# Predicting Debt Distress in Low-Income Countries

Clemens Graf von Luckner

Sebastian Horn

Aart Kraay

Rita Ramalho[1]

## Abstract

We develop an empirical model to predict episodes of debt servicing difficulties ("debt distress") in low-income countries, with three main contributions to the existing literature: First, we develop refined measures of external debt distress episodes that allow us to time the onset of distress episodes with increased precision. Second, we develop a systematic algorithm to comprehensively assess the out-of-sample predictive performance of more than 550,000 candidate binary prediction models based on J-K-fold cross-validation. Third, we test whether more sophisticated machine learning algorithms can outperform simple probit models. We find that simple single-equation models have better predictive power for debt distress than more sophisticated algorithms and are comparable to important policy benchmarks such as the current IMF and World Bank debt sustainability framework for low-income countries.

Keywords: sovereign debt, default, financial crisis, prediction, low-income countries

JEL codes: F34, F21, H63, C5, O10

---

[1] We thank Luca Bandiera, David Mihalyi, Carmen Reinhart, and seminar participants at DebtCon6 Princeton, the IMF and the World Bank for helpful comments and suggestions. We are grateful to Evis Rucaj and Rubena Sukaj of the World Bank Debtor Reporting System team for assistance with debt service data. The views expressed herein are those of the authors and do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent. Graf von Luckner: Sciences Po Paris and World Bank, cgrafvonluckner@worldbank.org; Horn: World Bank, shorn1@worldbank.org; Kraay: World Bank, akraay@worldbank.org; Ramalho: World Bank, rramalho@ifc.org

## 1. Introduction

The ability to predict and prevent sovereign debt crises is critical for economic development, as sovereign default episodes are associated with severe economic and social costs, including lower growth and productivity, increased poverty levels, and reduced foreign investment (Reinhart and Rogoff 2009, Aguiar and Amador 2021, Farah-Yacoub et al. 2024). It is therefore no surprise that an extensive academic literature has studied the drivers and correlates of sovereign default episodes and has attempted to build early warning systems (see Badia et al. 2022 for a comprehensive literature review).

This paper develops an empirical model to predict episodes of external sovereign debt servicing difficulties ("debt distress") in low-income countries (LICs). Three main features distinguish our approach from the existing literature. First, we propose refined and simplified measures of external debt distress episodes. Identifying debt distress episodes in low-income countries is challenging due to a lack of sovereign risk rating and financial market data. Much of the existing literature has therefore relied on sovereign debt restructurings and ad-hoc timing assumptions along with other signals to construct debt distress episodes (Kraay and Nehru 2006, Rodriguez and Rodriguez 2006, IMF and World Bank 2017). Debt restructurings, however, often mark the (attempted) resolution of debt distress rather than its onset, diluting the information in the outcome variable in distress prediction models (Das et al. 2012). Instead of relying on restructurings, we make use of novel and comprehensive datasets of direct payment defaults on private creditors in all low-income countries since the 1960s that allow us to time the onset of default events with increased precision (Asonuma and Trebesch 2016, Farah-Yacoub et al. 2023).

Second, we develop a systematic algorithm to comprehensively assess the out-of-sample predictive performance of all relevant combinations of explanatory variables in our dataset (resulting in more than 550,000 binary prediction models). Most of the existing academic literature relies on the ad-hoc selection of predictor variables and measures predictive accuracy in-sample (see e.g., Detragiache and Spilimbergo 2001, IMF and World Bank 2017). In contrast, our paper relies on a “brute force” comparison approach that first identifies candidate predictor variables from the academic literature and then estimates probit models with all relevant combinations of predictors. Predictive performance is assessed out-of-sample through J-K-fold cross-validation. Our algorithm provides a general empirical framework to identify top performing models, to assess trade-offs between model parsimony and predictive accuracy, and to study the relative importance of different variables for predicting debt distress.

Third, we use different estimation strategies and test whether more sophisticated machine learning algorithms can outperform the workhorse probit and logit models on which most of the existing literature relies. Specifically, we use a Random Forest classifier, one of the most commonly applied machine learning classifiers (Badia et al. 2022). Our results show that the best Random Forest models perform considerably worse than the best probit models. This suggests that more sophisticated prediction algorithms add little value in our setting with a comparatively small dataset and a limited number of predictor variables.

A key contribution of our paper is to systematically identify models with strong predictive performance. While there is a trade-off between model parsimony and predictive accuracy, even comparatively simple binary prediction models with five predictor variables or fewer can achieve strong predictive performance and outperform much more complex models. For example, a very simple model with just three variables (institutional strength, external debt service as a share of exports, and reserve holdings as a share of imports) generates very good predictive performance relative to most alternative models.

In addition to contributing to the academic literature on early warning systems of debt crises, our paper is motivated by the 2024 review of the IMF and World Bank debt sustainability framework for low-income countries (the “LIC DSF”). The LIC DSF is used by the IMF and the World Bank to assess the debt sustainability of low-income countries and serves as an early warning tool to inform the institutions' lending policies, country surveillance work and policy advice. At the heart of the LIC DSF is an empirical model that links the risk of debt distress to country characteristics and various debt burden indicators. Our research improves on this empirical model in several ways. In addition to proposing a refined outcome measure of debt distress and a systematic approach to assessing the out-of-sample predictive performance of candidate prediction models, we show that considerably simpler models can deliver predictive power comparable to that of the existing LIC DSF prediction model, and may be more amenable to policy use because of their simplicity.

In line with much of the existing academic literature, our paper first and foremost attempts to predict episodes of *external* debt distress (e.g., Manasse et al. 2003, Kraay and Nehru 2006, Manasse and Roubini 2009). At the same time, there is growing recognition that episodes of *domestic* debt distress in low-income countries have become more prevalent over time as sovereign domestic debt has expanded significantly in some LICs (IMF 2021a). This trend poses considerable measurement challenges given that systematic data on domestic debt, debt service and default are still scarce. We address this issue in two ways: First, we supplement our measure of external debt distress episodes with data on domestic defaults. We show that external and domestic distress episodes are highly correlated, suggesting that models that predict external debt distress with high accuracy are also suitable for the prediction of total public debt distress episodes. Second, we include measures of the total public debt-to-GDP ratio and total public debt service in our model selection algorithm to determine whether domestic debt or debt service is an important predictor of debt distress. While conceptually total public debt accumulation should matter for the risk of debt distress through the integrated government budget constraint, in practice we find that domestic debt adds little valuable information in predicting debt distress episodes. This finding suggests two non-exclusive interpretations: (a) sovereigns may have a wider range of options, including monetizing deficits and financial repression, to avoid outright default on domestic debt; and/or (b) our measures of domestic debt service are too noisy to identify a systematic relationship with debt distress episodes. Both interpretations suggest the need for better measurement, both of implicit defaults on domestic debt (to address (a)) and of domestic debt stocks and debt service (to address (b)). Finally, it is important to acknowledge that for much of the historical sample period since 1970 on which we estimate and evaluate prediction models, domestic debt was comparably small for most LICs. We therefore cannot rule out the possibility that the weak historical relationship between domestic debt and debt distress will change in the future as domestic debt grows in importance.

This paper proceeds as follows. Section 2 discusses the underlying data and explains how we construct measures of sovereign debt distress. Section 3 introduces our main prediction framework and the algorithm that we use to select predictor variables and top performing probit models. Section Section 4 presents additional empirical exercises based on alternative prediction algorithms, as well as for medium-term prediction horizons. Finally, Section 5 provides a systematic comparison of our preferred prediction model with the existing LIC DSF model used by the IMF and the World Bank. The Appendix provides additional details on the underlying data and presents selected robustness tests and additional results.

## 2. Data

This section explains how we define external sovereign debt distress episodes and introduces the variables that we use to predict its occurrence. Appendix A provides more detailed information on data sources, variable transformations, and sample coverage. We note at the outset that our sample of interest consists of 80 countries for which the IMF and the World Bank apply the LIC DSF to assess debt sustainability. This in turn consists of the set of countries eligible for concessional lending from the World Bank. This includes all World Bank client countries classified as low-income, as well as a number of mostly lower-middle-income countries that also are eligible for concessional lending from the World Bank. For terminological convenience, we refer to this group as “low-income countries”. For these countries, we consider all years with available data between 1970 and 2021. For additional details see Appendix A1.

### 2.1 Measuring external sovereign debt distress

We define episodes of external debt distress as periods in which at least one of the following two conditions holds:

(i) The sovereign is in **default** on its external creditors, as evidenced by a direct default or by the substantial and persistent accumulation of interest or principal arrears.

(ii) The sovereign has received exceptionally **large balance of payments assistance** in the form of non-concessional and rapidly disbursed IMF lending.

This section discusses each of the two conditions in turn and introduces the resulting external debt distress variable.

### 2.1.1 Default

Default is defined as a sovereign's failure to meet its repayment obligations to its creditors on time and beyond the grace period (Das et al. 2012).

**Direct default measure for private creditors:** For private external creditors, our default signal builds on newly compiled data by Farah-Yacoub et al. (2023) that identifies external defaults on private bondholders and commercial banks, leveraging international and local financial news, rating agency reports, IMF and World Bank reports and a range of existing academic sources. In this dataset, sovereigns enter default through a missed payment on their external creditors and exit default by concluding a comprehensive debt restructuring with their creditors.

**Significant and persistent arrears:** Low-income countries borrow from a diverse set of official and private creditors, but systematic data on the incidence of defaults only exists for private external creditors (Horn et al. 2022). In line with most of the literature, we therefore supplement our direct default measure for private creditors with an arrears-based proxy to capture defaults to other creditors (e.g., Detragiache and Spilimbergo 2001, Kraay and Nehru 2006, Medas et al. 2018, Arellano et al. 2022). This is challenging because the relationship between arrears and default is not straightforward: not every instance of payment arrears implies that the sovereign is in default or is experiencing debt servicing difficulties. Temporary and limited arrears accumulation may be due to other reasons, such as low debt management capacity, or might even be a cost-efficient way of obtaining short-term finance in the face of underdeveloped capital markets.[2] Indeed, episodes with positive outstanding arrears are much more frequent than most indicators of debt distress or default imply (see Farah-Yacoub et al. 2022 for a detailed discussion).

With these considerations in mind, we follow the literature by restricting attention to large and persistent arrears accumulation as a signal of debt distress. Specifically, our arrears-based distress signal is equal to one if the sum of principal and interest arrears to all external creditors exceeds five percent of the public and publicly guaranteed external debt stock and remains above this threshold for at least three years (see e.g. Detragiache and Spilimbergo 2001, Kraay and Nehru 2006, and World Bank and IMF 2017).[3] Data on arrears and debt are taken from the World Bank's International Debt Statistics.

