinformatics*, 7:3.
- Duval, B., and J. Hao. 2010. "Advances in Metaheuristics for Gene Selection and Classification of Microarray Data," *Briefings in Bioinformatics* 11 (1), pp. 127–41.
- 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. and Woo, J. 2017. "The Puzzle of Persistently Negative Interest-Rate–Growth Differentials: Financial Repression or Income Catch-Up?" *Fiscal Studies*, 38(2), pp.179–217.

- Feder, G., R.E. and Just. 1977. "A Study of Debt Servicing Capacity Applying Logit Analysis," *Journal of Development Economics*, 4(1), pp. 25–38.
- Feder, G., R. Just, and K. Ross. 1981. "Projecting Debt Servicing Capacity of Developing Countries," *Journal of Financial and Quantitative Analysis*, 16(5), pp. 651–69.
- Fernandez-Delgado, M., E. Cernadas, S. Barro, and D. Amorim. 2014. "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?" *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 3133–181.
- Fioramanti, M. 2008. "Predicting Sovereign Debt Crises using Artificial Neural Networks: A Comparative Approach," *Journal of Financial Stability*, Vol. 4, pp. 149–64.
- Fischer, S., R. Sahay, and C. Vegh. 2002. "Modern Hyper- and High Inflation," *Journal of Economic Literature*, 40, pp. 837–80.
- Frank, C. R., and Cline, W. R. 1971. "Measurement of Debt Servicing Capacity: An Application of Discriminant Analysis," *Journal of International Economics*, 1, pp. 327–44.
- Frankel, J., and G. Saravelos. 2012. "Can Leading Indicators Assess Country Vulnerability? Evidence from the 2008–09 Global Financial Crisis," *Journal of International Economics* 87, pp. 216–31.
- Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* Vol. 29, No. 5, pp. 1189–1232.
- Friedman, J. H., and B. E. Popescu. 2008. "Predictive Learning Via Rule Ensembles," *The Annals of Applied Statistics*, Vol. 2, No. 3, 916–54.
- Fuertes, A., and E. Kalotychou. 2006, "Early Warning Systems for Sovereign Debt Crises: The Role of Heterogeneity," *Computational Statistics & Data Analysis* 51, pp. 1420–41.
- . 2007. "Optimal Design of Early Warning Systems for Sovereign Debt Crises," *International Journal of Forecasting*, 23(1), pp. 85–100.
- Garín, J., R. Lester, E. Sims, and J. Wolff. 2019. "Without Looking Closer, It May Seem Cheap: Low Interest Rates and Government Borrowing," *Economic Letters* 180, pp. 28–32.
- Gelos, R., R. Sahay, and G. Sandleris, 2004, "Sovereign Borrowing by Developing Countries: What Determines Market Access?" IMF Working Paper No. 04/211 (Washington: International Monetary Fund).
- Genuer, R., J. Poggi, and C. Tuleau-Malot. 2015. "VSURF: An R Package for Variable Selection Using Random Forests," *The R Journal*, R Foundation for Statistical Computing, 7 (2), pp.19–33. hal-01251924v1

- Georgievska, A., L. Georgievska, A. Stojanovic, and N. Todorovic. 2008. "Sovereign Rescheduling Probabilities in Emerging Markets: A Comparison with Credit Rating Agencies' Ratings," *Journal of Applied Statistics*, 35(9), pp. 1031–51.
- Ghulam, Y., and J. Derber. 2018. "Determinants of Sovereign Defaults," *The Quarterly Review of Economics and Finance* 69, pp. 43–55.
- Gourinchas, P. and M. Obstfeld. 2012. "Stories of the Twentieth Century for the Twenty-First," *American Economic Journal: Macroeconomics*, Vol. 4(1), pp. 226–65.
- Greenwell, B. M. 2017. "Pdp: An R Package for Constructing Partial Dependence Plots." *The R Journal* 9 (1), pp. 421–36. <https://journal.r-project.org/archive/2017/RJ-2017-016/index.html>.
- Guscina, A., M. Sheheryar, and M. Papaioannou. 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," *Journal of Econometrics*, 36(1-2), pp. 205–30.
- . 1994. "A Simulation Estimation Analysis of the External Debt Crises of Developing Countries," *Journal of Applied Econometrics*, 9, pp. 109–131.
- Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer-Verlag, New York.
- Hilario, M., and A. Kalousis. 2008. "Approaches to Dimensionality Reduction in Proteomic Biomarker Studies," *Briefings in Bioinformatics* 9 (2), pp. 102–118.
- Hilscher, J. and Y. Nosbusch. 2010. "Determinants of Sovereign Risk: Macroeconomic Fundamentals and the Pricing of Sovereign Debt," *Review of Finance* 14, iss. 2: pp. 235–62.
- IMF. 2015. *The Fund's Lending Framework and Sovereign Debt—Further Considerations*. Board Paper (Washington: International Monetary Fund).
- . 2017. *Review of the Debt Sustainability Framework in Low-income Countries: Proposed Reforms*.
- Jordà Ò., M. Schularick, and A. M. Taylor. 2016. "Sovereigns versus Banks: Credit, Crises, and Consequences," *Journal of the European Economic Association* 14 (1), pp. 45–79.
- Kolscheen, E. 2010. "Sovereign Risk: Constitutions Rule." *Oxford Economic Papers*, New Series, Vol. 62, No. 1, pp. 62–85.

- Kose, A., S. Kurlat, F. Ohnsorge, and N. Sugawara. 2017. “A Cross-Country Database of Fiscal Space,” World Bank Development Prospects Group, Policy Research Working Paper No. 8157.
- Kraay, A., and V. Nehru. 2006. “When Is External Debt Sustainable?” *The World Bank Economic Review* 20 (3): pp. 341–65.
- Kursa, M., and W. Rudnicki. 2010. “Feature Selection with the Boruta Package,” *Journal of Statistical Software*, 36(11), pp. 1–13.
- Laeven, L. and F. Valencia. 2018. “Systemic Banking Crises Revisited,” IMF Working Paper No. 18/206 (Washington: International Monetary Fund).
- Lane, P., R., and G.M. Milesi-Ferretti. 2007. “The External Wealth of Nations Mark II.” *Journal of International Economics*, Vol. 73, pp. 223–50.
- . 2017. “International Financial Integration in the Aftermath of the Global Financial Crisis.” IMF Working Paper 17/115, International Monetary Fund, Washington, DC.
- Lee, S.H. 1991. “Ability and Willingness to Service Debt as Explanation for Commercial and Official Rescheduling Cases,” *Journal of Banking & Finance*, 15(1), pp. 5–27.
- Lloyd-Ellis, H., G.W. McKenzie, and S.H. Thomas. 1989. “Using Country Balance Sheet Data to Predict Debt Rescheduling,” *Economics Letters*, 31(2), pp. 173–177.
- Lundberg, S. M. and S. Lee. 2017. “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems*, pp. 4765–74.
- Ma, S., and J. Huang. 2008. “Penalized Feature Selection and Classification in Bioinformatics,” *Briefings in Bioinformatics* 9 (5), pp. 392–403.
- Maltritz, D., and A. Molchanov. 2014. “Country Credit Risk Determinants with Model Uncertainty,” *International Review of Economics & Finance*, 29, pp. 224–34.
- Manasse, P., N. Roubini, and A. Schimmelpfennig. 2003. “Predicting Sovereign Debt Crises,” IMF Working Paper No. 03/221 (Washington: International Monetary Fund).
- Manasse, P., and N. Roubini. 2009. “Rules of Thumb for Sovereign Debt Crises,” *Journal of International Economics*, 78(2), pp. 192–205.
- Marashaden, O. 1997. “A Logit Model to Predict Debt Rescheduling by Less Developed Countries,” *Asian Economies* 26, pp. 25–34.
- Mauro, P., R. Romeu, A. Binder, and A. Zaman. 2015. “A Modern History of Fiscal Prudence and Profligacy,” *Journal of Monetary Economics*, Vol. 76, pp. 55–70.

- Mauro, P., and J. Zhou. 2019. “ $r-g < 0$ : Can We Sleep More Soundly?” mimeo.
- 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).
- . 2018b. “The Global Debt Database: Methodology and Sources.” IMF Working Paper No. 18/111 (Washington: International Monetary Fund).
- McFadden, D., R. Eckaus, G. Feder, V. Hajivassiliou, V., and S. O’Connell. 1985. Is There Life After Debt? An Econometric Analysis of the Creditworthiness of Developing Countries. *International debt and the developing countries*, pp. 179–209.
- Medas, P., T. Poghosyan, Y. Xu, J. Farah-Yacoub, and K. Gerling. 2018. “Fiscal Crises,” *Journal of International Money and Finance*, Vol. 88, pp. 191–207.
- Messmacher, M. and M. Kruger. 2004. *Sovereign Debt Defaults and Financing Needs* (No. 4–53). International Monetary Fund.
- Mehrotra, N. 2017. “Debt Sustainability in a Low Interest Rate World,” Hutchins Center Working Paper No. 32
- Nogueira, S., and G. Brown. 2016. “Measuring the Stability of Feature Selection.” In: Frasconi, P., N. Landwehr, G. Manco, and J. Vreeken (eds) *Machine Learning and Knowledge Discovery in Databases*. ECML PKDD 2016. Lecture Notes in Computer Science, vol 9852. Springer.
- Pamies Sumner, S. and K. Berti. 2017. “A Complementary Tool to Monitor Fiscal Stress in European Economies,” EC Discussion Paper, 49 (June).
- 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.
- Reinhart, C. M. 2002. “Default, Currency Crises, and Sovereign Credit Ratings,” *World Bank Economic Review*, Oxford University Press, vol. 16(2), pages 151–70.
- Reinhart, C. M., V. R. Reinhart, and K. Rogoff. 2012. “Public debt Overhangs: Advanced Economy Episodes since 1800s,” *Journal of Economic Perspectives*, Vol. 26, No. 3 (Summer), pp. 69–86.
- Reinhart, C. M., and K. Rogoff. 2009, *This Time is Different: Eight Centuries of Financial Folly* (Princeton, NJ: Princeton University Press).

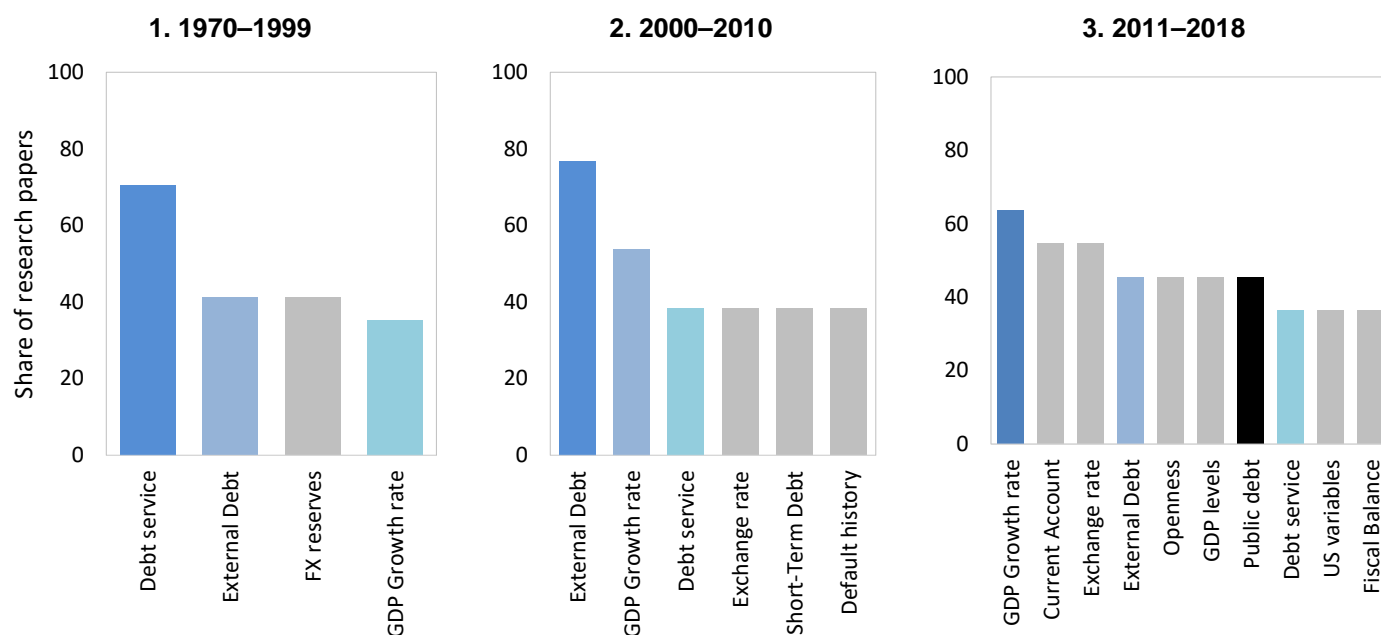
- . 2011a. “From Financial Crash to Debt Crisis,” *American Economic Review* 101(5), pp. 1676–1706.
- . 2011b. “The Forgotten History of Domestic Debt,” *Economic Journal* 121 (552), pp. 319–350.
- Reinhart, C. M., K. Rogoff and M. Savastano. 2003. “Debt Intolerance,” *Brookings Papers on Economic Activity*, 34, 2003-1, pp. 1–74.
- Rivoli, P. and Brewer, T.L. 1997. “Political Instability and Country Risk,” *Global Finance Journal*, 8(2), pp. 309–21.
- Rodriguez A., and P. N. Rodriguez. 2006. “Understanding and Predicting Sovereign Debt Rescheduling: A Comparison of the Areas Under Receiver Operating Characteristic Curves,” *Journal of Forecasting*, 25, pp. 459–79.
- Saeys, Y., Inza, I., and P. Larranaga. 2007. “A Review of Feature Selection Techniques in Bioinformatics,” *Bioinformatics* 23 (19), pp. 2507–17.
- Sargen, N. 1977. “Economic Indicators and Country Risk Appraisal,” *Economic Review*, Federal Reserve Bank of San Francisco, issue Fall, pp. 19–35.
- Savona, R., and M. Vezzoli. 2015. “Fitting and Forecasting Sovereign Defaults using Multiple Risk Signals,” *Oxford Bulletin of Economics and Statistics*, 77 (1), pp. 66–91
- Savona, R., Vezzoli, M., and E. Ciavolino. 2015. “A Data-Driven Explanation of Country Risk: Emerging Markets vs. Eurozone Debt Crises,” SYRTO Working Paper n.17.
- Snider, L.W. 1990. “The Political Performance of Third World governments and the Debt Crisis,” *American Political Science Review*, 84(4), pp. 1263–80.
- Solberg, R.L. 1988. *Sovereign Rescheduling: Risk and Portfolio Management*. Routledge.
- Strumbelj, E. and I. Kononenko. 2010. “An Efficient Explanation of Individual Classifications Using Game Theory,” *Journal of Machine Learning Research*, Vol. 11, pp. 1–18.
- Sturzenegger, F. and J. Zettelmeyer. 2006. *Debt Defaults and Lessons from a Decade of Crises*, Table 1, Chapter 1 (Cambridge: MIT Press).
- Sumner, S.P. and K. Berti. 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?" *Journal of Banking and Finance* 28(11), pp. 2845–67.
- Taffler, R.J. and B. Abassi. 1984. "Country Risk: A Model for Predicting Debt Servicing Problems in Developing Countries," *Journal of the Royal Statistical Society: Series A (General)*, 147(4), pp. 541–61.
- Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B*, Vol. 58, No. 1, pages 267–88.
- Wolpert, D. H., and W. G. Macready. 1997. "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 1, pp. 67–82.

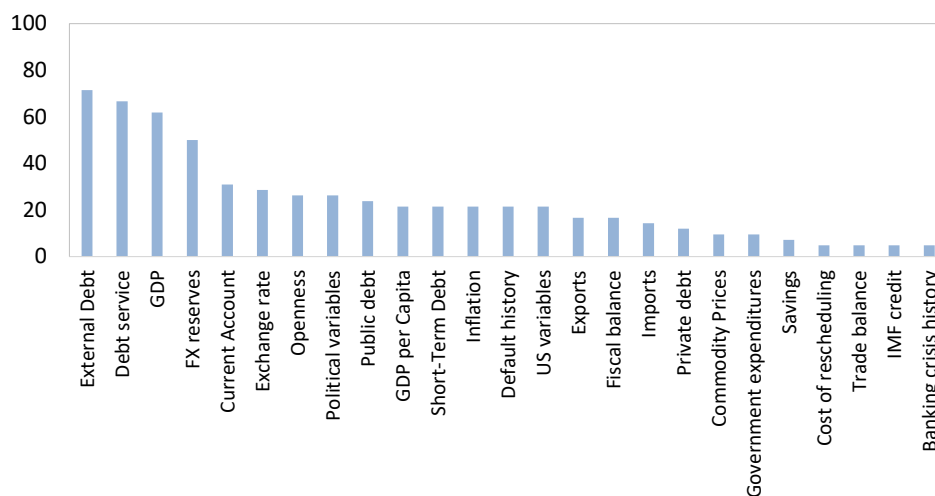
## FIGURES

**Figure 1. Predictors of Fiscal Crises in the Literature**



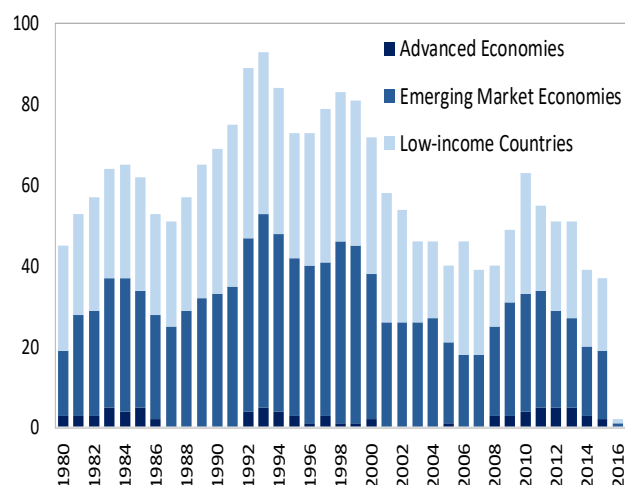
Note: The charts are based on a literature review of 42 empirical papers (for more details, see Appendix 1). Variables plotted are those that are statistically significant in at least a third of the papers published during the period of reference.

**Figure 2. Most Common Predictors in the Literature**  
(Share of surveyed papers, 1970–2018)

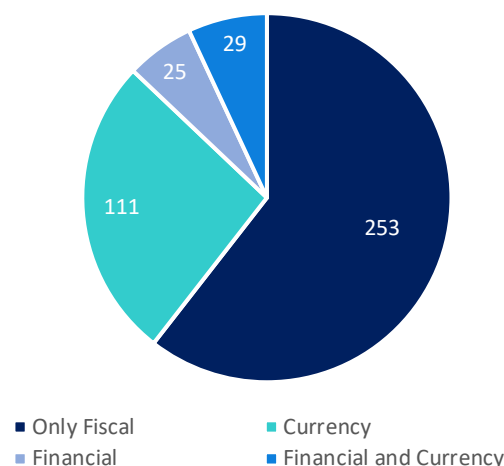


Notes: The chart is based on a literature review of 42 empirical papers (for more details, see Appendix 1).

**Figure 3. Countries with Fiscal Crises, 1980–2016**  
(Number)



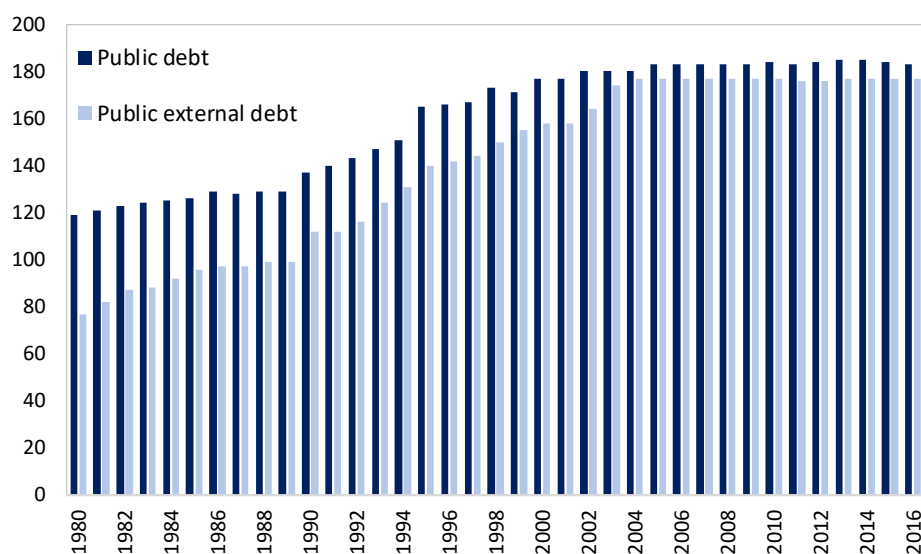
**Figure 4. Overlap with Other Crises, 1980–2016**  
(Number of crises episodes)



Sources: Bloomberg; Datastream; Eurostat; Gelos, Sahay, and Sandleris (2004); Guscina, Sheheryar, and Papaioannou (2017); IMF, *International Financial Statistics*; Laeven and Valencia (2018); OECD; Reuters; and authors' calculations.

1/ Two crises are identified as overlapping if they start within two years of each other. Financial crises are banking crises episodes as defined in Laeven and Valencia (2018).

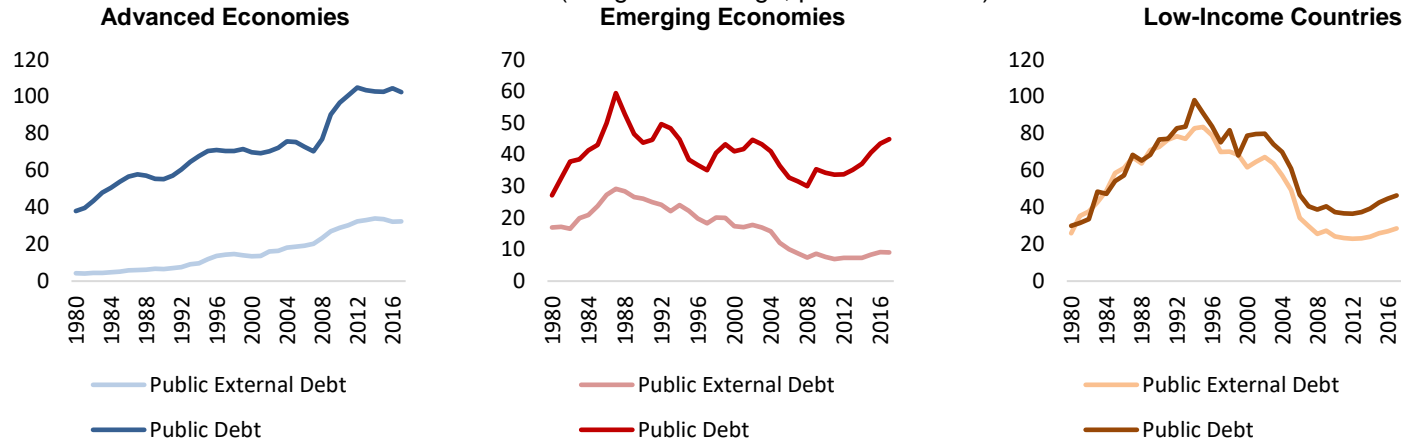
**Figure 5. Debt Statistics: Country Coverage, 1980–2016**  
(Number of countries)



Sources: IMF, *Global Debt Database*; IMF, *World Economic Outlook*; World Bank, *World Development Indicators*; U.S. Bureau of Economic Analysis; Haver; Arslanalp and Tsuda (2014); and authors' calculations.