**Debt restructurings:** In contrast with most of the existing literature, we do not define debt restructurings with official or private creditors as separate distress signals.[4] Although sovereign defaults and sovereign debt restructurings are closely linked, they are not identical. A default is the failure of a sovereign to meet its repayment obligations to its creditors when they are due. In contrast, a sovereign debt restructuring is an exchange of an existing debt instrument for a newly issued debt instrument through a legal process, most often with net present value losses for the creditor (Das et al. 2012).

In most distress episodes, a default marks the onset of debt distress, while a debt restructuring marks the resolution (or an attempted resolution) of the distress episode (Asonuma and Trebesch 2016). Default and restructuring can be multiple years apart. For our debt distress prediction exercise, it

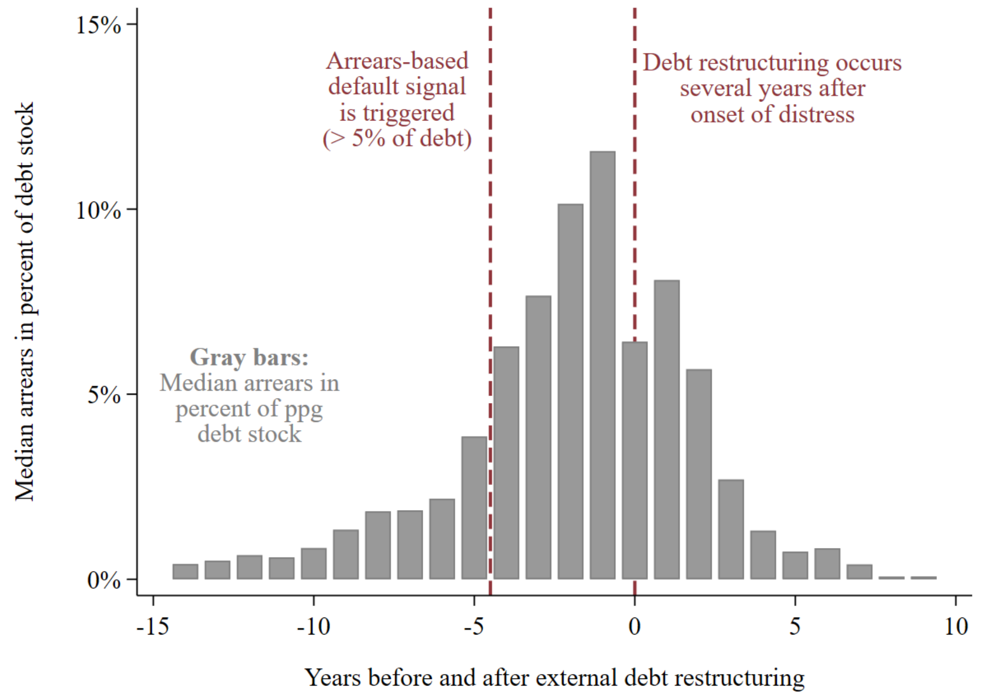
[2] This could be the case if a sovereign's marginal borrowing rate is higher than the late payment penalty rate specified in one of its loan contracts (see for example Gelpern et al. 2021).

[3] If the threshold is surpassed for at least three consecutive years, each of these years is treated as an external debt distress episode. The median arrears to debt level is around 1 percent in our sample and our threshold roughly corresponds to the upper third of the distribution.

[4] One exception are preemptive debt restructurings with net present value losses for creditors that we treat as direct default events and that are captured in the data on private external defaults compiled by Farah-Yacoub et al. (2023). See Asonuma and Trebesch (2016) for a detailed discussion of preemptive sovereign debt restructurings.

therefore is crucial to identify the initial default rather than the subsequent restructurings. Figure 1 illustrates this point by showing the accumulation and resolution of arrears around external debt restructurings with private and Paris Club creditors in our low-income country sample. In the median country and episode, arrears cross the 5 percent threshold four years before the restructuring.[5]

Figure 1. Build-up and resolution of arrears around external debt restructurings




*Note:* This figure shows the evolution of median principal and interest arrears in percent of the outstanding ppg debt stock for years before and after an external debt restructuring with private or Paris Club bilateral creditors. Data on arrears is from the World Bank's International Debt Statistics and data on restructurings is from Asonuma and Trebesch (2016) and from Horn et al. (2022).

### 2.1.2 Large IMF balance of payments assistance

Not all debt distress episodes end in default. In the past several decades, an increasing share of debt distress episodes in emerging markets and advanced countries have been resolved without any missed payments on external creditors (Pescatori and Sy 2007, Mitchener and Trebesch 2023). Many recent contributions have therefore identified debt distress episodes as periods in which sovereign

---

[5]A detailed analysis of external debt distress episodes and Paris Club debt restructurings underlines this point. In Appendix Section A2, we show that the incorporation of Paris Club debt restructurings into our external debt distress episode definition yields only few additional episodes. Furthermore, additional episodes lag rather than lead our definition of external debt distress episodes, thereby diluting the accuracy of the default signal.

bond spreads have exceeded specific thresholds (Aguiar et al. 2016, Dvorkin et al. 2020, Krishnamurthy and Muir 2017, Horn et al. 2020, IMF 2021b).

This approach is not feasible in low-income countries where many sovereigns do not borrow on commercial terms, and financial market data is available only for a small subset of countries and years. Instead, we rely on the incidence of large IMF programs to signal distress periods that do not involve outright default or large arrears accumulation. Specifically, we consider all years in which sovereigns draw on more than 30 percent of their quota during the first six months of a non-concessional IMF lending program to be a signal of debt distress.[6] This focus on large upfront disbursements follows IMF and World Bank (2017) and attempts to capture episodes of urgent financing needs that cannot be met by alternative sources.[7] Data on the use of IMF lending comes from the IMF's International Financial Statistics.

### 2.1.3 Defining external sovereign debt distress episodes

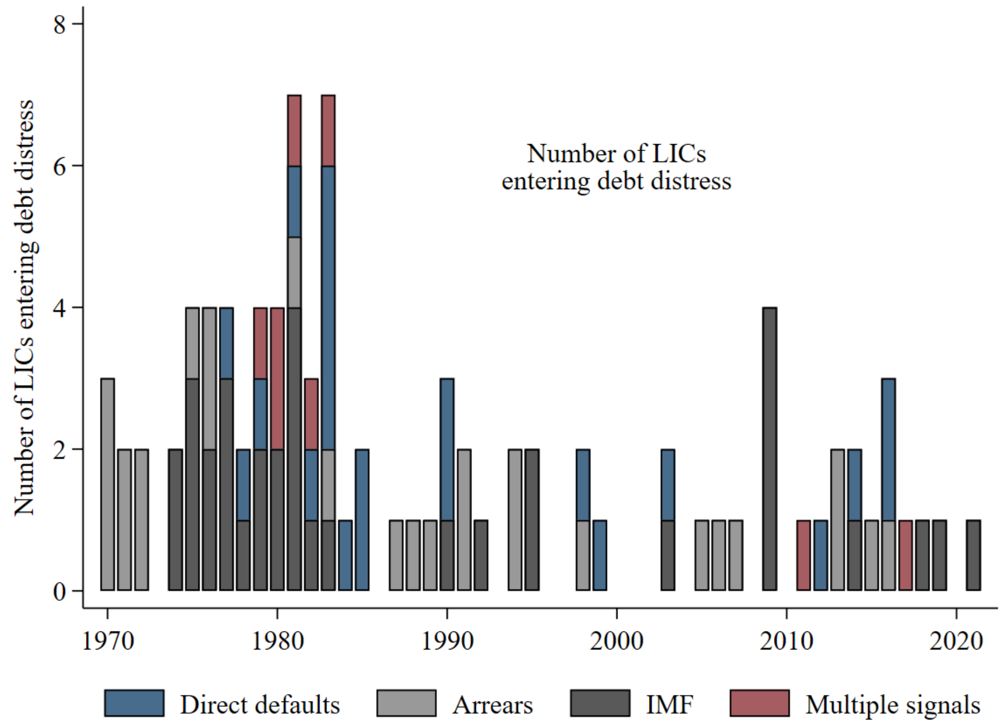
We define debt distress episodes as years in which one or more of the debt distress signals are observed. This definition results in 90 episodes of debt distress. Figure 2 shows that the three signals are roughly equally important for identifying the onset of debt distress episodes and that the relative importance of the signals has remained largely stable over time. In line with academic research, we identify fewer new distress episodes in the second half of our sample, i.e, after the mid-1990s (see Reinhart et al. 2016 for a discussion). Appendix Section A2 provides additional details on the

[6] In line with most contributions in the literature, we only consider non-concessional IMF lending programs and exclude rapid disbursements made under the IMF's Poverty Reduction and Growth Trust (PRGT). In many low-income countries, lending under the PRGT has been used for extended periods of time and is therefore not necessarily indicative of sovereign distress episodes (IMF Independent Evaluation Office 2002, Reinhart and Trebesch 2017). Furthermore, PRGT lending is extended at highly concessional interest rates and might give recipient countries a financial incentive to apply for these facilities, even if they have access to alternative (more expensive) sources of market financing and are not in debt distress. We further exclude rapid disbursements under the IMF's Rapid Financing Instrument, which often addresses funding needs of a limited or transitory nature and which has primarily been used by countries with strong fundamentals and at low or moderate risks of debt distress. The IMF distress signal is activated only for the year in which the rapid disbursement takes place.

[7] As for our arrears threshold, the 30 percent of quota cut-off value corresponds to roughly the upper third of the disbursement distribution in LICs during the first six months of a non-concessional program.

incidence and duration of debt distress by presenting external debt distress episodes by signal, year, and debtor country.

Figure 2. Debt distress episodes by signal and over time




*Note:* This figure shows the onset of debt distress episodes over time and by the signal, which initially triggered the debt distress episode. For a distress episode to be triggered by a given signal, the signal needs to be the only signal to indicate distress at the onset of the episode.

#### 2.1.4 Domestic debt distress

Thus far we have focused on constructing *external* debt distress episodes. This however ignores episodes of *domestic* debt servicing difficulties. Conceptually this is problematic, because governments of many countries in our sample borrow from both external and domestic creditors to finance budget deficits. Expanding the set of distress episodes to all instances of *public* debt distress is challenging, since comprehensive data on episodes of *domestic* debt servicing difficulties does not exist for low-income countries.