**Figure 6. Public and Public External Debt, 1980–2016**

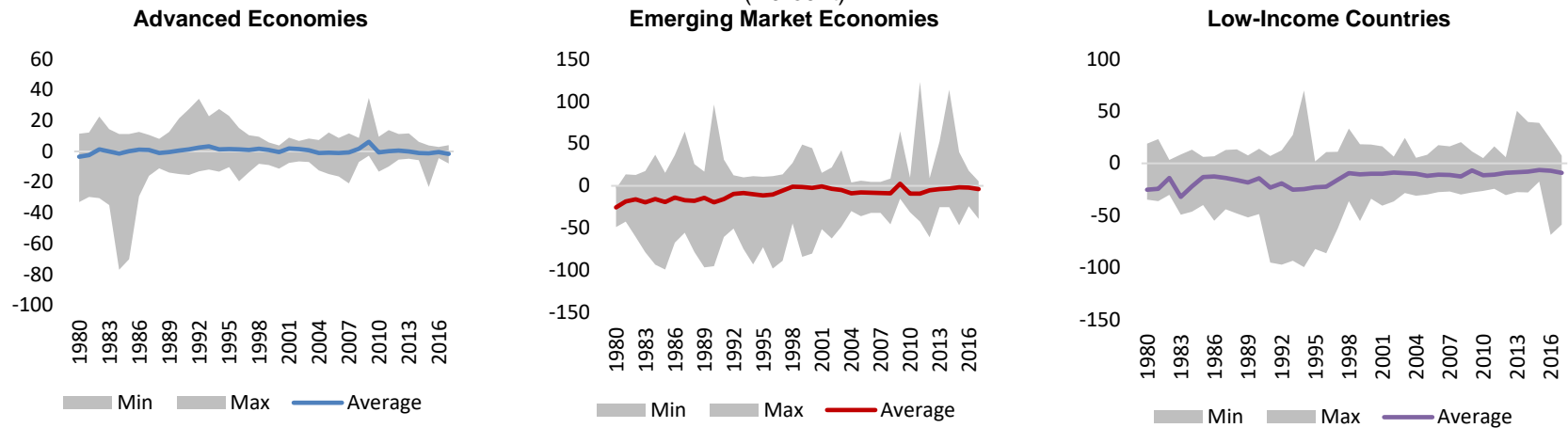
(Weighted average, percent of GDP)



Sources: IMF, *Global Debt Database*; IMF, *World Economic Outlook*; World Bank, *World Development Indicators*; U.S. Bureau of Economic Analysis; Haver; Arslanalp and Tsuda (2014); and authors' calculations. Note: Public external debt refers to public and publicly guaranteed debt.

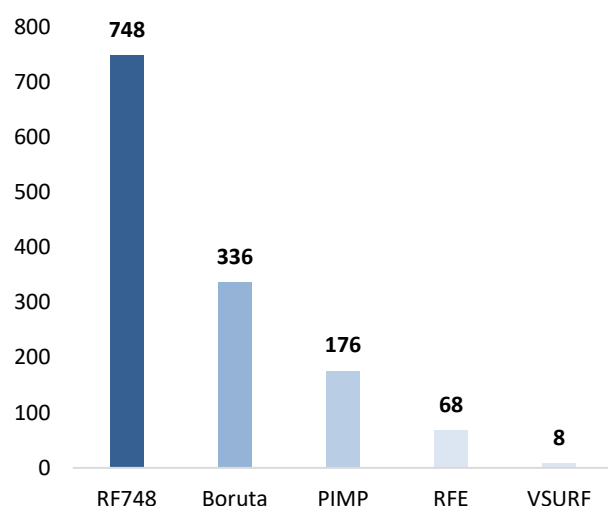
**Figure 7. Interest-Growth Differential, 1980–2016**

(Percent)



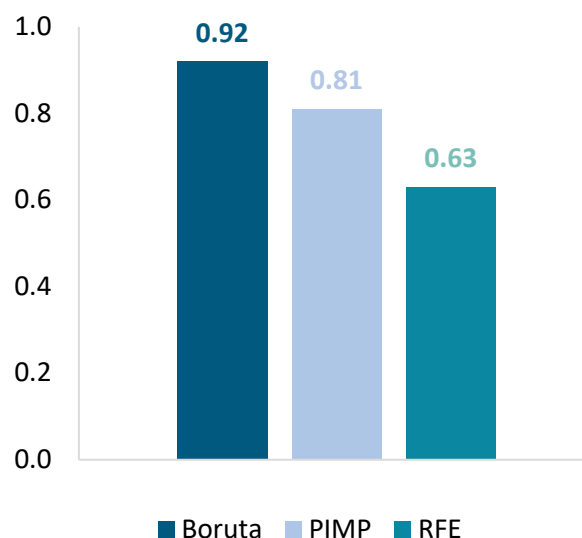
Sources: IMF, *Global Debt Database*; IMF, *World Economic Outlook*; and authors' calculations.

**Figure 8. Feature Selection Algorithms**  
(Number of variables selected)



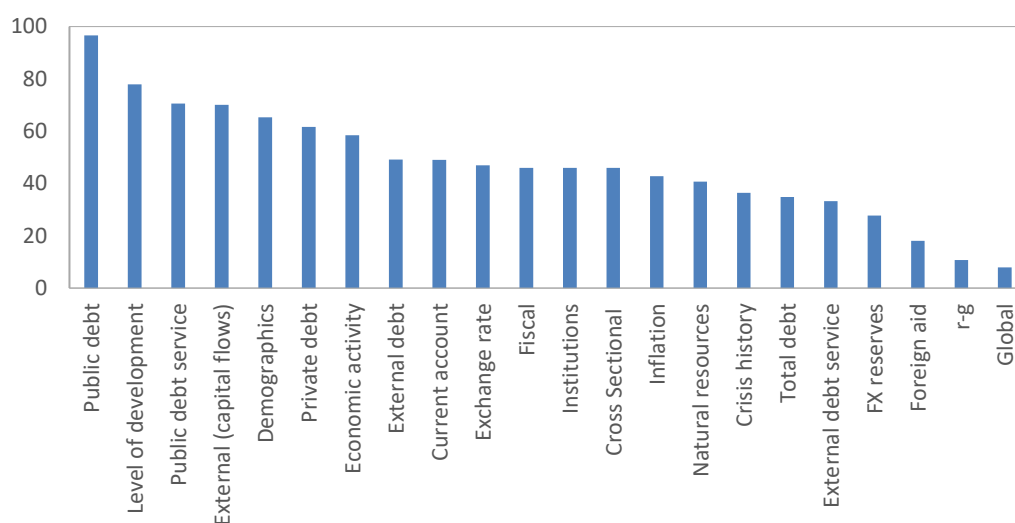
Note: The chart shows the number of variables selected by each feature selection algorithm and the full RF model estimated over the full sample.

**Figure 9. Robustness in Variable Selection**  
(Pearson correlation coefficient)



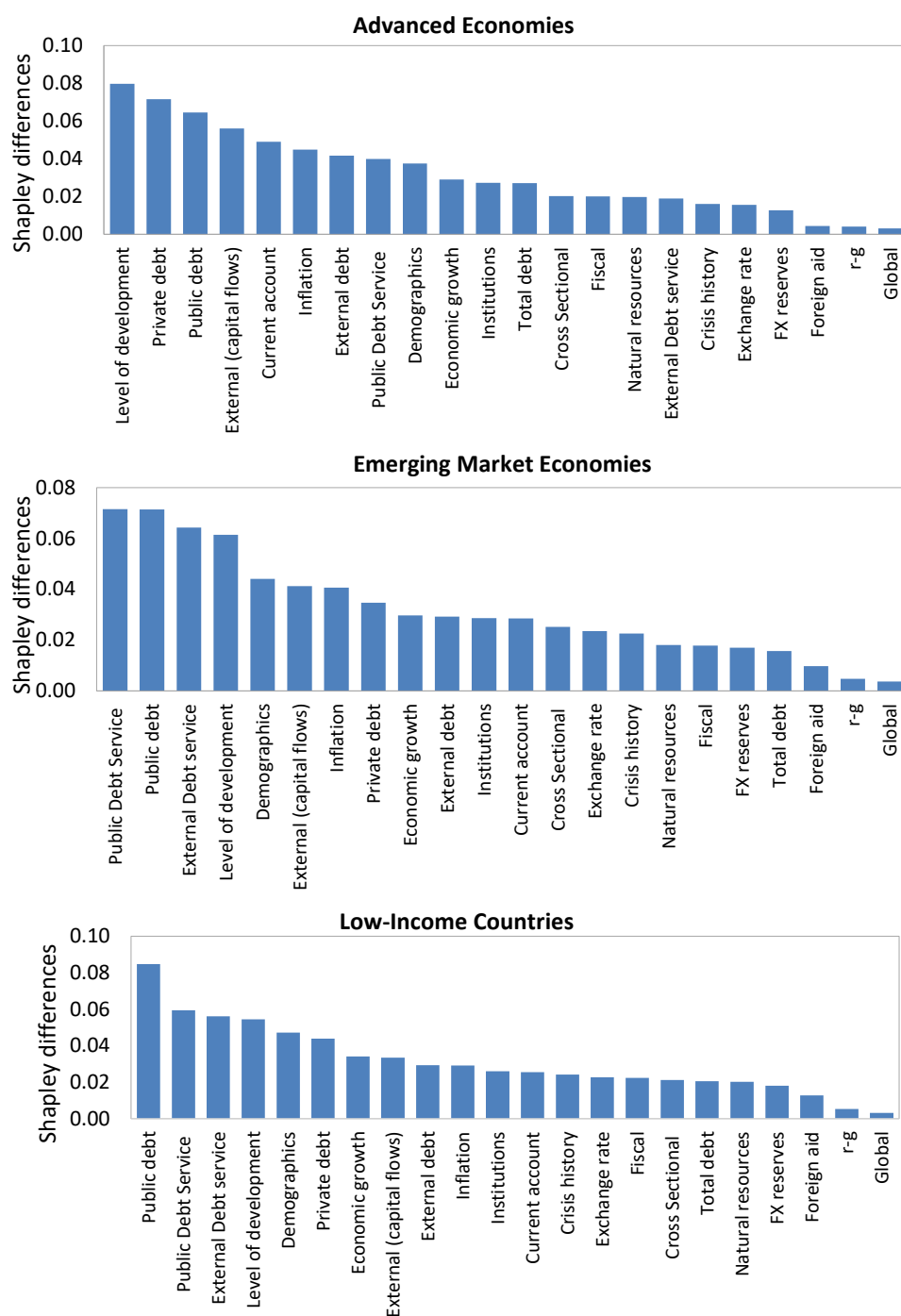
Note: The Pearson index provides a measure the stability of the chosen feature set to variations in the training data. Alternative samples are drawn randomly dropping 5 percent of observations from the original dataset. The Pearson Correlation Coefficient gives an indication of the overlap of features across samples (takes values between -1, no overlap, and 1, full overlap).

**Figure 10. Variable Importance by Group of Predictors**

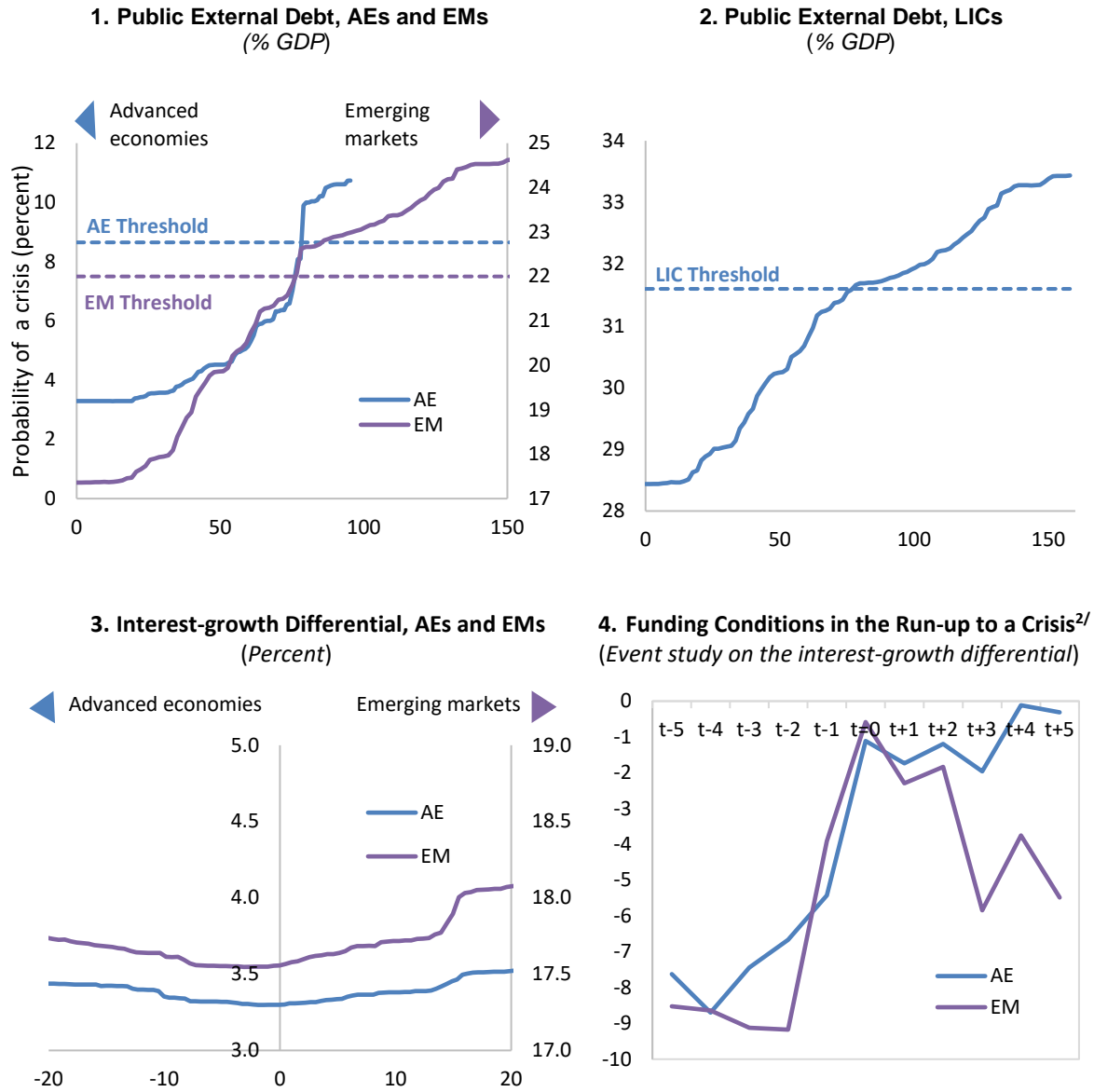


Notes: Variable importance is calculated using an in-built out of bag permuted predictor importance function in R based on the RF estimated with the variables selected by Boruta.

**Figure 11. Contribution to Probability of a Crisis**  
(Shapley Values)

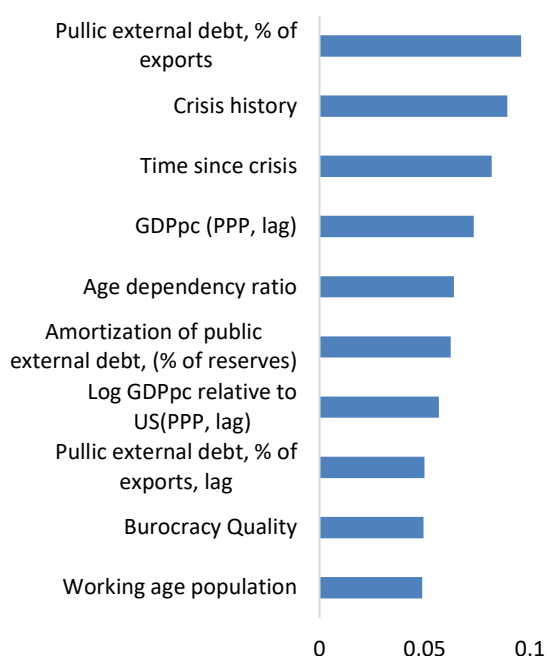


Notes: These charts display the mean Shapley value difference (crisis versus non-crisis observations) by income groups.

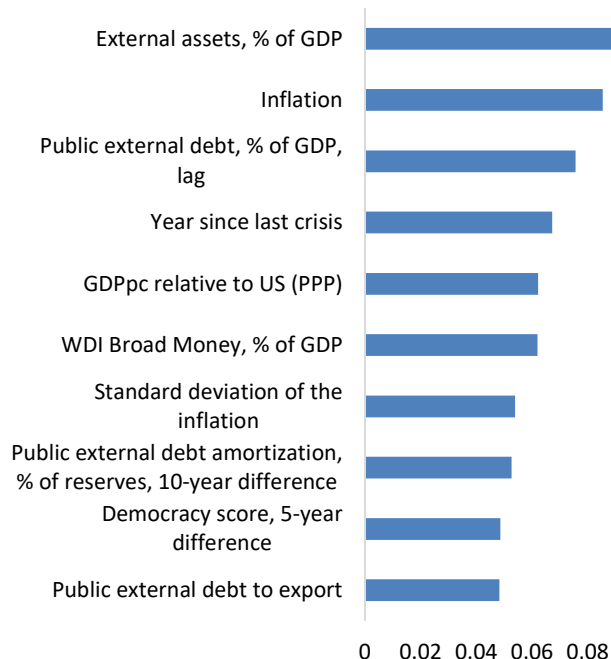
**Figure 12. Partial Dependence Plots<sup>1/</sup> and Event Studies**

1/ Charts (1)-(3) display PDPs 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 (missed crises and false alarms).

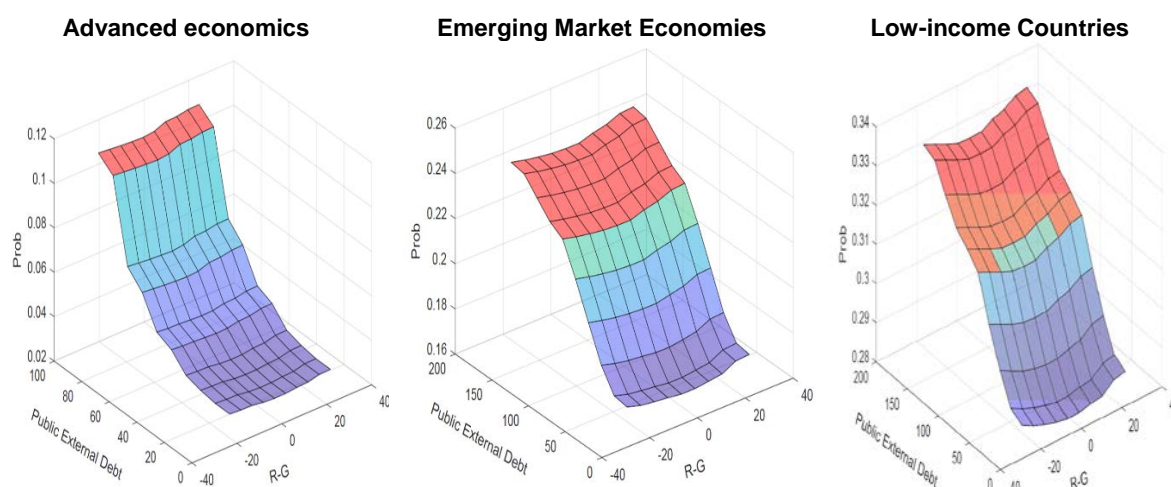
2/ Chart (4) display an event study based on the framework developed by Gourinchas and Obstfeld (2012) where  $t=0$  is the start of the fiscal crisis. We estimate the equation  $y_{i,t} = \alpha_i + \sum_{j=-5}^5 \beta_{t+j} D_{i,t+j} + \varepsilon_{i,t}$  where  $y$  is the interest-growth differential, and  $D_{i,t}$  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 should be interpreted as the interest-growth differential at time  $t+k$ , relative to "non-crisis" times benchmark.

**Figure 13. Overall Interaction Strength**

Notes: The chart shows the interaction strength (H-statistic) for each feature with all other features for the Boruta RF. Public external debt has the highest relative interaction effect with all other features.

**Figure 14 Top-10 Interactions with Public External Debt**

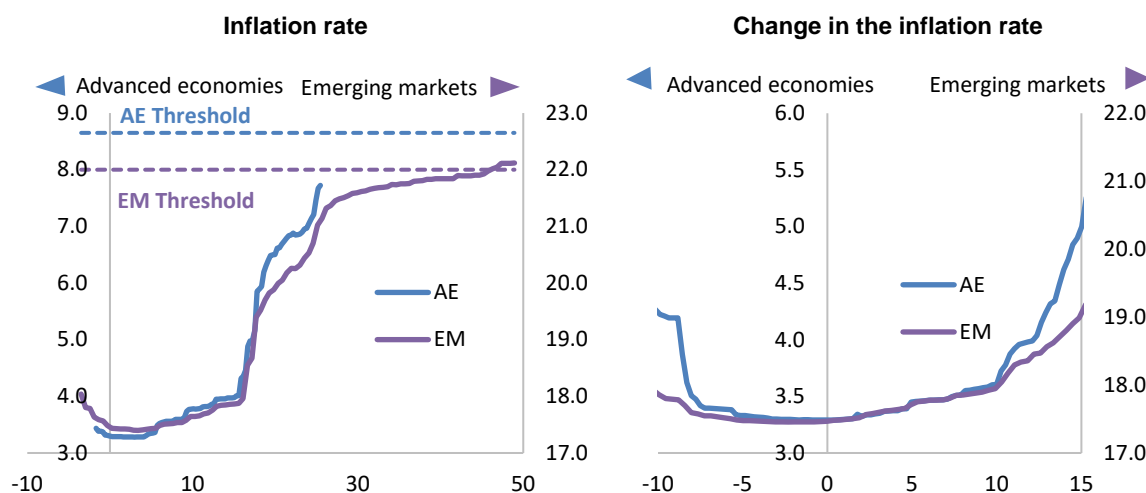
Notes: The chart shows the 2-way interaction strengths (H-statistic) between public external debt and each other feature.

**Figure 15. Bivariate Partial Dependent Plots: Public External Debt and r-g**

Notes: Charts display bivariate partial dependent plots for different country groupings. Cells highlighted in red depicts 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.

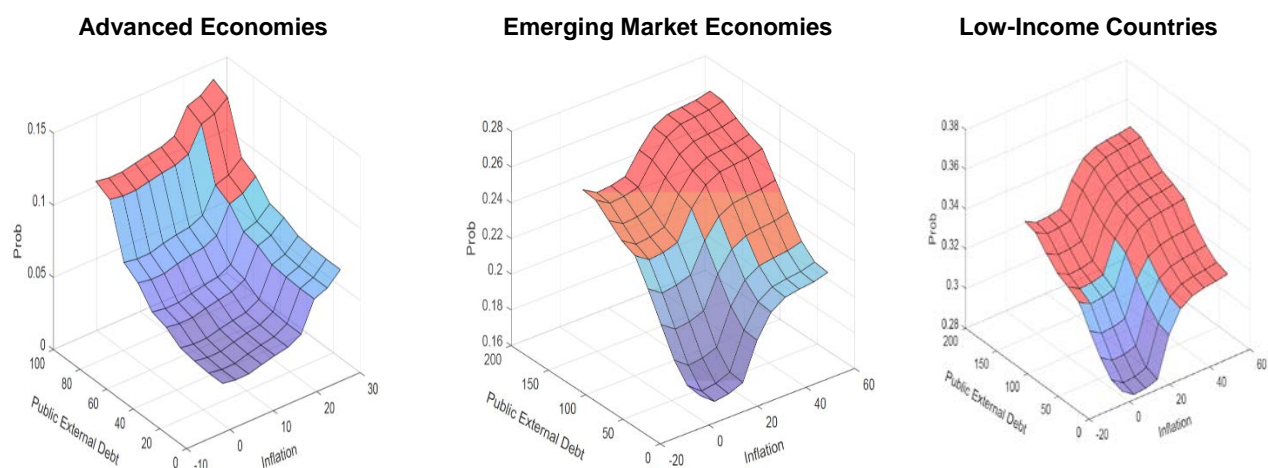


**Figure 16. Inflation: Univariate Partial Dependence Plots**  
(Percent)



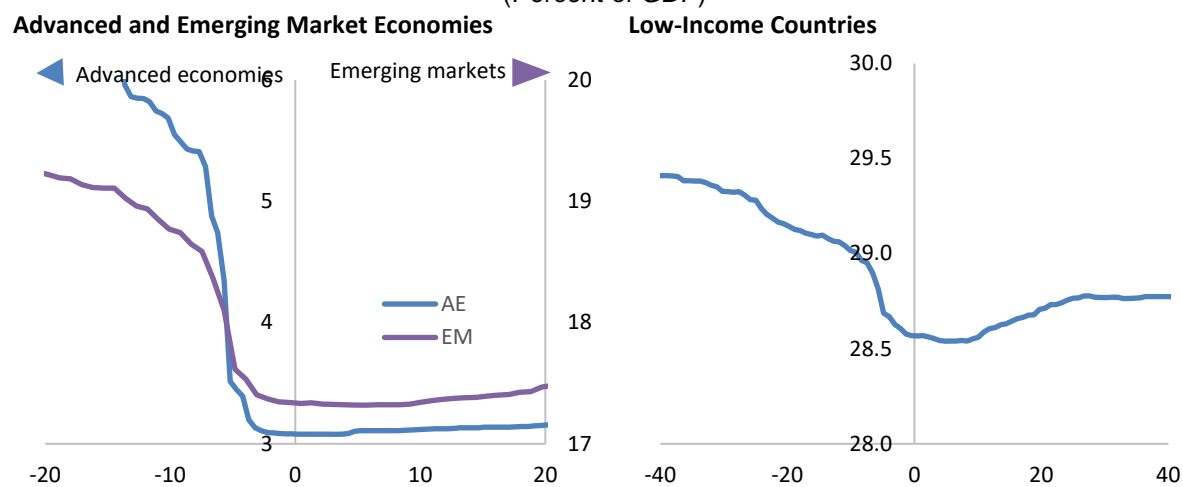
Note: The charts display PDPs based on the Boruta random forest. Estimated probabilities are plotted in the vertical axis and inflation in the horizontal axis. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt.