In this subsection we expand our debt distress measure by including domestic distress episodes and analyze how the identified distress patterns change. To measure domestic debt distress episodes, we build on the dataset of domestic debt restructurings in IMF (2021a) that identifies 67 domestic debt restructuring episodes in low-income countries from 1970 to 2020. Since there is no systematic data on domestic *defaults* for our full sample of low-income countries, we need to map these restructuring events into defaults by imposing a timing assumption. In line with recent empirical research, we assume that domestic defaults occur one year prior to the restructuring.[8]

In Appendix Section A2, we provide a detailed account on how our measure of debt distress changes once we move from a purely external metric (as defined above) to a public debt distress metric that additionally takes domestic debt defaults into account. We find that the incorporation of the 67 domestic default events leads to only four new debt distress episodes. This is because only few domestic debt defaults occur outside of the external debt distress episodes that we have identified above, suggesting that episodes of domestic and external debt distress are highly correlated within countries. This is in line with the academic literature on financial crises that has emphasized the "conglomerate" nature of different types of financial distress (Reinhart et al. 2015, Reinhart 2022). For the purpose of our debt distress prediction analysis, this finding implies that by focusing on well-documented external debt distress episodes we are likely to capture the vast majority of domestic debt distress episodes as well.[9]

---

[8] A duration of one year between domestic default and restructuring is in line with the median duration of domestic law defaults in a sample of 52 low-income, emerging market and advanced economies studied by Erce, Mallucci and Picarelli (2022).

[9] In addition to *de jure* debt restructurings, domestic debt distress episodes can also take the form of *de facto* defaults through inflation that erodes the real value of debt denominated in domestic currency, or through financial repression (see for example IMF 2021a, 2021b). We find that de facto domestic defaults are highly correlated with external debt distress episodes. Incorporating de facto domestic defaults into the analysis, however, raises additional measurement challenges and risks obfuscating the empirical distress prediction analysis, since most variables that are commonly used to identify de facto domestic defaults (inflation, interest rates) are also important predictors of debt distress, i.e., right-hand side variables.

## 2.2. Predictors of debt distress

To identify candidate predictors of debt distress with good data coverage for our sample of low-income countries, we begin with the existing LIC DSF model of the World Bank and the IMF (IMF and World Bank 2017) and the comprehensive literature review provided by Badia et al. (2022). In this subsection, we provide a short overview of all variables and point interested readers to Appendix Section A3 for a detailed account of all variable definitions, transformations, and data sources. Overall, we consider 28 candidate predictors of debt distress that fall into five distinct categories.

First, we consider a broad range of debt burden indicators that measure both debt stocks and repayment flows and that focus on external as well as on total public (external plus domestic) debt. Each debt burden indicator is either scaled by GDP and exports (in the case of stocks) or by exports and fiscal revenue (in the case of repayment flows). Measures of external debt (service) are taken from the World Bank's International Debt Statistics. Constructing series of total public debt and debt service is challenging since data on domestic public debt (service) in LICs is scarce and particularly limited for early parts of the sample. To expand coverage, we piece together data from different sources, including the IMF's World Economic Outlook database and the work by Abbas et al. (2010), Mauro et al. (2015) and Reinhart and Rogoff (2011) (see Appendix A3 for details).

Second, we consider measures of institutional quality and other slow-moving country characteristics as emphasized by Reinhart et al. (2003) and Kraay and Nehru (2006). Our primary measure of institutional quality is the World Bank's Country Policy and Institutional Assessment (CPIA) which is a composite indicator of a country's strength in economic policies and public sector management. As in IMF (2021), we also use different measures for a country's credit and default history, openness to trade and GDP per capita to measure structural country characteristics and overall resilience to debt distress. Third, we consider GDP growth rates and inflation rates to assess the cyclical position of the debtor economy and therefore its short-term ability to repay. Fourth, we use two variables from the Database of Political Institutions (Cruz et al. 2021) to measure political instability. Specifically, we use the number of years that the chief executive has been in office and the number of years that are left in the chief executive's term as proxies for the position of the country in the political cycle and potential signals for vulnerabilities from heightened political uncertainty. Fifth, we rely on different measures to capture the impact of external macroeconomic and financial conditions on debt vulnerabilities in developing countries (see e.g. Reinhart and Rogoff 2009 or Johri et al. 2022). Our measures include different indicators for capital inflows (the current account balance, net FDI inflows and remittances)[10], changes in exchange rates and the terms of trade, as well as proxies for global financial conditions (US Treasury yields, global real GDP growth). Finally, we include a country's stock of foreign exchange reserves as a measure for the country's capacity to absorb external shocks.

## 3. Predicting debt distress

### 3.1 Prediction strategy

Our focus in this paper is on predicting future debt distress, conditional on not currently being in debt distress, i.e., we are interested in predicting the *onset* of debt distress episodes, but not their *duration*. Formally, define $S_{ct} = 1$ if any one of the three signals of debt distress is observed in country c at time t, and zero otherwise. Our sample of interest consists of all country-year observations for which $S_{ct} = 0$, $S_{ct-1} = 0$ and $S_{ct-2} = 0$, i.e., the country currently is not in debt distress and has not been in debt distress in either of the two previous years.

For these observations, we are interested in predicting whether a debt distress episode begins in any of the following $k$ years. To measure this outcome, we define a binary variable $Y_{ct+k} = 1$ if $max\{S_{ct+1}, \dots, S_{ct+k}\} = 1$ and $Y_{ct+k} = 0$ otherwise. We refer to this binary outcome as a debt distress *event*, i.e., the onset of a debt distress episode. Our goal is to model $P[Y_{ct+k} = 1 | X_{ct}]$ where $X_{ct}$ is a vector of explanatory variables available at time $t$. In our main results in this section, we focus on one year-ahead predictions, i.e., the case where $k = 1$ and the event of interest is whether

[10] Data on remittances exhibits various peculiarities and is discussed in greater detail in Appendix Section A4.

debt distress occurs in the following year. In Section 4, we also consider prediction horizons of $k =$ 5 years.

In our main results, we model the probability of a debt distress event using a very simple probit specification:

$$P[Y_{ct+k} = 1|X_{ct}] = \Phi(\beta_k' X_{ct})$$

where $\Phi(.)$ denotes the normal cumulative distribution function, $X_{ct}$ is a vector of predictors (including an intercept), and $\beta_k$ is a conformable vector of parameters.

We use the estimated model to generate binary predictions of debt distress events. For each observation in our dataset, we first generate a predicted probability $\hat{p}_{ct+k} = \Phi(\hat{\beta}_k' X_{ct})$ where $\hat{\beta}_k$ denotes the vector of estimated slope coefficients. We then generate a binary prediction $\hat{Y}_{ct+k} = 1$ if $\hat{p}_{ct+k} > p^*$ and $\hat{Y}_{ct+k} = 0$ otherwise, for a given threshold probability $p^*$.

We choose the threshold probability $p^*$ to minimize a prediction loss function that penalizes the rate of false positives *FPR* (incorrectly predicting distress events when distress does not occur, as a share of non-distress events) and the rate of false negatives *FN* (incorrectly not predicting distress events when distress occurs, as a share of distress events), as:

$$FPR = \left(\sum_{c,t} (1 - Y_{ct+k})\ \hat{Y}_{ct+k}\right) / \sum_{c,t} (1 - Y_{ct+k})$$

$$FNR = \left(\sum_{c,t} Y_{ct+k}\left(1 - \hat{Y}_{ct+k}\right)\right) / \sum_{c,t} Y_{ct+k}$$

We use the following quadratic mean loss function to evaluate predictions:

$$L = \sqrt{wFNR^2 + (1 - w)FPR^2}$$

where $w$ is a weight parameter that we set at $w = 0.5$ throughout. This loss function penalizes the rate of false positive and false negative errors. In addition, it penalizes models that generate very different false positive and false negative error rates. For example, a model with $FNR = 0.8$ and $FPR = 0.2$ is penalized more heavily than a model with $FPR = FNR = 0.5$ even though the average of the false positive and false negative rates is the same for both models. We rely on this functional form for the prediction loss function to capture the fact that in many policy applications, users are unlikely to be indifferent between these models, and specifically are unlikely to be comfortable with a prediction model that tolerates very high rates of false negatives (positives) to achieve low rates of false positives (negatives).[11]

Since we are primarily interested in the out-of-sample predictive performance of the model, we implement this approach using J-K-fold cross-validation. Specifically, we divide our sample into $K = 10$ randomly selected, stratified, equally sized, mutually-exclusive and exhaustive subsets. We successively hold back one subset of the data at a time (the "test" sample) and estimate the probit model in the remainder of the dataset (the "training" sample). Using the estimated parameters from the training sample, we generate predicted probabilities for each observation in the test sample. We then choose the cutoff probability $p^*$ to minimize the prediction loss function in the test sample. We repeat this process over all 10 subsets of the data, which produces binary predictions for all observations in the sample. This in turn allows us to calculate the false positive and false negative rates for the entire sample. Finally, since the partition of our model into subsamples is random, to reduce variability we repeat the K-fold cross-validation $J = 10$ times and summarize the performance of the model with the mean of the prediction loss function across each of these $J$ repetitions. We also retrieve the standard deviation of the loss function across these $J$ replications

[11] We focus on *rates* of false positive and false negative errors as a straightforward method to address class imbalance. In our core estimation sample, only 4 percent of country-year observations, correspond to debt distress events. If the accuracy of predictions were to be evaluated using only the counts of false positive and false negative errors, a 96 percent accuracy rate could be achieved simply by predicting no distress for all observations. Focusing on *rates* of false positive and false negative errors in effect assigns greater weight to false negative errors and serves as a simple and transparent alternative to techniques to address class imbalance such as up-sampling.

and use it to construct a confidence interval around the mean value of the prediction loss function for the model.[12]

### 3.2 Variable selection

One of the main contributions of this paper is that we implement a systematic approach to the selection of the set of predictors of debt distress included in the vector of right-hand-side variables $X_{ct}$. We do this by "brute force" consideration of all relevant combinations of the 28 explanatory variables described in Section 2.2. Without further constraints, this would result in over 268 million ($2^{28} = 268{,}435{,}456$) models to consider. To limit the sheer number of models to estimate, we impose three constraints on the set of models considered.[13]

- First, we require all models to include the CPIA measure of policy. The rationale for this is that we would like to focus attention on models that could be used in the World Bank and IMF LIC DSF, in which countries' policy performance as measured by the CPIA plays a central role.[14]
- Second, since we are analyzing debt sustainability, we require all models to include at least one variable related to debt (either stock or flow).
- Third, many of the categories of explanatory variables discussed in Section 2.2 contain multiple proxies for the same concept. For example, debt service as a share of exports and debt service as a share of government revenues are conceptually similar measures of the flow of debt service obligations. We impose the constraint that models can contain at most one

---

[12] Specifically, the confidence interval is defined as $\bar{L} \pm zSD(L)/J$ where $\bar{L}$ and $SD(L)$ are the mean and standard deviation of the prediction loss function across the $J$ replications and $z$ is the percentile of the normal distribution that delivers the desired confidence level.