**Figure 17. Inflation and Public External Debt: Bivariate Partial Dependence Plots**



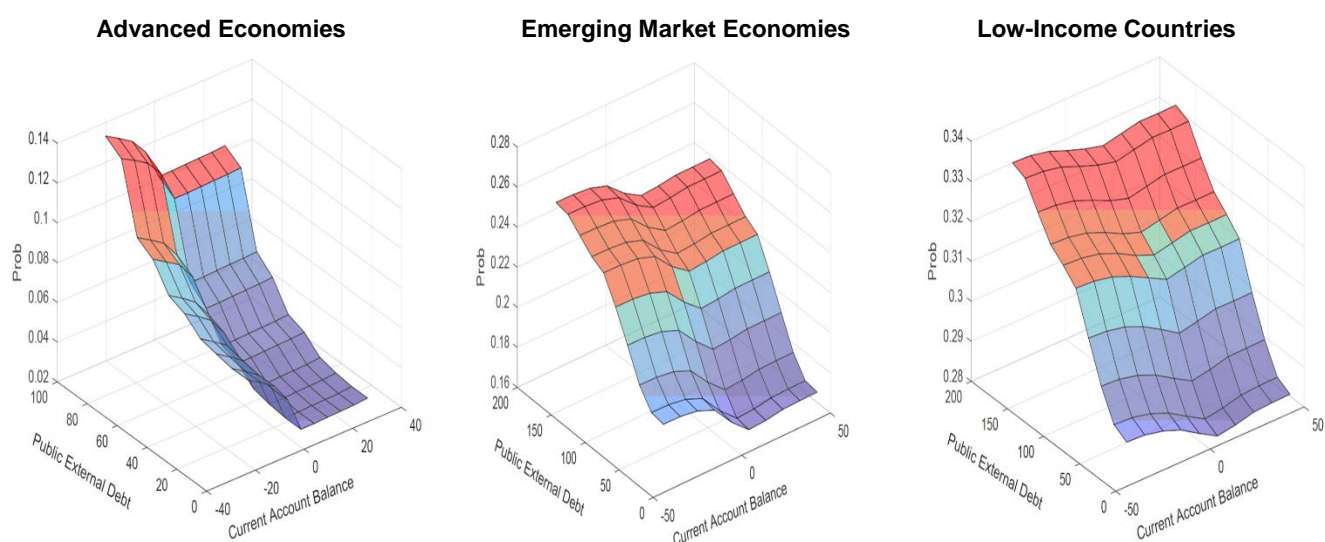
Notes: Charts display bivariate PDPs for different country groupings. Cells highlighted in red depicts combinations of public external debt and inflation for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing the sum of type I and type II errors (missed crises and false alarms). The darker the blue color, the lower the probability of a crisis.

**Figure 18. Current Account: Univariate Partial Dependence Plots**  
(Percent of GDP)



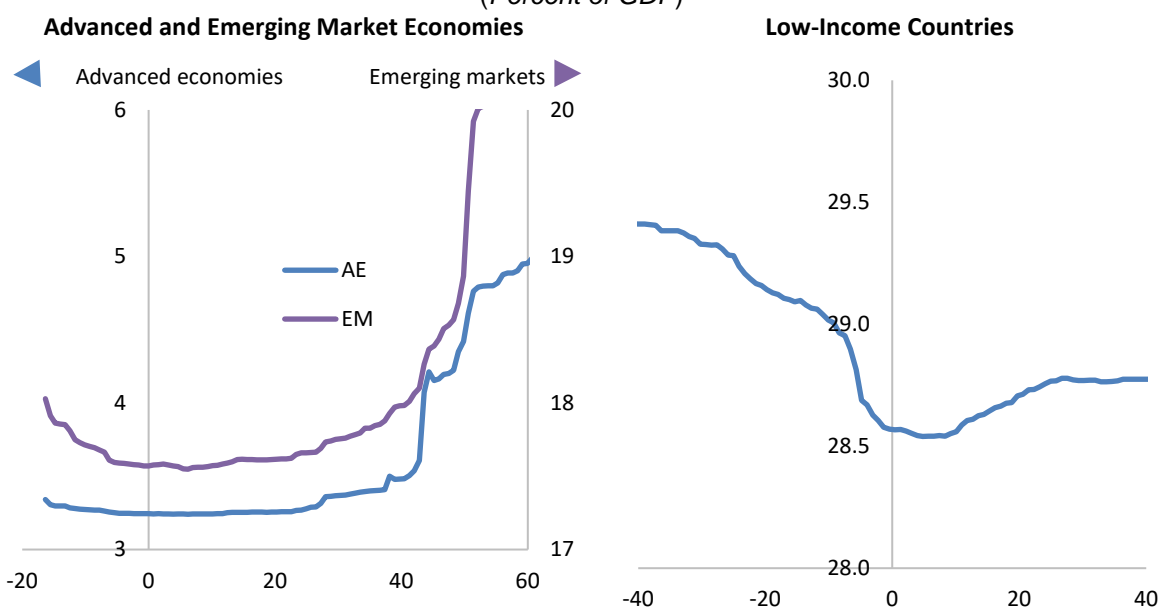
Note: The charts display PDPs based on the Boruta random forest. Estimated probabilities are plotted in the vertical axis and the current account in the horizontal axis. Solid lines show the PDP curve which represents the average prediction across all levels of public external debt.

**Figure 19. Current Account and Public External Debt: Bivariate Partial Dependence Plots**  
(Percent of GDP)



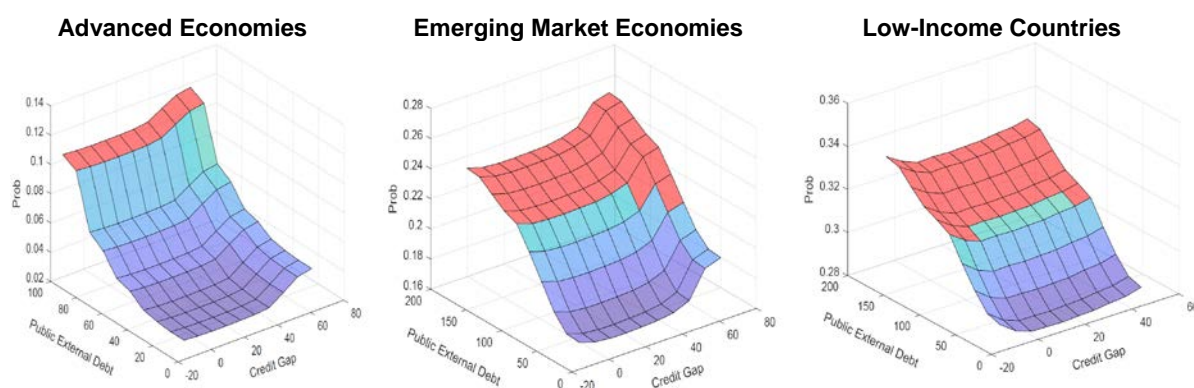
Notes: Charts display bivariate PDPs for different country groupings. Cells highlighted in red depicts combinations of public external debt and the current account balance for which the estimated probability of a crisis is above the probability thresholds calculated for that income group based on minimizing 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.

**Figure 20. Credit Gap: Univariate Partial Dependence Plots**  
(Percent of GDP)



Note: The charts display 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.

**Figure 21. Credit Gap and Public External Debt: Bivariate Partial Dependence Plots**  
(Percent of GDP)



Notes: Charts display bivariate PDPs for different country groupings. Cells highlighted in red depicts 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.

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Frank and Cline (1971)		26 countries over 1960–68 (13 rescheduling in 8 countries)	Linear discriminant function	Debt service to export earnings; debt amortization to total debt outstanding (inverse of average maturity); imports to reserves
Sargen (1977)	Debt rescheduling	44 countries over 1960–1976	Linear discriminant function	Inflation; debt service to exports
Feder (1977)	When PPG payments to lending institutions are delayed or rescheduled	41 countries, 1965–1972	Logit	Debt service/exports, imports/reserves, amortization/debt, income/capita, capital inflows/debt service, GDP growth, export growth
Feder, Just, Ross (1981)	Debt rescheduling or arrears	56 countries 1965–1976	Logit	Debt service to exports; debt service to exports squared; FX reserves to imports; FX reserves to imports squared; ratio of net noncommercial FX inflows to debt service payments; ratio of commercial FX inflows to debt service payments squared; exports to GNP; relative per capita GNP to US per capita GNP; regional dummies
Cline (1984)	Debt rescheduling	60 countries for 1968–1982 (22 rescheduling episodes)	Logit	Debt service to exports of goods and services; reserves to imports; per capita economic growth; current account squared (correcting for sign); gross debt minus reserves to exports of goods and service; amortization rate of debt; total net external borrowing by all nonoil developing countries/total imports
Taffler and Abassi (1984)	Debt rescheduling	95 developing countries from 1968–1978 (55 rescheduling of 14 countries)	Linear discriminant model	Commitments per capita; debt to exports; average rate of inflation; domestic credit to GDP

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Callier (1985)	Debt rescheduling, moratorium on servicing debt, arrears of payments of imports or interest	62 countries for 1977–1982	Logit	Public external debt/GDP; current account/GDP; difference between GNP and GDP growth; investment/GDP; index of relative productivity of the country compared to industrial countries; openness index (exports+imports/GDP+imports); population (natural log)
McFadden, Eckaus, Feder, Hajivassiliou, and O'Connell (1985)	Arrears on interest on long-term debt, arrears on principal of long-term debt, higher tranche IMF support, or rescheduling	93 developing countries over the period 1970–1982	Probit and one-factor models; logit	PROBIT-- debt service due to exports; reserves to imports; LOGIT-- total reserves minus gold to GDP; imports/GDP; debt/exports; repayment indicator
Citron and Nickelsburg (1987)	Default (a country renegotiates the terms or conditions of a loan)	5 countries, 1960–1983		Reserves plus IMF credits, political instability
Hajivassiliou (1987)	Arrears on external debt (separates between 3 regimes, arrears but no crisis, debt crisis)	79 countries, 1970–1982	Logit	Debt/exports, debt service/exports, reserves/imports, history of past restructuring, GDP growth, real GNP per capita
Berg and Sachs (1988)	Whether a country had to reschedule debt in the period 1982-87	35 countries	Probit	Income distribution and agriculture are statistically significant at the five percent level. The outward-orientation variable is not significant for sample of 35 countries.
Solberg (1988)	Rescheduling of sovereign debt	67 countries, 1971–1984	Linear and logit	Composite indicators of debt service cost, rescheduling cost, rescheduling state, long-term structural, medium-term adjustment, ST liquidity
Lloyd-Ellis (1989)	Debt rescheduling	Annual sample of 27 countries, 1977–1981; 6-monthly data for 59 countries from 1977–1985	Logit	Loans to assets, medium-term debt/total debt, undisbursed credit commitments/total bank lending, unallocated credit, FX reserves/IMF Quota, number of rescheduling, average maturity period, value of rescheduling, grace period