[13] While variable selection can also be performed through shrinkage methods like the least absolute shrinkage and selection operator (lasso), its objective function, which balances the sum of squared residuals with a penalty on the absolute values of the coefficients, does not align with our desired prediction loss function based on type 1 and type 2 errors. Moreover, the lasso does not allow for a direct comparison of each model's performance, which is a core element of our model selection exercise.

[14] Imposing this constraint does not affect the results of our model selection algorithm since prediction models without the CPIA measure generally do not achieve significantly higher predictive accuracy than models with the CPIA measure.

debt stock measure, at most one debt flow measure, at most one credit history variable, at most one measure of the political cycle, and at most one measure of the change in the value of local currency.

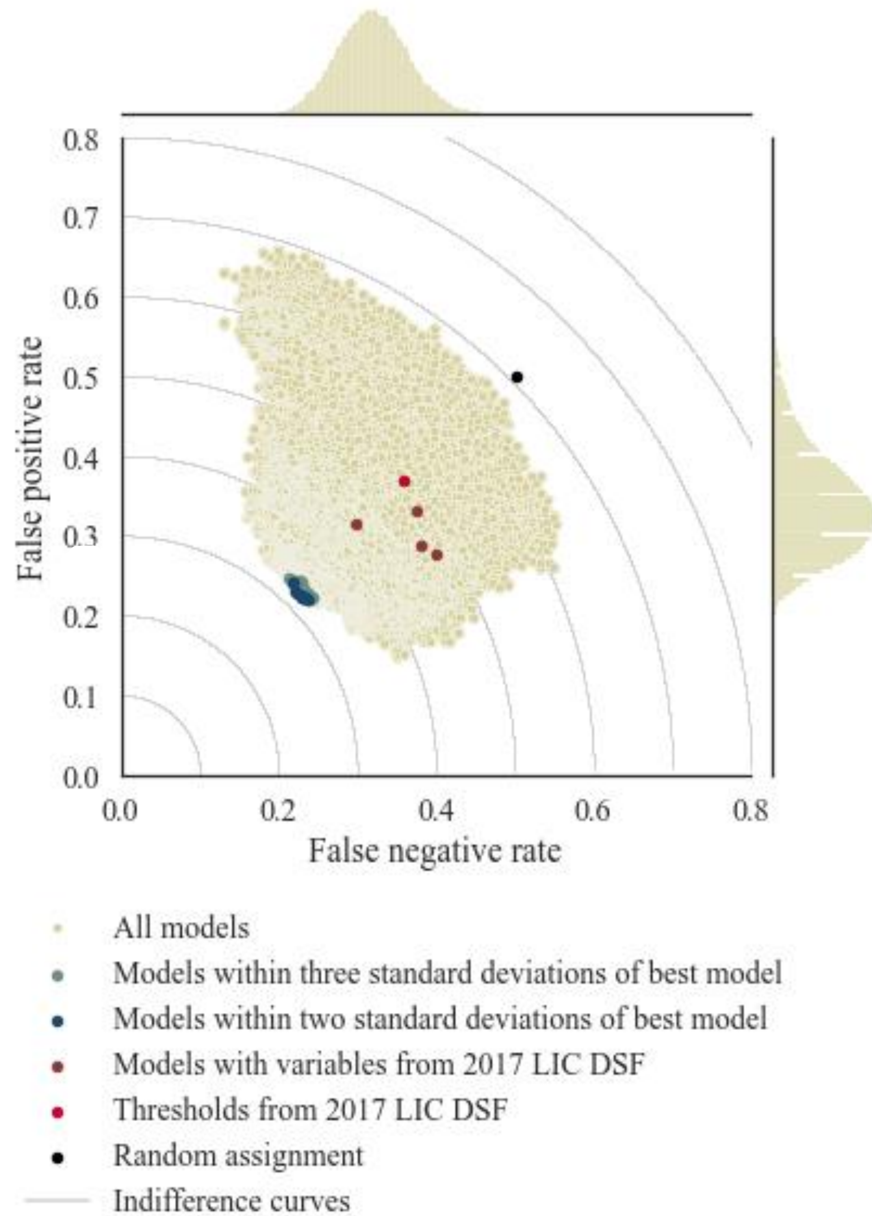
Even with these constraints, we have a very large set of 559,872 candidate models to consider. To facilitate the interpretation of the results, we restrict our sample of observations to those for which all 28 explanatory variables are available. This results in a dataset of 1,002 observations, of which 40 correspond to the onset of a debt distress episode.

### 3.3 Predictive performance

Figure 3 summarizes the main results of our prediction exercise. It plots the false positive and the false negative rates for each of the 559,872 probit models, with each marker representing a different model. Grey lines are isoquants showing all combinations of false positive and false negative rates that achieve the same value for the loss function. Points closer to the origin correspond to better predictive accuracy and a lower prediction loss function. The model that reaches the lowest isoquant has the best predictive performance.

As explained above, our assessment of model performance is based on a J-K-fold cross validation approach that trains and tests models on different randomly-selected segments of the data, resulting in measures of performance that themselves are random. In addition to the top performing model, Figure 3 therefore also shows the set of 22 models whose loss functions fall within the 99% confidence interval of the best model's loss function and the set of 10 models with loss functions within two standard deviations of the best model's loss function.

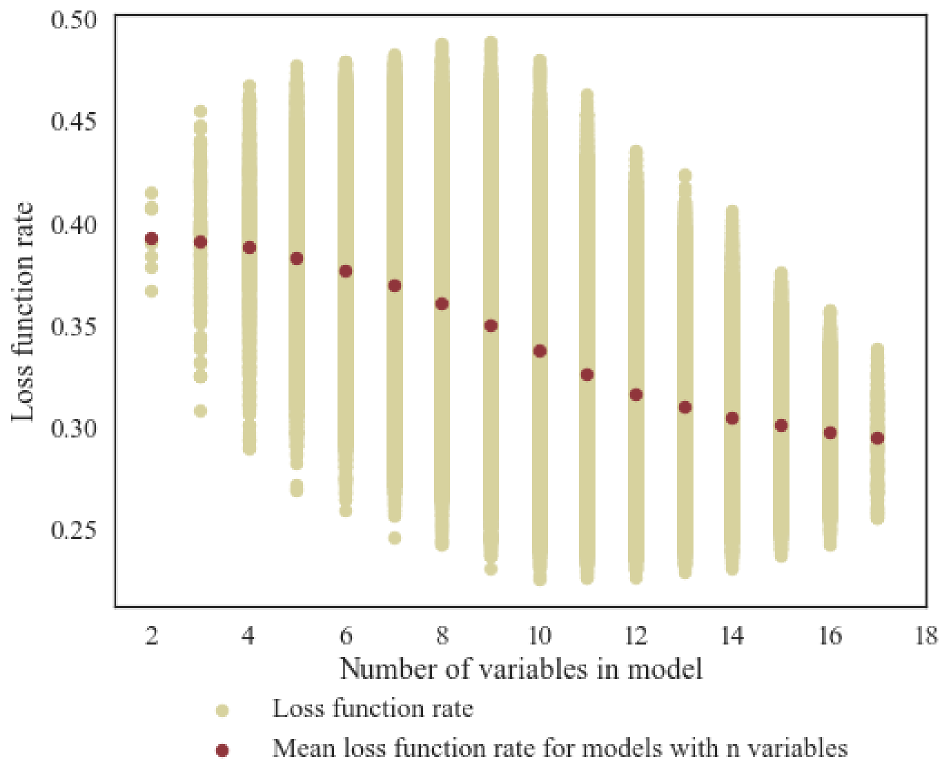
Figure 3. Error rates and loss functions for full model space




*Note:* This figure shows false positive and false negative error rates for the full model set of all 559,872 probit models. Grey lines indicate combinations of false positives and false negatives that yield the same quadratic loss function. The histograms on the horizontal and vertical axis show the marginal distribution of models across false and negative rates separately.

Figure 4 illustrates the tradeoff between model parsimony (i.e., the number of explanatory variables) and predictive performance. In our model comparison setup, the most parsimonious model with the best performance contains only the CPIA and debt service over exports and generates a prediction loss function of 0.36. Adding explanatory variables generally improves model performance on average, as indicated by the downward-sloping pattern of red dots indicating average model performance for each model size, with the biggest marginal benefit arising from adding a third variable to a model with just two variables. Thereafter, giving up parsimony for accuracy yields more limited returns. The performance of the best model for each model size improves rapidly for models up to five variables, but then flattens out considerably. In other words, with only five variables, one can arrive at prediction models that perform only slightly worse in predicting debt distress than significantly more elaborate model with a total of more than ten variables. Furthermore, starting with 12 variables, adding more information no longer increases the predictive performance even of the best models.

Figure 4. Loss functions and model parsimony




*Note:* This figure plots the prediction loss function (vertical axis) against the number of included regressors in the debt distress prediction model (horizontal axis). Each dot represents one of the 559,872 models.

Table 1 displays the top performing models and presents loss functions, error rates and standardized regression coefficients for the most accurate prediction models by different degrees of model parsimony.[15] The regression results reveal interesting insight on the relative strength of different RHS variables in predicting external debt distress. The external debt service over exports ratio and the reserves over imports ratio are included in all top performing models and enter the regressions with economically sizeable and statistically highly significant coefficients. This suggests that these two variables contain particularly valuable information for the prediction of debt distress. None of the other predictor variables consistently enters the top performing models.[16] Finally, previewing our subsequent results below, we note that the best-performing model with only three explanatory variables includes the CPIA, total debt services, and the reserves-to-imports coverage ratio, and achieves quite respectable predictive power.

While all models shown in Table 1 exhibit strong predictive performance, not all are amenable for policy use. Some of the models include predictor variables with counter-intuitive coefficient signs that would lead to perverse incentives when applied in a policy setting. The public debt to export ratio in models 3 and 4, for example, enters the regressions with a negative sign, suggesting that higher debt burdens are associated with lower distress risk.17 Other models include variables with very small coefficient sizes, e.g., the external public debt stock as a share of GDP in model 7, and therefore contribute very little to the predicted probability of distress. Finally, some models include variables with low data coverage or severe measurement issues, such as the remittances over GDP

[15] Standardized regression coefficients in this probit setting are defined as $\beta_x^{standardized} = \frac{\beta_x^{marginal}\sigma_x}{\sigma_{\hat{p}}}$, where $\beta_x^{marginal}$ is the estimated marginal effect of variable $x$; $\sigma_x$ is the standard deviation of $x$, and $\sigma_{\hat{p}}$ is the standard deviation of the predicted probabilities for the model. These coefficients measure the number of standard deviations the predicted probability of distress increases with a one-standard deviation increase in the explanatory variable, holding fixed all the other explanatory variables at their sample means.

[16] The CPIA score is required to be included in all models, since we want to restrict attention to models that could potentially be used in the context of the IMF and World Bank debt sustainability analysis. On the importance of institutional quality and the CPIA for debt distress prediction see Reinhart et al. (2003) and Kraay and Nehru (2006).

[17] This somewhat peculiar finding happens surprisingly often in our unconstrained model space when models contain one debt service variable – which typically enters negatively and highly significantly – and one debt stock variable. This may reflect the fact that, conditional on the level of debt service which raises the risk of debt distress, higher debt stocks may reflect persistent stronger fundamentals that encouraged creditors to lend to the country, which would lower the risk of debt distress.

ratio included in models 6 and 7 (for a detailed discussion of the measurement issues in remittances, see Appendix A4). We address these undesirable features of some of the top models systematically in the next subsection.

Table 1. Top performing prediction models by number of predictors

| | Dependent variable: Incidence of external sovereign debt distress in t+1 | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| CPIA | -0.15** | -0.10* | -0.10* | -0.08* | -0.07* | -0.06* | -0.06** |
| Ext. debt service / exports | 0.22*** | 0.19*** | 0.25*** | 0.22*** | 0.22*** | 0.12*** | 0.12*** |
| Reserves / imports | | -0.24*** | -0.23*** | -0.20*** | -0.20*** | -0.13** | -0.09** |
| Public debt / exports | | | -0.10 | -0.11 | | | |
| Inflation | | | | 0.08 | | | |
| GDP p.c. | | | | | 0.12*** | 0.12*** | 0.11*** |
| NPV of ext. debt / exports | | | | | -0.08 | | |
| GDP growth | | | | | -0.10** | | |
| Remittances / GDP | | | | | | -1.74** | -1.64** |
| Post-2001 dummy | | | | | | -0.07 | 0.01 |
| Remittances / GDP x post-2001 | | | | | | 1.62** | 1.54** |
| US 10-year yield | | | | | | | 0.08* |
| Ext. debt stock / GDP | | | | | | | -0.03 |
| Years left in current term | | | | | | | -0.06 |
| Number of variables | 2 | 3 | 4 | 5 | 6 | 7 | 10 |
| Loss function | 0.37 | 0.31 | 0.29 | 0.27 | 0.26 | 0.25 | 0.23 |
| False positive rate | 0.37 | 0.32 | 0.30 | 0.25 | 0.20 | 0.25 | 0.22 |
| False negative rate | 0.36 | 0.30 | 0.28 | 0.28 | 0.31 | 0.24 | 0.23 |
| Number of observations | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 |

*Note:* This table shows standardized marginal effects for the top prediction models, i.e., the number of standard deviations increase in the predicted probability corresponding to a one standard deviation increase in the explanatory variable, holding all other variables constant at their sample means. Each of the models shown in this table minimizes the loss function for a given number of predictor variables. Stars refer to statistical significance of the underlying probit coefficient estimates: * p<0.1, ** p<0.05, *** p<0.01. All regressions include a constant not reported here.

### 3.4 Model selection

In the previous subsection we measured the out-of-sample predictive performance of 559,872 probit models and identified a large set of models with strong predictive performance. For practical policy applications, however, we need to narrow down to one or a few models with the "best" predictive performance. Doing so is not straightforward, for two reasons. First, although we perform repeated J-K-fold cross-validation, the random partitioning of data into training and test data sets inevitably creates some random variation in the prediction loss function. Second, and as discussed, there are considerations beyond raw predictive power that are relevant in policy applications of a debt distress prediction model. To address this second concern, we impose three additional filters to restrict attention to models with desirable properties beyond prediction accuracy. The following subsections address each of the three filters in turn. Appendix Table B1 provides additional details by documenting how each of the predictor variables is affected by the filters.

#### 3.4.1 Ruling out perverse policy incentives

As explained in the introduction, the debt sustainability framework for low-income countries not only functions as an early warning system but also informs a wide range of macroeconomic and financial policies. The model, for example, is used to inform low-income countries' borrowing decisions and serves as a guide on how to balance funding needs with a country's ability to repay. In this context, we need to ensure that the debt distress model does not create perverse policy incentives. For example, the application of a debt distress prediction model with a negative coefficient on a debt stock variable, for example, would imply that low-income countries should (infinitely) increase their debt stocks in order to reduce distress vulnerability.

To rule out such perverse policy incentives, we consider only models in which the estimated coefficients on debt stock and flow measures are positive, and the estimated coefficients on the reserves over imports ratio and the CPIA are negative, in any models in which any of these variables appear. Applying this filter eliminates 27% of all models. Although this might appear as fairly restrictive, it is important to note that a large share of these models would otherwise also have been dropped by one or both of the filters described below.

#### 3.4.2 Data availability

The debt distress prediction model needs to be applicable to all low-income countries in a timely manner. It is therefore crucial that the model only requires data inputs that are readily available for a broad cross-section of LICs. To rule out models with infeasible data requirements, we construct a data availability factor that measures the share of country-year observations since 2000 for which all data inputs for a model are available. We use this measure to rule out all models with data availability below 90 percent, i.e. models in which more than 10 percent of country-year observations since 2000 are missing. 14% of the models that remained in the sample after applying the perverse incentives filter pass this threshold.

#### 3.4.3 Economically meaningful effect sizes

Finally, the inclusion of additional variables is costly in the practical application of the model, e.g., for data compilation and projection, as well as increasing model complexity. To justify these efforts, one would therefore prefer a model in which all estimated coefficients have economically meaningful effect sizes. This also ensures that policy advice is targeted at variables with significant impact on the overall debt distress rating, rather than at variables that have quantitatively negligible effects on the predicted probability of distress. To rule out models with economically meaningless effect sizes, we require all normalized marginal effects to be greater than 0.05, corresponding to roughly the 20$^{th}$ percentile of the distribution of this measure across all variables in all models.[18] Of the remaining models, after having applied the perversive-incentives and the data availability filter, 17% of the models pass this threshold. Note that one economically insignificant variable in a model

[18] In the case of the three remittances variables, economic meaningfulness requires the product of the standard deviation of remittances and the sum of the marginal effects of remittances relative to GDP and the interaction with the post-2001-dummy to exceed 5% of the standard deviation of the predicted probabilities of the model in question (see Appendix A4 for details on our treatment of remittances data).

is sufficient for this rule to bind, which explains why a threshold at the 20th percentile excludes 83% of all models. After the application of this filter, we are left with a total of 9,429 models.

### 3.5 Top models

This subsection presents the top performing models from our model selection exercise, i.e., the models that have the highest predictive performance among the set of models that fulfil the desirable properties described in Section 3.4. We present top models for different degrees of parsimony, taking into account the trade-off between predictive accuracy and the number of included variables illustrated in Figure 4 above.

All models presented in Table 2 combine high predictive accuracy with high data coverage. All estimated coefficients have economically meaningful effect sizes and none of the models implies perverse incentives for policy trade-offs. In this sense, the proposed models are all suitable for policy application. Table 2 confirms the importance of projected debt service flows in percent of exports which, in contrast to debt stock measures, consistently enter the top performing models. Additionally, the reserves over import ratio, GDP per capita and inflation are shown to be valuable predictor variables for debt distress in low-income countries.

It also is striking to notice how few explanatory variables enter in this set of top-performing models. While we have considered all relevant combinations of the set of 28 potential predictors of debt distress, the best-performing models with two through seven predictors jointly include only 9 of these variables. Most notably, measures of the NPV of external debt, and measures of the stock of domestic debt and the flow of domestic debt service are not included in any of the models with the best predictive performance. For the NPV of external debt, this in part reflects the one-year-ahead prediction horizon of interest. Over this short horizon, it is not surprising that immediate debt service due the following year is a strong predictor of debt distress, while the flow of debt service in the distant future captured in the NPV of debt is not.

Table 2. Best prediction models for one-year ahead debt distress prediction

| | Dependent variable: Incidence of external sovereign debt distress in t+1 | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| CPIA | -0.15** | -0.10* | -0.12** | -0.08* | -0.08** | -0.06 | -0.11** |
| Ext. debt service / exports | 0.22*** | 0.19*** | 0.18*** | 0.17*** | 0.17*** | 0.15*** | |
| Reserves / imports | | -0.24*** | | -0.21*** | -0.17*** | -0.15** | -0.17* |
| GDP p.c. | | | 0.18*** | 0.14*** | 0.16*** | 0.13** | 0.25*** |
| Inflation | | | 0.11** | | 0.09** | | 0.11 |
| GDP growth | | | | -0.09** | | -0.09* | |
| Credit history | | | | | -0.07 | | -0.07 |
| Commodities terms of trade | | | | | | -0.08* | -0.09 |
| US 10 year yield | | | | | | 0.08* | 0.12* |
| Openness | | | | | | | -0.10 |
| CA balance / GDP | | | | | | | -0.06 |
| Ext. debt stock / exports | | | | | | | 0.09 |
| Number of variables | 2 | 3 | 4 | 5 | 6 | 7 | 10 |
| Loss function | 0.37 | 0.31 | 0.29 | 0.27 | 0.26 | 0.27 | 0.29 |
| False positive rate | 0.37 | 0.32 | 0.33 | 0.21 | 0.25 | 0.19 | 0.30 |
| False negative rate | 0.36 | 0.30 | 0.24 | 0.32 | 0.28 | 0.34 | 0.27 |
| Data coverage since 2000 | 0.96 | 0.94 | 0.91 | 0.93 | 0.91 | 0.92 | 0.92 |
| Number of observations | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 | 1,002 |

*Note:* This table shows standardized marginal effects for the top prediction models, i.e., the number of standard deviations increase in the predicted probability corresponding to a one standard deviation increase in the explanatory variable, holding all other variables constant at their sample means. Each of the models shown in this table minimizes the loss function for a given number of predictor variables and satisfies the additional model selection constraints described in Section 3.4. Stars refer to statistical significance of the underlying probit coefficient estimates: * p<0.1, ** p<0.05, *** p<0.01. All regressions include a constant not reported here.

For domestic debt, recall from our discussion of debt distress events in Section 2 that episodes of external and domestic debt distress tend to coincide, at least to the extent that we can measure the latter well. Despite this, measures of stocks of total public debt and total public debt service, which include domestic debt, do not enter any of the top-performing prediction models. This finding admits two overlapping interpretations: (a) sovereigns may have a wider range of options, including monetizing deficits and financial repression, to avoid outright default on domestic debt; and/or (b) our measures of domestic debt stocks and debt service are too noisy to identify a systematic relationship with debt distress episodes. Both interpretations suggest the need for better measurement, both of implicit defaults on domestic debt (to address (a)) and of domestic debt stocks and debt service (to address (b)).