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Snider (1990)	Suspension in payments of external debt	58 countries, 1970–1984	Logit	Relative political capacity, capital flight (net short-term capital outflows of the nonbank private sector plus recorded errors and omissions), debt service ratio to exports, institutional investor country credit rating, change in world price agriculture commodities, change in world price of minerals, change in world price of oil, 6-month LIBOR (previous year)
Lee (1991)	Rescheduling of sovereign external debt	75 countries, 1970–1985	Logit	For commercial creditors all variables are significant except for government domestic debt; for official creditors: total external debt, growth GDP, interest rate and domestic debt are significant at 15% level. Get similar results as for official creditors when pool all creditor together.
Hajivassiliou (1994)	Defined by a rescheduling arrangement of the obligations or involvement of the IMF.	79 countries, 1970–1982	Nested trinomial logit	Debt/exports, debt service/exports, GDP growth, GNP/capita
Rivoli and Brewer (1997)	Debt rescheduling from Global Development Finance WB database	80 developing countries, 1980–1990	Logit mode	Debt service/exports, reserves/imports, external debt/GNP, amortization, lagged rescheduling, political instability
Detragiache and Spilimbergo (2001)	Arrears>5% of debt or rescheduling of debt	69 countries (1971–1998)	Logit	Short-term debt, debt service due, reserves, total debt, multilateral share, overvaluation, openness
Catao and Sutton (2002)	Lindert and Mortan (1989), Reinhart (2001)	25 EMs, 1970–2001	Logit	Volatility of Tot, GDP growth, debt service/exports, nir/debt, fiscal balance, US interest rate, REER
Peter (2002)	Changes in the levels of debt arrears and amounts rescheduled.	78 EMs (1984–1997)	Panel logit	Default lagged, variability of real GNP-per capita growth and a political risk index, the percentage deviation of the real exchange rate from long-run trend, the current account/GDP ratio, the amount of arrears accumulated in the previous year, the M2/international reserves ratio, the CPI inflation rate, and a debt stock indicator (comprised of the debt/GDP and the debt/exports ratio)

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Manasse, Roubini, and Schimmelpfennig. (2003)	Default by Standard and Poor's or if access to non-concessional IMF financing > 100 percent of quota	47 EMs access countries (1970–2002)	Logit and binary recursive tree analysis	Ratio of short-term debt to reserves, ratio of debt services to reserves, ratio of current account balance to GDP, interest rate on US treasury bills, growth rate of GDP, dummy for inflation rate above 50%, dummy for past defaults, index of political freedom, dummy for years with presidential election
Kruger and Messmacher (2004)	Standard and Poor's to determine when a country is in default	42 countries spanning the period 1970–2001	Multivariate logit	PNF, external debt/GDP, CAB, GDP growth, external debt/export
Ciarlone and Trebeschi (2005)	Debt crises (outright default or failure of a country to be current on external obligations; and IMF program with 100% or above quota)	28 EMs, 1980–2001	Logit, multinomial logit	For entering a crisis: Interest payments on external debt; openness; export growth rate; FX reserves to total external debt. For remaining in a crisis: total external debt and ST debt-to GDP; FX reserves to total external debt; openness. Multinomial logit: entering a crisis: interest rate payments as share of FX reserves; openness; total external debt
Kraay and Nehru (2006)	Any of three forms of exceptional finance: significant arrears on external debt, Paris Club rescheduling, and non-concessional International Monetary Fund lending	132 Low and middle income, 1970–2002	Probit	Share of public or publicly guaranteed debt, debt to official creditor, concessionally, growth, CPIA, default history
Rodriguez and Rodriguez (2006)	Sovereign debt restructuring	53 middle-income countries, 1986–1994	Logit, MARS, tree-based, neural network	Logit: debt/GDP; previous payment history; MARS: previous payment history, debt/GNP; Tree: previous payment history, debt/GNP; Neural network: Previous payment history
Fuertes and Kalotychou (2006)	Debt crises: if (1) arrears on external debt; (2) rescheduling agreement	75 EMs and developing countries 1983–2000	Multivariate and univariate logit; k-means clustering; combination of three	External: volatility of export growth and trade balance/GDP. External debt/GDP, short-term to total debt, IMF credit to exports, degree of trade openness; Domestic conditions: private credit/GDP, GDP growth, growth in GNP volatility, real exchange rate; Others: short-term debt to reserves, government expenditure, gross domestic savings.

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Fioramanti (2008)	Sovereign debt crises (sovereign default or IMF loan with access above 100% quota)	46 EMs for 1980–2004	Artificial neural networks	No discussion (**not possible to interpret the marginal effect of an increase in an independent variable)
Georgievska et al (2008)	Sovereign debt rescheduling	124 EMs, 1981–2002	Panel logit	Lagged rescheduling total debt/GNP, icrg, interest arrears /exports, import/GDP, export/GDP, CA balance/GDP
Hilscher and Nosbusch (2010)	Yield spread over U.S. Treasuries	32 EMs, 1994–2007	Hazard model	Volatility of ToT, external debt/GDP, reserves/GDP
Manasse and Roubini (2009)	Sovereign debt crisis; default; large IMF program (100% or more)	47 EMs, 1970–2002	Classification and Regression Tree	External debt to GDP; ST debt to reserves; real GDP growth; public external debt as share of fiscal revenue; CPI inflation; number of years to next presidential election; US treasury bills rate; external financial requirements; exchange rate overvaluation; exchange rate volatility
Kolscheen (2010)	Rating agency considers that country defaulted on its external debts	59 countries (13 years)	Pooled probit	Growth, external debt/GDP, debt service, T-bill, Parliamentary, Executive turnover, polcon
Reinhart and Rogoff (2011a)	Sovereign debt crises	70 countries, 1824–2009	Multinomial logit	Domestic banking crisis, change in public debt, financial center crisis
Chakrabarti and Zeaiter (2014)	Sovereign default (total principal and interest arrears to external debt)	190 countries, 1970–2010	Extreme Bound Analysis	Robust predictors: credit worthiness, growth, leverage on export earnings, debt service ratio, reserves, inflation, exchange rate, trade deficit, corruption, and democratic accountability. Not robust: openness, central bank liabilities, interest payments, cost of borrowing, imports, exports, per capita GNP, and government stability
De Cos, Pablo Hernández, Koester, Moral-Benito, Nickel (2014)	Any of the following four criteria is satisfied: i) inflation rate above 35%, ii) significant sovereign bond yield spreads (two standard deviations above their country - specific mean), iii) public debt default/restructuring/rescheduling and/or iv) a large - scale IMF - supported program.	European Economic and Monetary Union countries (excluding Luxembourg) 1970–2010	Signaling approach	Fiscal balance, the cyclically adjusted balance, value-added in construction, the net international investment position, the real exchange rate and GDP, short-term debt, the current account and household debt



## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Maltritz and Molchanov (2014)	Sovereign yield spreads	31 EMEs, 19 OECD (1996–2011)	Bayesian model averaging	Per capita GDP, GDP growth, Inflation, Budget balance/GDP, Debt service ratio (the ratio of debt payments to exports), Total debt/GDP, One year US interest rate, Capital investment/GDP, FX reserves/GDP, Money supply change, Openness indicator Defined as (exports + imports) / GDP, Trade balance/GDP Defined as (exports – imports) / GDP, Currency depreciation, Growth rate of imports, Growth rate of exports, Growth rate of FX reserves, FX reserves/imports, Overall freedom, Governance practices and restrictions on economic activities in countries, Business freedom, Trade freedom, Fiscal freedom, Government spending, Monetary freedom, Investment freedom, Financial freedom, Property rights Freedom from corruption
Savona and Vezzoli (2015)	Sovereign default	66 EMEs and Greece, Italy, Portugal, and Spain	Final regression tree	Short-term debt to reserves; default history; real GDP growth; US interest rates; exchange rate overvaluation
Dawood, Horsewood and Strobel (2017)	Sovereign debt crises (arrears or IMF program; or debt restructuring)	38 AEs and EMEs, 1980–2012	Binary logit; multinomial logit	For binomial logit: Total debt, IMF credit, global interest; FX reserves; Current account; FDI; Real GDP growth; REER overvaluation; Gov expenditures; National savings; Gov bank loans. At 5%: openness; domestic credit. Not all robust to different specifications. For multinomial logit: global interest; current account; at 5%: total debt; real GDP growth; REER overvalue;
IMF (2017)	Defaults, restructuring of external debt, large IMF programs	80 LICs, 1970–2014	Probit	Institutions, external debt to GDP, debt to exports, debt service to revenue, debt service to exports, domestic growth, reserves, remittances, world growth, GDP per capita, openness, FX income, risk premium, Dummy for conflicts

## APPENDIX 1. LITERATURE REVIEW

Reference	Definition of crisis	Sample	Model	Importance/Significance
Sumner and Berti (2017)	Fiscal crises	35 countries; 1970–2015	Signal extraction; logit	For the logit1 (less data): gross public debt; Current account; real GDP growth; World GDP growth; at 10%: change in gross public debt; private sector credit flow. For logit2 (larger sample): private sector credit flow; world GDP growth; at 5%: current account; real GDP growth; at 10% change in gross public debt. For the S0 (signal): yield curve; private sector credit flow; net savings of households; current account; net international investment position; gross financing needs; cyclical adjusted balance; share construction on GDP; GDP per capita in PPP; ST debt of households; ST debt of nonfinancial corporates; ST debt GG; Net debt; private sector debt; change in nominal unit labor costs; primary balance; gross debt; change in gross debt; change in expenditures GG; change in REER; Real GDP growth.
Bruns and Poghosyan (2018)	Fiscal crises	29 AEs and 52 EMEs, 1970–2015	Extreme Bound Analysis; Logit	Output gap; External Current Account; FX reserves growth; FX reserves as share of GDP; Openness; Primary balance gap (share of GDP from VEE); Real GDP growth; Overall fiscal balance; Primary balance; FX debt (GG from VEE)
Cerovic, Gerling, Hodge and Medas (2018)	Fiscal crises	188 countries, 1970–2015 (2007–15 out of sample)	Signals and logit	Logit results (in sample): For AE/EM: Current Account, Primary expenditures growth, Output gap, interest expenses/revenue, public debt/revenue. For LICs: FDI, World food prices, GDP growth.
Ghulam and Derber (2018)	External sovereign default	70 countries 1970–2010	Survival analysis (general logit/specified logit incl. significant increase in hazard up to 15 years)	Volatilities of US treasury bills rates and USD-denominated LIBOR; Political uncertainty; Export/Import growth; Increase in inflation; Debt/GDP; Increase in external; GDP per capita; Previous banking; US treasury rate; Central government debt/GDP; High current account deficit and exchange rate volatility.

Notes: AEs: Advanced Economies; EMEs: Emerging Market Economies; LICs: Low-income Countries.

## APPENDIX 2. FISCAL CRISIS: DEFINITIONS AND DATA SOURCES

Event	Criterion	Thresholds			Literature	Sources	Notes
		Minimum two years gap between crises	AEs	EMEs	LIDCs		
(1)	<b>Credit Event</b>	<b>Default, restructuring, or rescheduling</b>					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.
		(i) of substantial size (in percent of GDP p.a.); AND  (ii) defaulted nominal amount grows by a substantial amount (in percent p.a)		> 0.5  ≥ 10	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	
(2)	<b>Exceptionally large official financing</b>	(i) <b>High-access IMF financial arrangement</b> with fiscal adjustment objective in place (in percent of quota); OR		≥ 100	Baldacci and others (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) <b>EU program</b>					
(3)	<b>Implicit domestic public default</b>	(i) <b>High inflation rate</b> (in pct. of growth of annual average CPI p.a.) OR	≥ 35	≥ 100	Baldacci and others (2011); Sturzenegger and Zettelmeyer (2006); and Fisher, Sahya and Vegh (2002)	IMF (World Economic Outlook)	
		(ii) <b>Steep increase in domestic arrears</b> (in first difference of the ratio of 'other account payables (OAP)' to GDP in percentage points)		≥ 1	Checherita-Westphal, Klem, and Viefers (2015); Reinhart and Rogoff (2011a)	Eurostat; OECD (data on other accounts payable)	
(4)	<b>Loss of market confidence</b>	(i) <b>High price of market access</b> (in basis points of sovereign spreads or CDS spreads) OR					
		(a) Level of spreads (bps)		≥ 1,000 bps	Sy (2004); Baldacci and others (2011)	Reuters Datastream; Bloomberg	
		(b) Annual change in spreads (bps)	≥ 300	≥ 650	na		
		(ii) <b>Loss of market access</b>		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 and others (2017)	Guscina, Sheheryar, and Papaioannou (2017); Gelos, Sahay, and Sandleris (2004);	

## APPENDIX 3. SAMPLE OF COUNTRIES

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

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>Public Debt</b>			
Public debt in percent of GDP <sup>1/</sup>	IMF, <i>Global Debt Database</i> (Mbaye, Moreno Badia, and Chae 2018a).	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, L_wavg, L2_wavg, 10yr_L_wavg
Public debt in percent of general government revenue <sup>1/</sup>	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L, fd_L2, L2_wavg
General government short-term external debt in percent of GDP <sup>2/</sup>	Authors' calculations based on IMF, <i>World Economic Outlook</i> (various vintages); IMF desk data; and BIS Debt Securities Data.	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L
General government short-term external debt in percent of reserves <sup>2/</sup>	Authors' calculations based on IMF <i>World Economic Outlook</i> (various vintages); IMF desk data; IMF, <i>International Financial Statistics</i> ; and BIS Debt Securities Data. Name: International Liquidity, Total Reserves excluding gold. Series Code: RAFA_USD.A.	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L, fd_L2
Public external debt in percent of GDP <sup>2/</sup>	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> ; US Bureau of Economic Analysis; Haver Analytics; Arslanalp and Tsuda (2014); IMF, <i>Quarterly External Debt Statistics</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, L_wavg, L2_wavg, 10yr_L_wavg
Public external debt to export <sup>2/</sup>	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> ; US bureau of economic analysis; Haver Analytics; Arslanalp and Tsuda (2014); IMF, <i>Quarterly External Debt Statistics</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, fd_L, fd_L2, d3_L
<b>Private Debt</b>			
One sided credit gap based on the GDD loans and securities	IMF, <i>Global Debt Database</i> (Mbaye, Moreno Badia, Mbaye, and Chae 2018a).	L, L2, fd_L, fd_L2, wavg	
Total Debt, loans and securities, in percent of GDP	IMF, <i>Global Debt Database</i> (Mbaye, Moreno Badia, Mbaye, and Chae 2018a).	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, 5yr_L, 10yr_L
10-year Average credit gap: current level of debt – 10-year average level	IMF, <i>Global Debt Database</i> (Mbaye, Moreno Badia, Mbaye, and Chae 2018a).	L, wavg	L, wavg

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
Domestic credit to private sector by banks, percent of GDP	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L, wavg	L, L2, fd_L, d3_L, 5yr_L, 10yr_L
External debt stocks, private nonguaranteed, percent of GDP	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L
Broad Money, percent of GDP	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L, wavg	L, L2, 10yr_L, L_wavg, L2_wavg
<b>External Debt<sup>3/</sup></b>			
Short-term external debt in percent of GDP <sup>3/</sup>	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L, fd_L2
Short-term external debt in percent of reserves <sup>3/</sup>	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> —Name: International Liquidity, Total Reserves excluding gold. Series Code: RAFA_USD.A.	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L, fd_L2
Total external debt, percent of GDP in USD	World Bank, <i>World Development Indicators</i> ; Lane and Milesi-Ferretti (2007, 2017).	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg
Total external debt, percent of export <sup>3/</sup>	World Bank, <i>World Development Indicators</i> ; Lane and Milesi-Ferretti (2007, 2017).	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L
<b>Total Debt</b>			
Total Debt (public plus private debt), percent of GDP	IMF, <i>Global Debt Database</i> (Mbaye, Moreno Badia, and Chae 2018a).	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, L_wavg, L2_wavg
<b>Public Debt Service</b>			
General government interest expenses in percent of GDP	IMF, <i>World Economic Outlook</i> . Medas et al (2018). Abbas et al (2011).	L, L2, fd_L, fd_L2, 5yr_L, d3_L, wavg	L, L2, fd_L, d3_L
Amortization of external public debt in percent of GDP	IMF, <i>World Economic Outlook</i> ; <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L
Amortization of external public debt in percent of reserves	IMF, <i>World Economic Outlook</i> and <i>International Financial Statistics</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, d3_L
Public debt service in percent of revenue	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, fd_L, fd_L2, d3_L, L2_wavg

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>External Debt Service</b>			
Public debt service in percent of export	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, fd_L, fd_L2, d3_L
External gross financing needs	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L	L, L2, fd_L, fd_L2, d3_L
Debt service on total external debt, percent of GDP	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L
Debt service on total external debt, percent of export	IMF, <i>World Economic Outlook</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L
Debt service on total external debt, percent of reserves	IMF, <i>International Financial Statistics</i> ; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L
<b>Fiscal</b>			
General government expenditures in percent of GDP	IMF, <i>World Economic Outlook</i> ; Medas et al (2018); Abbas and others (2011).	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
General government primary expenditures in percent of GDP	IMF, <i>World Economic Outlook</i> ; Medas et al (2018); Abbas and others (2011).	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2
Overall balance in percent of GDP	IMF, <i>World Economic Outlook</i> ; Medas et al (2018); Abbas and others (2011).	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
General government primary balance, percent of GDP	IMF, <i>World Economic Outlook</i> ; Medas et al (2018); Abbas and others (2011); Mauro et al (2011).	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2
General government revenues in percent of GDP	IMF, <i>World Economic Outlook</i> ; Medas et al (2018); Abbas and others (2011).	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2
Stock and flow adjustment	Authors' calculations based on IMF, <i>World Economic Outlook</i>	L, L2, fd_L, fd_L2, mean_L, wavg	L, L2, fd_L2, mean_L
Standard deviation of the stock and flow adjustment	Authors' calculations based on IMF, <i>World Economic Outlook</i>	L	L
<b>R-G</b>			
R minus G	Authors' calculations based on IMF, <i>World Economic Outlook</i> , and <i>Global Debt Database</i> (Mbaye, Moreno Badia, and Chae 2018a)	L, L2, fd_L, 5yr_L, 10yr_L, wavg	L, L2, 5yr_L, 10yr_L

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>Economic Activity</b>			
Percent change in Material impact: all natural disaster hazards. In National Currency, Million	CRED, 's The International Disaster Database (EM-DAT available at <a href="https://www.emdat.be/">https://www.emdat.be/</a> )	L, L2	
Percent change of real GDP per capita	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L	L, L2, pc3_L
Percent change in real GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L	L, L2
Growth deviation from past 5-year average	IMF, <i>World Economic Outlook</i> ; Medas et al (2018).	L, wavg	L, L_wavg
Percent change in nominal GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L	L, L2, fd_L
GDP growth rate relative to the past 5-year average growth rate	IMF, <i>World Economic Outlook</i> .	L	L
Geometric average of the last 3-year GDP growth	IMF, <i>World Economic Outlook</i> .	L	L
domestic savings, private (current US\$)	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L, mean_L, wavg	L, L2, d3_L, mean_L, mean_L_wavg
Standard deviation of real GDP growth	IMF, <i>World Economic Outlook</i> .	L	L
<b>Inflation</b>			
Percent change of Consumer Price Index, period average	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L, wavg	L, L2, fd_L, pc3_L
Percent change of Consumer Price Index, end of period Units	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L, wavg	L, L2, fd_L, pc3_L
Standard deviation of the inflation	IMF, <i>World Economic Outlook</i> .	L	L
<b>External</b>			
Personal remittances in percent of GDP	World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, fd_L, fd_L2, d3_L
Net foreign direct investment in percent of GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
Other investment, net (loans, deposits, insurance, pensions, trade credits, SDR), percent of GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2



#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
Portfolio investment, net	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	
Percent change in terms of trade (of goods and services) Index	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L	L, L2, pc3_L
Percent change in trading partner growth	IMF, <i>Global Economic Environment</i> (internal database).	L, L2, fd_L, pc3_L	L, pc3_L
Percent change of trading Partner Import Demand	IMF, <i>Global Economic Environment</i> (internal database).	L, L2, fd_L, pc3_L	pc3_L
Standard deviation of the change in terms of trade	IMF, <i>World Economic Outlook</i> .	L	L
External assets in percent of GDP	IMF, <i>World Economic Outlook</i> ; Lane and Milesi-Ferretti (2007, 2017)	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, 5yr_L, 10yr_L
<b>Current Account</b>			
Current account balance in percent of GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L	L, L2
Export of goods and services in percent of GDP	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
Import of goods and services in percent of GDP	IMF, <i>World Economic Outlook</i>	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
Current account excluding imports	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, d3_L	L, L2, d3_L
Average of the last 10 year of the sum of export and import of goods and services	IMF, <i>World Economic Outlook</i> .	L, wavg	L, L_wavg
<b>Exchange Rate</b>			
Percent change of exchange rate (NC units per U.S. dollar, period average)	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L	L, L2, fd_L, pc3_L
Percent change of exchange rate (NC units per U.S. dollar, end of period)	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L	L, pc3_L

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
Percent change in real exchange rate, period average	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, pc3_L	L, L2, fd_L, pc3_L
Log of PPP exchange rate	World Bank, <i>International Comparison Program</i> database.	L, L2, wavg	L, L2
PPP overvaluation (relative to US)	IMF, <i>World Economic Outlook</i> .	L, wavg	L
Standard deviation of the percent change in exchange rate (period average)	IMF, <i>World Economic Outlook</i> .	L	L
Standard deviation of the percent change in exchange rate (end of period)	IMF, <i>World Economic Outlook</i> .	L	L
<b>FX Reserves</b>			
Percent change in total reserves excluding gold in dollars	IMF, <i>International Financial Statistics</i> . Name: International Liquidity, Total Reserves excluding gold. Series Code: RAFA_USD.A.	L, L2, fd_L, pc3_L, wavg	L, L2, fd_L, pc3_L
Total reserves in terms of the number of months of imports	IMF, <i>International Financial Statistics</i> and desk data files; World Bank, <i>World Development Indicators</i> .	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, d3_L
<b>Foreign Aid</b>			
Net official development assistance in percent of GDP	Development Assistance Committee of the Organization for Economic Co-operation and Development, Geographical Distribution of Financial Flows to Developing Countries, Development Co-operation Report, and International Development Statistics database.	L, L2, fd_L, fd_L2, d3_L, wavg	L, L2, fd_L, fd_L2, d3_L
<b>Global</b>			
Percent change of crude oil price	IMF, <i>World Economic Outlook</i> .	L, L2, pc3_L, fd_L	
Percent change of Non-fuel price	IMF, <i>Primary Commodity Prices</i> (based on monthly data; average of the year); Medas et al (2018).	L, L2, pc3_L, fd_L	

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
Percent change of food price	IMF, <i>Primary Commodity Prices</i> (based on monthly data; average of the year); Medas et al (2018).	L, L2, pc3_L, fd_L	
US T-Bill rate Percent	IMF, <i>International Financial Statistics</i> .	L, L2, fd_L, fd_L2	L, L2
VIX Index Period Average	Bloomberg Finance L.P.	L, L2	
VIX Index Period End	Bloomberg Finance L.P.	L, L2	
Percent change of VIX Index (Period Average)	Bloomberg Finance L.P.	L, L2, pc3_L, fd_L	
Percent change of VIX Index (End of Period)	Bloomberg Finance L.P.	L, L2, pc3_L, fd_L	
US T-Note 5-year rate Percent, Period Average	Bloomberg Finance L.P.	L, L2, fd_L, fd_L2	L, L2
US T-Note 10-year rate Percent, Period Average	Bloomberg Finance L.P.	L, L2, fd_L, fd_L2	L, L2
US T-Note 5-year rate Percent, End of Period	Bloomberg Finance L.P.	L, L2, fd_L, fd_L2	
US T-Note 10-year rate Percent, End of Period	Bloomberg Finance L.P.	L, L2, fd_L, fd_L2	
World real GDP growth, in percent	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L	
Geometric average of the last 3-year world GDP growth	IMF, <i>World Economic Outlook</i> .	L	

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>Crisis History</b>			
Average crisis history	Authors' calculations based on Medas et al (2018).		Average crisis history
Number of crisis			
Number of crisis by year, Advanced and Emerging Economies (AE+EM)			
Number of crisis by year, Emerging and Low-Income Economies (EM+LIC)			
Number of crisis by year, AEs			
Number of crisis by year, EMEs			
Number of crisis by year, LICs			
Number of crisis by year (sum of the current and previous year)			
Number of crisis by year (sum of the current and previous year), AE+EME			
Number of crisis by year (sum of the current and previous year), EME+LIC			
Number of crisis by year (sum of the current and previous year), AE			
Number of crisis by year (sum of the current and previous year), EME			
Number of crisis by year (sum of the current and previous year), LIC	Authors' calculations based on Medas et al (2018).		
Number of crisis by country			
Number of crisis by country, AE+EME			
Number of crisis by country, EME+LIC			
Number of crisis by country, AE			
Number of crisis by country, EM			
Number of crisis by country, LIC			
Year since last crisis			Year since last crisis