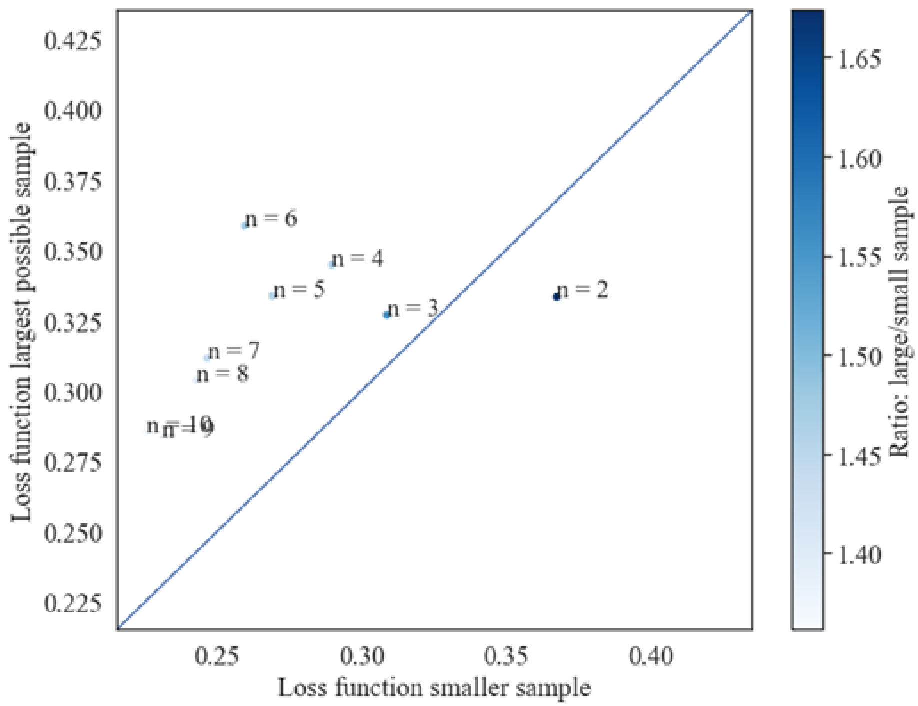
Choosing a final model from this set of top performing models boils down to trading off the desired degree of parsimony and data coverage with the desired degree of predictive accuracy. A simple model with just the CPIA, debt service over exports and the reserve over imports ratio reaches an impressive loss function of only 0.31. This loss function can be further reduced to 0.27 with the addition of two regressors, GDP per capita and the GDP growth rate, with only small costs to data availability and parsimony. Adding more variables to this model with 5 predictors yields only minimal gains in predictive accuracy but comes with the cost of (slightly) lower data coverage and parsimony.

Comparing Table 1 and 2, it is notable that the best models with two and three predictors are the same, meaning that the application of the model selection filters in Section 3.4 does not affect these two models. In other words, these are the unconstrained best models with two and three predictors in our entire model space. However, for models with four or more predictors, imposing the model selection filters in Section 3.4 does come at some cost of predictive performance. For example, the best-performing model in Table 1 achieves a prediction loss function of 0.26, while among the unconstrainted set of models in Table 1, the lowest value of the prediction loss function is 0.23.

### 3.6 Robustness

In the previous subsections, all 559,872 probit models were estimated on the same dataset of 1,002 country-year observations to ensure comparability of performance across models. In this section, we provide a robustness test that re-estimates each of the best models in the largest possible data sample for each model. In contrast to the previous exercises, model performance here is not directly comparable across models, but provides valuable insights in how stable the predictive performance of each model is to the inclusion of additional data points.

Figure 5. Re-estimation of top models in larger data sample




*Note:* This figure shows how the loss function of each of the best models (see Table 2) changes when the model is re-estimated in the largest possible data sample. "n" refers to the number of predictor variables included in each model. The loss function on the horizontal axis is estimated in the core sample of 1,002 country-year observations, for which data exists for all candidate models. The loss function on the vertical axis is estimated in the largest data sample available for each model. The blue color scale indicates the extent to which sample size increases.

Figure 5 presents the results from this exercise. It depicts the loss function achieved in the previous, fixed sample on the horizontal axis and the loss function estimated in the new, larger sample on the vertical axis. Our results show large differences in the robustness of models to changes in the increase in sample size. The parsimonious models (with 2 and 3 predictor variables) show comparatively robust predictive accuracy in the enlarged sample, whereas most of the more complex models experience significant reductions in their predictive performance. This is remarkable given that sample size increases most for models with few predictor variables (as indicated by the blue color scale). While it is difficult to generalize definitively, the evidence in Figure 5 suggests that model stability is a further reason to prefer more parsimonious models over less parsimonious models.

## 4 Model extensions

This section extends the one-year ahead probit prediction models analyzed in the previous section by considering two alternative approaches: machine learning models (Section 4.1) and five-year ahead predictions that are a common use case in policy applications (Section 4.2).

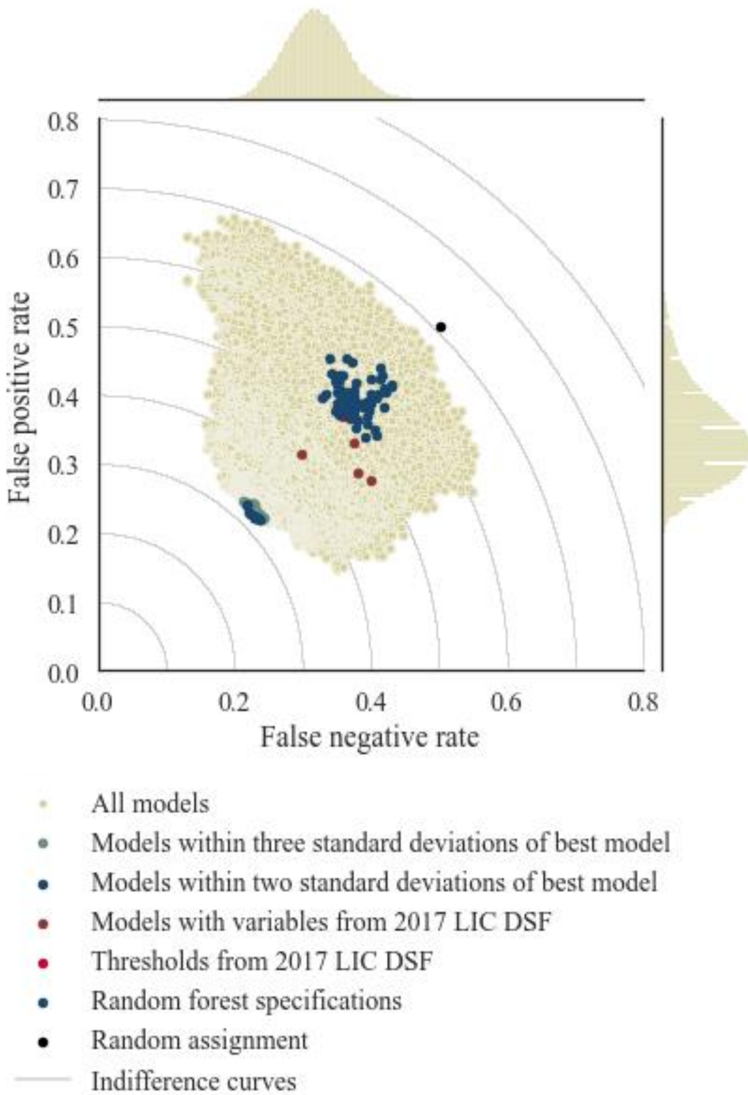
### 4.1 Alternative algorithms

Probit (and Logit) models have long been the standard methodology used in debt prediction models. Yet, recent decades have seen a rise in the use of alternative machine learning methods in the literature on crisis prediction (Badia et al. 2022).

To test whether alternative methods might outperform the Probit algorithm in the prediction of debt distress in low-income countries, we perform a horse race with a Random Forest classifier, one of the most commonly applied machine learning classifiers. The Random Forest classifier has multiple attributes that render it a potentially well-suited alternative to the Probit models in the context of debt distress. Random Forest models are unaffected by collinearity in the data, handle data with outliers well,[19] and although sometimes described as a "black box", Random Forest models do provide information on the most important variables (or *features*). Furthermore, the algorithm is relatively easy to explain in non-specialist terms, which would be of value for the application of the distress prediction model in policy contexts.

[19]This is only true when outliers are in the training sample. Random Forests perform less well when extrapolating to data in ranges not existent in the training set.

Figure 6. Error rates and loss functions for full model space – alternative algorithms




*Note:* This figure shows false positive and false negative error rates for the full model set of all 559,872 probit models and for all Random Forest specifications. Grey lines indicate combinations of false positives and false negatives that yield the same quadratic loss function. The histograms on the horizontal and vertical axis show the distribution of models across false and negative rates.

To test whether the Random Forest outperforms the Probit method applied in Section 3, we perform a grid search over 56 potential specifications of the Random Forest (a process also referred to as "parameter tuning"). Because of the imbalanced data structure, all Random Forest models apply balanced class weights in all specifications, meaning the algorithm applies weights that are inversely proportional to each class's frequency in the input data. We test all combinations of the following parameters:

- As a split-quality-measure, we run the algorithm with a gini-measure, as well as entropy for the Shannon information gain.
- For the number of estimators (the number of trees), we test four values between 25 and 100.
- For the maximum depth of the forest (maximum number of nodes), we consider six steps between 10 and 110, as well as a specification without any constraint on the maximum depth.

For each specification we then apply J=10 stratified K-fold cross validations (K=10), ensuring that the frequency of the positive class is the same across folds. Unlike in the approach applied with the Probit classifier, we do not run the Random Forest for each combination of right-hand-side variables separately, since the Random Forest algorithm omits variables that do not provide added information, and hence chooses the optimal combination of variables organically.

Figure 6 plots the resulting error rates and loss functions. It is evident that the best performing Random Forest Models perform significantly worse than the best probit models identified in Section 3 of the main text.

### 4.2 Predicting debt distress five years into the future

Our baseline empirical model is designed to predict debt distress one year into the future. However, most policy applications focus primarily on predicting debt distress over medium term horizons. Practitioners often meet this need by generating projections of future debt burden indicators under alternative scenarios, and then examine whether these projections cross thresholds derived from one-year-ahead prediction models (see for example Appendix Section C for a description of the IMF and World Bank debt sustainability framework). A shortcoming of this approach is that it creates space for optimism bias. Projected debt ratios are naturally sensitive to assumptions about future borrowing needs, borrowing costs, exchange rate movements, and future growth which affects the denominator when projecting debt ratios into the future. This opens space for optimistic assumptions about the future, leading to lower projected debt burden indicators and a lower projected risk of debt distress (see e.g., Frankel and Schreger 2013, Beaudry and Willems 2022, Estefani-Flores et al. 2023).

An alternative approach that eliminates scope for optimism bias is to base predictions of debt distress at year $t + k$ only on data that is available at time $t$. While predictions of debt distress further into the future inevitably will be less accurate, they may nevertheless be more reliable when compared with the alternative of combining a more accurate one-year-ahead prediction model with auxiliary predictions of future debt burden indicators that may suffer from overoptimism bias.

In this section we investigate how well our setup predicts debt distress five years into the future. Specifically, we seek to predict the binary indicator $Y_{ct+5}$ which is equal to one if there is a signal of debt distress in country $c$ in any of the five years subsequent to $t$, using explanatory variables dated at time $t$, i.e. $X_{ct}$. We first form a sample of observations for which data on $Y_{ct+5}$ and $X_{ct}$ are available. Necessarily, this sample stops in 2016 (because the last year in our dataset is 2021). This results in a sample of 899 observations, of which 175 correspond to a transition into debt distress.[20]

In this sample, we run our model selection algorithm based on J-K-fold cross-validation of all possible models satisfying the constraints described in Section 3.1. Once we have established the predictive accuracy for each candidate model, we apply the model selection procedure described in Section 3.4 to rule out models that provide perverse policy incentives, include economically meaningless coefficients, or require variables with unsatisfactory data availability.[21]

---

[20] Note that although the rate of true positives in the five-year-ahead sample is higher (19 % vs 4 %), the loss function remains a suitable metric to compare the best performing models across the two approaches. This is because the baseline loss function against which they are compared, is independent of the true positive rate.

[21] See Table B2 in Appendix Section B for a regression table with the unconstrained top prediction models.

Table 3. Top performing prediction models for five-year ahead debt distress prediction

| | Dep. variable: Incidence of external sovereign debt distress within next five years | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| CPIA | -0.41*** | -0.41*** | -0.36*** | -0.37*** | -0.31*** | -0.32*** | -0.33*** |
| Ext. debt service / exports | 0.66*** | 0.72*** | 0.66*** | 0.64*** | 0.55*** | 0.51*** | 0.63*** |
| GDP p.c. | | 0.59*** | 0.60*** | 0.64*** | 0.53*** | 0.60*** | 0.57*** |
| Inflation | | | 0.19** | 0.18** | | | |
| Openness | | | | -0.09 | | -0.15 | -0.17 |
| CA balance / GDP | | | | | -0.25*** | -0.29*** | -0.21*** |
| Credit history | | | | | -0.30*** | -0.28*** | -0.28*** |
| US 10 year yield | | | | | 0.31*** | 0.31*** | |
| Reserves / imports | | | | | | | -0.27*** |
| Y.o.y. change in FX rate | | | | | | | -0.24*** |
| Number of variables | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Loss function | 0.40 | 0.30 | 0.29 | 0.29 | 0.28 | 0.28 | 0.28 |
| False positive rate | 0.41 | 0.30 | 0.30 | 0.30 | 0.29 | 0.26 | 0.27 |
| False negative rate | 0.40 | 0.29 | 0.27 | 0.27 | 0.28 | 0.29 | 0.28 |
| Data coverage since 2000 | 0.96 | 0.93 | 0.91 | 0.91 | 0.93 | 0.93 | 0.93 |
| Number of observations | 899 | 899 | 899 | 899 | 899 | 899 | 899 |

*Note:* This table shows standardized marginal effects for the top prediction models, i.e., the number of standard deviations increase in the predicted probability corresponding to a one standard deviation increase in the explanatory variable, holding all other variables constant at their sample means. Each of the models shown in this table minimizes the loss function for a given number of predictor variables and satisfies the additional model selection constraints described in Section 3.4. In contrast to the tables shown in Section 3, the dependent variable in this table measures the incidence of debt distress in any of the next five years. Stars refer to statistical significance of the underlying probit coefficient estimates: * p<0.1, ** p<0.05, *** p<0.01. All regressions include a constant not reported here.

Table 3 summarizes the results of the model selection algorithm and presents the most accurate prediction models for different degrees of parsimony. As in the one-year ahead prediction exercise, we include the CPIA measure of institutional quality in all models. Beyond this, the results confirm the importance of debt service over exports, which is consistently included in the top performing models. In contrast to the one-year prediction exercise, reserves over imports do not play the same prominent role as a top predictor. This is in line with economic intuition since reserves only provide a buffer against short-term liquidity issues but cannot provide an effective remedy against more structural debt issues. Instead, slow-moving variables such as GDP per capita or a country's credit history play more valuable roles in debt distress prediction. GDP per capita enters the prediction models with a significant and positive coefficient, implying that – all else equal – countries with higher income are more likely to experience external sovereign debt distress. This correlation is likely to be driven by the fact that more developed countries rely more heavily on market finance and are therefore exposed to higher volatility in borrowing costs and market access.

The top models for five-year ahead debt distress prediction perform only slightly worse than their counterparts for the one-year ahead horizon. In the one-year ahead prediction exercise, the loss function was minimized at 0.26 by the best model with six predictor variables. This compares to a slightly higher loss function of 0.28 in the top performing model with six, seven or eight predictor variables for the five-year horizon. For more parsimonious models, we even see a small increase in predictive accuracy. While the top model with only three predictor variables reached a loss function of 0.31 in the one-year ahead prediction exercise, the top model with three predictors in the five-year ahead prediction achieves a loss function 0.30. Overall, these results suggest that predicting debt distress over longer horizons can be done with limited loss of predictive accuracy. This could be a promising alternative to the existing practice of predicting medium-term debt distress by projecting debt burden indicators into the future.

## 5. Comparison with the existing IMF and World Bank LIC DSF

In this section, we compare the predictive performance of our preferred model to the predictive performance of a key policy benchmark - the existing IMF and World Bank LIC DSF model which we describe in detail in Appendix Section C. We focus our comparison on the best parsimonious model with only three predictor variables (the "BPM"). As noted in Section 3, this model combines high predictive accuracy with high amenability for policy use and high stability across different samples.

Conducting a systematic and fair comparison of both models is challenging as the models differ along several dimensions and as both models have been trained on different data samples, with different dependent variables and different loss functions (see Appendix Section C2). We therefore perform two distinct empirical exercises. The first exercise is based on our full dataset, uses our definition of external debt distress episodes and our quadratic loss function. The second exercise takes place in the original data sample used in IMF and World Bank (2017), with the original debt distress episode definition and the linear loss function of IMF and World Bank (2017).

### 5.1 Comparing predictive performance in the full sample

In this exercise, we take our preferred parsimonious model from Section 3 – the BPM - and generate a binary prediction for each observation in the largest possible sample of data.[22] Specifically, (a) we run J-K fold cross-validation for our BPM in the full sample and retrieve the average p* threshold probability across the J-K fold replications; (b) we estimate the probit model in the full sample and retrieve the predicted probabilities; (c) we classify observations by comparing the predicted probability with the average p*. To generate predictions for the existing IMF and World Bank framework, we mechanically apply the LIC DSF model and identify countries and years at high risk

[22] Specifically, we include all observations for which the data for the best parsimonious model and for the existing LIC DSF specification exist, so that a fair comparison is possible. To create mechanical predictions from the LIC DSF model (see Appendix C for details), we use a backward looking 10-year moving average (rather than an average centered on year t with projections as in the actual LIC debt sustainability analyses). We require a minimum of five observations to compute the moving average. This procedure results in a sample size of 1356 observations.

of debt distress, i.e., countries where the projected debt burden indicators exceed any of the country-specific debt thresholds prescribed in the current LIC DSF (see Appendix Section C1).

Table 4 shows confusion matrices summarizing the predictions of both our parsimonious model and the LIC DSF model. The confusion matrices show that the best parsimonious model achieves a significantly lower loss function than the LIC DSF model. The BPM does not just reduce the number of missed crises from 19 to 15 (false negatives), it also manages to substantially reduce the number of false alarms from 483 to 394 (false positives).

Table 4. Confusion Matrix – BPM and LIC DSF Classification versus Actual

| | **Predicted by Best Parsimonious Model** | | **Predicted by 2017 LIC DSF Model** | |
|---|---|---|---|---|
| **Actual** | No distress | Distress | No distress | Distress |
| No distress | 903 | 394 | 814 | 483 |
| Distress | 15 | 44 | 19 | 40 |
| False positive rate | 0.30 | | 0.37 | |
| False negative rate | 0.25 | | 0.32 | |
| Quadratic loss function | 0.28 | | 0.35 | |

Note: Sample Size = 1356, defined by all observations for which the data for the best parsimonious model and the LIC DSF 2017 specification exist, so that a fair comparison is possible (see Appendix C for details).

### 5.2 Comparing predictive performance in the original LIC DSF sample

The previous empirical exercise favored the BPM, as the exercise was conducted in the full sample, and with the definition of the dependent variable and the quadratic loss function on which the BPM was trained. In a second exercise, we therefore focus on the original LIC DSF sample with the original definition of external debt distress episodes and rely on the linear loss function used in IMF and World Bank (2017) to compare both models.[23] To generate predictions for the BPM, we estimate it in the original sample and identify debt distress predictions by comparing predicted probabilities

[23] As before, we restrict the sample to those country-year observations for which sufficient data is available to compute predictions for both models.

with the p* value that minimizes the linear loss function in-sample, as opposed to J-K-fold cross-validation. We do this to maximize comparability with the LIC DSF model, which was also tuned to minimize an in-sample linear loss function. To generate predictions from the LIC DSF model, we again mechanically compare realized debt burden indicators with the LIC DSF thresholds and identify countries considered to be at high risk of debt distress (see Appendix Section C1).

Table 5 shows confusion matrices for the BPM and LIC DSF framework. When trained to minimize the original prediction loss function that is a linear weighted average of the false positive and false negative rates, with double the weight on false negatives relative to false positives, both models produce highly imbalanced false positive and false negative error rates. To arrive at low rates of missed crises (14% for BPM and 21% for LIC-DSF), both models generate a very large number of false alarms: They misclassify 37% (LIC-DSF) or even 48% (BPM) of all tranquil years as crisis years. With respect to the overall loss function, the BPM marginally outperforms the existing LIC-DSF framework by achieving a loss function of 0.25 (compared to 0.26 for the LIC-DSF).

Table 5. Confusion Matrix – BPM and LIC DSF Classification in 2017 LIC DSF Sample

| | **Predicted by Best Parsimonious Model** | | **Predicted by 2017 LIC DSF Model** | |
|---|---|---|---|---|
| **Actual** | No distress | Distress | No distress | Distress |
| No distress | 172 | 156 | 206 | 122 |
| Distress | 9 | 54 | 13 | 50 |
| False positive rate | 0.48 | | 0.37 | |
| False negative rate | 0.14 | | 0.21 | |
| Linear loss function | 0.25 | | 0.26 | |

Note: Sample Size = 391, defined by all observations for which the data for the best parsimonious model and the LIC DSF 2017 specification exist, so that a fair comparison is possible (see Appendix C for details).