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
Banking crisis start year	Authors' calculations based on Laeven and Valencia (2018).	L, L2	
Financial Crisis			
Financial Crisis, AE+EM			
Financial Crisis, EM+LIC			
Financial Crisis, AE			
Financial Crisis, EME			
Financial Crisis, LIC			
<b>Level of development</b>			
Urban population (% of total)	World Bank, <i>World Development Indicators</i>	L, 5yr_L, 10yr_L, wavg	L, 5yr_L, 10yr_L
Log of real GDP per capita (in PPP dollars, Units), relative to US	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, wavg	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, L_wavg, L2_wavg
Log of nominal GDP in USD, relative to US	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, wavg	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, L_wavg, L2_wavg, 5yr_L_wavg
<b>Institutions, elections</b>			
Revised Combined Polity Score (single regime score, runs from 1 (full democracy) to -1 (full autocracy))	Center for Systemic Peace	L, L2, fd_L, fd_L2, 5yr_L, 10yr_L, wavg	L, L2, 5yr_L, 10yr_L
Checks and balances index	Data base on Political Institutions (DPI) 2015.	L, 5yr_L, 10yr_L, wavg	L
Bureaucracy Quality	PRS Group.	L, 5yr_L, 10yr_L	L, 5yr_L, 10yr_L
Corruption	PRS Group.	L, 5yr_L, 10yr_L	L, 5yr_L, 10yr_L
Years remaining in current chief executive's term	Data base on Political Institutions (DPI) 2015.	L	
Legislative election held dummy	Cruz, Keefer and Scartascini (2018).	L	
Executive election held dummy variable	Cruz, Keefer and Scartascini (2018).	L	
Political Stability and Absence of Violence/Terrorism: Estimate	World Bank, <i>World Development Indicators</i> .	L, 5yr_L, 10yr_L	L, 5yr_L, 10yr_L
Regulatory Quality: Estimate	World Bank, <i>World Development Indicators</i> .	L, 5yr_L, 10yr_L	L, 5yr_L, 10yr_L

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>Demographics</b>			
Population ages 15-64, total	World Bank, <i>World Development Indicators</i> .	L, 5yr_L, 10yr_L, wavg	L, 5yr_L, 10yr_L, 5yr_L_wavg, 10_yr_L_wavg
Percent change of population ages 15-64, total	World Bank, <i>World Development Indicators</i> .	L, wavg	L
Age Dependency Ratio, % of working-age population	World Bank, <i>World Development Indicators</i> .	L, 5yr_L, 10yr_L, wavg	5yr_L, 10yr_L, 5yr_L_wavg
Population density (people per sq. km of land area)	World Bank, <i>World Development Indicators</i> .	L, wavg	L
Log of population relative to US	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, wavg	L, L2, fd_L, d3_L, 5yr_L, 10yr_L, L_wavg, L2_wavg
<b>Natural Resources</b>			
Dummy: Fuel exporter	IMF, <i>World Economic Outlook</i> .		
Dummy: Fuel exporter or VELIC commodity exporter	IMF		
Value of oil export, % of GDP in USD	IMF, <i>World Economic Outlook</i> .	L, L2, fd_L, fd_L2, wavg	L, L2, fd_L, fd_L2
Mineral rents (% of GDP)	World Bank, <i>World Development Indicators</i> .	L, L2, 5yr_L, 10yr_L, mean_L, wavg	L, L2, 5yr_L, 10yr_L, mean_L
Oil rent (% of GDP)	World Bank, <i>World Development Indicators</i> .	L, L2, 5yr_L, 10yr_L, mean_L, wavg	L, L2, 5yr_L, 10yr_L, mean_L
Total natural resources rent (% of GDP)	World Bank, <i>World Development Indicators</i> .	L, L2, 5yr_L, 10yr_L, mean_L, wavg	L, L2, 5yr_L, 10yr_L, mean_L
Agriculture, forestry, and fishing, value added (% of GDP)	World Bank, <i>World Development Indicators</i> .	L, wavg	L

#### APPENDIX 4. DATA: DEFINITION, SOURCES, AND PREDICTOR GROUPINGS

Variable	Source	Permutation	Boruta Selected Variables
<b>Country Category</b>			
Dummy: Monetary Union	IMF, <i>World Economic Outlook</i> .	L	L
Dummy: Island country	Wikipedia		Dummy: Island country
Dummy: Landlocked country	CIA, <i>World Factbook</i> .		
Dummy: Small state	Authors' calculations based on IMF, <i>World Economic Outlook</i> and World Bank, <i>World Development Indicators</i> .		Dummy: Small state

Note: L = lag; L2 = second lag; fd\_L = lag of first difference; fd\_L2 = second lag of first difference; 5yr\_L = lag of the trailing 5 year difference; 10yr\_L = lag of the trailing 10 year difference; d3\_L = lag of the trailing 3 year difference; pc3\_L = lag of percentage change over trailing 3 year; mean\_L = lag of trailing 10 year moving average; wavg = cross sectional weighted average for all the permutations

1/ Public debt includes total debt liabilities of the government with domestic and foreign creditors. In compiling public debt series for each country, we look at the different perimeters of government (non-financial public sector, general government, and central government) for which the Global Debt Database reports data, choosing the debt category for which the time series is the longest. In many cases, particularly, among LICs, this results in a narrow definition of debt (central government) but ensures the consistency of the series across time. In contrast, previous studies have often used a hybrid approach to compile debt statistics, switching debt concepts depending on availability which may have yielded longer but inconsistent time series.

2/ Public external debt is defined in terms of the residency of holder. It includes general government debt and debt guaranteed by the government and, as such, it may have a wider sectoral coverage than our measure of total public debt. An attempt was made to construct alternative measures based on currency-denomination, but time series availability was limited for most of the sample and that metric was excluded from the analysis. Equal constraints applied to other debt characteristics such as the type of holder (e.g. banks, official sector) or the maturity of debt.

3/ External debt includes total debt liabilities of a country (both for the government and private sector) with foreign creditors.

## APPENDIX 5. METHODOLOGICAL DETAILS

### Empirical model

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 Bussi 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.

To estimate the probability of a crisis, we rely on a RF—an ensemble learning method based on decision trees. Each decision tree 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  $m_{try}$  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.

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). The tuning parameter  $m_{try}$  is chosen from a grid of candidate values through cross-validation.<sup>12</sup> 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.

### Sample splitting

Models were evaluated based on 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 evaluation). To avoid possible information spillovers from the test to the training sample, we use a rolling cutoff year (beginning in 2000) to separate between the two. That is, we start by estimating a model with data for 1980–2000 and then roll forward both the estimation and the testing periods, adding one year at a time in each iteration. Therefore, we end up estimating 15 models, each one based on a larger training sample than the previous one with the hyperparameter retuned in each round.

### Hyperparameter tuning

For each of the 15 training samples, we use k-fold cross validation to choose the optimal  $m_{try}$ , where  $k$  is the number of years of that training sample. k-fold cross validation simulates out-of-sample prediction. We choose the  $m_{try}$  to minimize out-of-sample log-likelihood loss. The

---

<sup>12</sup> The same procedure for hyperparameter tuning is followed for the variable selection algorithms.



tuning length parameter is set to 10 (i.e. we do a blind search 10 times to search for the optimal  $m_{try}$ ). Our procedure is as follows:

1. First, the training sample is partitioned into  $k$  equal sized subsamples. A model is then fit to  $k-1$  subsamples and used to predict results for the  $k$ th. This is repeated for each of the  $k$  subsamples, so that there is an “out-of-sample” prediction of each observation. The best performing  $m_{try}$ , in terms of log-likelihood loss function, is chosen.
2. Using the selected tuning parameter values, we fit the model to the entire training set.
3. With this fitted model, we produce predicted probability of a crisis for the corresponding testing set.

### Evaluation measures

In machine-learning classification models, a standard measure of model accuracy is the AUC. One of its advantages is that it does not require that we specify a probability threshold above which we predict a fiscal crisis will occur. However, the AUC has some limitations, especially in the presence of class-imbalanced data (as it is our case, where crisis episodes are relatively low-frequency events). Thus, we also report two alternative measures commonly used in the literature, the mean squared error and the log-likelihood (see, for example, Fuertes and Kalotychou 2006; Dawood et al. 2017). Specifically:

- The mean squared error (MSE) calculated as:

$$\frac{1}{N} [\sum_{i \in I_{crisis}} (1 - p_i)^2 + \sum_{i \in I_{non-crisis}} p_i^2], \text{ and}$$

- The log-likelihood calculated as

$$\frac{1}{N} [\sum_{i \in I_{crisis}} \log(p_i) + \sum_{i \in I_{non-crisis}} \log(1 - p_i)]$$

where  $p_i$  denotes the predicted probability of crisis for observation  $i$ ,  $I_{crisis}$  and  $I_{non-crisis}$  denote the sets of crisis starts and non-crisis observations in the test sample, and  $N$  is the number of observations in the test sample. These two metrics differ on how they penalize false classifications (MSE will have a value of 1 if our model misses entirely all crises in the test sample, while the log-likelihood will be  $-\infty$ ). All evaluation measures are computed as the average of each metric for the 15 test samples.

### Comparison of Out-of-sample Performance across Prediction Models

The theoretical literature suggests that different models work well for different prediction problems—what is known in the machine learning literature as “the no free lunch theorem” (Wolpert and Macready 1997). Nonetheless, a large scale empirical study found that RF is

the best performer across a broad set of prediction tasks (Fernandez-Delgado et al. 2014) and the early warning literature has recently found similar results (Bluwstein et al. 2019). To assess relative performance in our particular setting (i.e. forecasting fiscal crises), we evaluate the RF (using the original 748 variables) against two other popular machine learning algorithms: LASSO (Tibshirani 1996) and Support Vector Machine (SVM, Cortes and Vapnik 1995).

- *LASSO* is a shrinkage and selection method for linear regression that minimize the usual sum of squared errors with a bound of the sum of the absolute values of the coefficients. For all practical purposes, it is just a logistic regression model (like the standard one used in the early warning literature), the difference being that it penalizes large coefficients and forces all but the most important ones to zero. Thus, it is less susceptible to overfitting although is not well suited to identify complex interactions.
- *SVM* is a popular machine learning classification algorithm for small and medium-sized datasets. After performing non-linear transformation of the features, the algorithm estimates and separating hyperplane (in our particular case discriminating between crisis and non-crisis events).

While there is a large variety of additional algorithms we might have tested, we focus on these because they are well established in the machine learning literature, and they are commonly used for forecasting in economics. On the basis of the three metrics we use to evaluate out-of-sample performance, we find that the RF does always as well or better than the other two algorithms (Table A5.1). In terms of the AUROC, one of the most commonly used metrics, RF is clearly superior than both LASSO and the SVM for advanced and emerging market economies. For other metrics, RF is broadly better in all income groups although differences are not statistically significant.

**Appendix Table 5.1. Out-of-Sample Performance**

	RF	Lasso	SVM
<b><i>Advanced and emerging market economies</i></b>			
AUROC	0.806 (0.026)	0.763** (0.03)	0.767*** (0.025)
Log(likelihood)	-0.299 (0.026)	-0.299 (0.036)	-0.295 (0.032)
MSE	0.087 (0.01)	0.089 (0.013)	0.087 (0.012)
<b><i>Low-income countries</i></b>			
AUROC	0.706 (0.043)	0.676 (0.043)	0.682 (0.04)
Log(likelihood)	-0.556 (0.039)	-0.581 (0.052)	-0.587 (0.056)
MSE	0.189 (0.017)	0.2* (0.021)	0.201 (0.022)

Note: Bootstrapped standard deviations (based on 100 random resamples of the test sample with replacement) in parentheses. Stars indicate degree of confidence that a model outperforms the Random Forest estimated with the full set of 748 variables: \* 90%, \*\*=95% , and \*\*\*=99%. Models predict probability of crisis start occurring in year t+1 or t+2. Out-of-sample performance obtained from 15 rolling regressions.

### 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  $X^C$ .

$$f^S(X^S) = E_C[f(X^S, X^C)] = \int f(X^S, X^C) p_C(X^C) dX^C$$

Where  $p_C(X^C)$  is the marginal probability of  $X^C$ . PDP work 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.

$$f^S(X^S) \approx \frac{1}{N} \sum_{i=1}^N f(X^S, X_i^C)$$

Where  $N$  is the number of observations and  $X_i = (X_i^S, X_i^C)$  in the  $i$ th observation.  $f^S(X^S)$  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.

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