While the large rate of false alarms may not be considered a satisfactory prediction outcome, it is remarkable that the much simpler BPM outperforms the existing framework even in this second exercise that relied entirely on the original definitions and sample in which the LIC DSF framework was trained. This finding points to considerable scope in simplifying the existing IMF and World Bank LIC DSF without compromising its predictive performance.

## 6. Conclusion

This paper has developed an empirical model to predict episodes of external sovereign debt servicing difficulties in low-income countries based on refined measures of debt distress episodes and a systematic algorithm to assess the out-of-sample predictive performance of more than 550,000 probit models through J-K fold cross-validation. Our paper and findings have several implications for the design and application of debt sustainability frameworks in developing countries.

First, we have demonstrated that highly parsimonious linear models with as few as three predictor variables combine high predictive accuracy with high amenability for policy application. Most importantly perhaps, these simple linear prediction models perform at least as well as if not better than much more complex specifications such as Random Forest algorithms or the existing debt sustainability framework employed by the IMF and World Bank. This finding points to considerable scope in simplifying the existing IMF and World Bank LIC DSF without compromising its ability to provide reasonably informative early warning signals.

Second, we show that simple linear models can also be applied to the prediction of debt distress over longer time horizons. When predicting debt distress five years rather than one year into the future, our top performing probit models experience only small reductions in predictive accuracy. This finding points towards a potential remedy for the overoptimism bias that has often plagued debt sustainability frameworks in practice. Instead of projecting predictor variables into the future, linear models can be trained to predict distress over the medium term using only information available at time t.

Third, our results shed light on how domestic debt, and therefore total public debt, could be integrated into debt distress prediction models. In terms of the outcome of debt distress, we have shown that domestic debt distress episodes (measured as outright defaults on domestic creditors) coincide almost perfectly with external debt distress episodes – debt servicing difficulties vis-à-vis both domestic and foreign creditors tend to occur at the same time. In terms of predictors of debt distress, our systematic model selection exercise indicates that external debt service as a share of exports is the best (among all domestic and external debt burden indicators) at predicting debt distress. While we acknowledge significant measurement challenges for both domestic debt burden indicators as well as domestic debt distress, the results here suggest that a prediction model designed to accurately predict external debt distress based on external debt burden indicators will be quite effective at also predicting overall public debt distress.

Finally, it is important to bear in mind the very real limitations of even the best-performing models. While we have seen that quite parsimonious prediction models perform at least as well as more complex models, the absolute level of performance of all models considered in this paper remains modest. For example, the best-performing parsimonious model with three explanatory variables (second column of Table 2) achieves false positive and false negative rates of 0.32 and 0.30, respectively. In other words, even very well-performing models will still fail to predict nearly one third of debt distress events, and conversely will incorrectly raise false alarms of debt distress in one third of cases where distress does not occur. Finally, the models we consider, like all prediction models, have the limitation that their usefulness for predicting debt distress in the future depends on the underlying forces driving debt distress, and their relationship with measured variables in the prediction model, are both stable over time. For example, our findings on the limited role of domestic debt in predicting debt distress may no longer hold in the future if LICs substantially shift their portfolio of sovereign debt towards domestic debt. These caveats indicate that when using such models to make policy decisions, it is important to rely on additional information, careful application of judgment, and above all a healthy dose of caution.

## References


Abbas, S. Ali, Nazim Belhocine, Asmaa ElGanainy, and Mark Horton. 2010. A Historical Public Debt Database. IMF Working Paper 10/245.

Aguiar, Mark, and Gita Gopinath. 2007. Emerging Market Business Cycles: The Cycle is the Trend. *Journal of Political Economy*, 115(1), 69-102.

Aguiar, Mark, Satyajit Chatterjee, Harold L. Cole, and Zachary Stangebye. 2016. Quantitative Models of Sovereign Debt Crises. In *Handbook of Macroeonomics,* Vol. 2, edited by John B. Taylor and Harald Uhlig, 1697-1755.

Aguiar, Mark, and Manuel Amador. 2021. *The Economics of Sovereign Debt and Default*. New Jersey: Princeton University Press.

Arellano, Cristina, Xavier Mateos-Planas, and José-Víctor Ríos-Rull. 2022. Partial Default. *Journal of Political Economy,* 131(6), 1385-1439.

Asonuma, Tamon, and Christoph Trebesch. 2016. Sovereign Debt Restructurings: Preemptive or Post-Default. *Journal of the European Economic Association*, 14(1), 175-214.

Badia, Marialuz Moreno, Paulo Medas, Pranav Gupta, and Yuan Xiang. 2022. Debt Is Not Free. *Journal of International Money and Finance*, 127, 102654.

Beaudry, Paul, and Tim Willems. 2022. On the Macroeconomic Consequences of Over-Optimism. *American Economic Journal: Macroeconomics.* 14(1), 38-59.

Citron, Joel-Thomas, and Gerald Nickelsburg. 1987. Country Risk and Political Instability. *Journal of Development Economics*, 25(2), 385–92.

Clemens, Michael A., and David McKenzie. 2018. Why don't remittances appear to affect growth? *The Economic Journal,* 128(612), 179-209.

Cruz, Cesi, Philip Keefer, and Carlos Scartascini. 2020. The Database of Political Institutions 2020. Washington, DC: Inter-American Development Bank Research Department.

Das, Udaibir, Michael Papaioannou, and Christoph Trebesch. 2012. Sovereign Debt Restructurings 1950-2010: Literature Survey, Data and Stylized Facts. IMF Working Paper 12/203.

Detragiache, Enrica and Antonio Spilimbergo. 2001. Crises and Liquidity: Evidence and Interpretation. IMF Working Paper No. 01/02.

Dvorkin, Maximiliano, Juan M. Sanchez, Horacio Saprizia, and Emircan Yurdagul. 2020. News, sovereign debt maturity, and default risk. *Journal of International Economics,* 126, 103352.

Erce, Aitor, Enrico Mallucci, and Mattia Picarelli. 2022. A Journey in the History of Sovereign Defaults on Domestic-Law Public Debt. ESM Working Paper Series No. 51.

Estefania-Flores, Julia, Davide Furceri, Siddarth Kotari, and Jonathan D. Ostry. 2023. Worse than you think: Public debt forecast errors in advanced and developing countries. *Journal of Forecasting,* 42(3), 685-714.

Farah-Yacoub, Juan, Clemens Graf von Luckner, and Carmen M. Reinhart. 2022. Chasing the Storm: Debt Distress, Arrears, and Defaults. Mimeo.

Farah-Yacoub, Juan, Clemens Graf von Luckner, and Carmen M. Reinhart. 2023. Eternal external debt crises: A long view. Mimeo.

Farah-Yacoub, Juan, Clements Graf von Luckner, and Carmen M. Reinhart. 2024. The Social Costs of Sovereign Default. NBER Working Paper No. 32600.

Frankel, Jeffrey, and Jesse Schreger. 2013. Over-optimistic official forecasts and fiscal rules in the Eurozone. *Review of World Economics,* 149, 247-272.

Gelpern, Anna, Sebastian Horn, Scott Morris, Bradley Parks, and Christoph Trebesch. 2021. How China Lends: A Rare Look Into 100 Debt Contracts with Foreign Governments. *Economic Policy,* eiac054.

Gruss, Bertrand, and Suhaib Kebhaj. 2019. Commodity Terms of Trade: A New Database. IMF Working Paper No. 2019/021

Ha, Jongrim, Ayhan Kose, and Franziska Ohnsorge. 2021. One-Stop Source: A Global Database of Inflation. World Bank Policy Research Working Paper 9737.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2020. Coping with Disasters: Two Centuries of International Official Lending. NBER Working Paper 27343.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2022. Hidden Defaults. *AEA Papers & Proceedings*, 112, 531-535.

Horn, Sebastian, David Mihalyi, Philipp Nickol, and Cesar Sosa-Padilla. 2024. Hidden Debt Revelations. Mimeo.

International Monetary Fund and World Bank. 2017. Review of the Debt Sustainability Framework in Low-Income Countries: Proposed Reforms. IMF Policy Paper 08/2017.

International Monetary Fund. 2021a. Issues in Restructuring Sovereign Domestic Debt. Policy Paper No. 2021/071.

International Monetary Fund. 2021b. Review of the Debt Sustainability Framework for Market Access Countries. IMF Policy Paper No. 2021/003.

IMF Independent Evaluation Office. 2002. Evaluation of Prolonged Use of IMF Resources. Evaluation Report. Washington D.C.: International Monetary Fund.

Johri, Alok, Shahed Khan, and Cesar Sosa-Padilla. 2022. Interest Rate Uncertainty and Sovereign Default Risk. *Journal of International Economics,* 139, 103681.

Kaufmann, Daniel, and Aart Kraay. 2023. Worldwide Governance Indicators, 2023 Update. (www.govindicators.org), Accessed on 10/19/2023.

Kraay, Aart, and Vikram Nehru. 2006. When Is External Debt Sustainable? *The World Bank Economic Review*, 20 (3), 341–65.

Krishnamurthy, Arvind, and Tyler Muir. 2017. How Credit Cycles across a Financial Crisis. NBER Working Paper 23850.

Mauro, Paolo, Rafael Romeu, Ariel Binder, and Asad Zaman. 2015. A Modern History of Fiscal Prudence and Profligacy. *Journal of Monetary Economics,* 76, 55-70.

Manasse, Paolo, and Nouriel Roubini. 2009. “Rules of thumb” for sovereign debt crises. *Journal of International Economics*, 78(2), 192-205.

Manasse, Paolo, Nouriel Roubini, and Axel Schimmelpfennig. 2003. Predicting Sovereign Debt Crises. IMF Working Paper No. 03/221.

Medas, Paulo, Tigran Poghosyan, Yizhi Xu, Juan Farah-Yacoub, and Kerstin Gerling. 2018. Fiscal Crises. *Journal of International Money and Finance*, 88, 191–207.

Mitchener, Kris, and Christoph Trebesch. 2023. Sovereign Debt in the Twenty-first Century. *Journal of Economic Literature,* 61(2), 565-623.

Pescatori, Andrea, and Amadou N. R. Sy. 2007. Are Debt Crises Adequately Defined? *IMF Staff Papers* 54 (2): 306–37.

Reinhart, Carmen M., Kenneth Rogoff, and Miguel Savastano. 2003. Debt intolerance. *Brookings Papers on Economic Activity,* (1), 1-74.

Reinhart, Carmen, and Kenneth Rogoff. 2009. *This time is different.* Princeton: Princeton University Press.

Reinhart, Carmen, and Kenneth Rogoff. 2011. From Financial Crash to Debt Crisis. *American Economic Review,* 101(5), 1676-1706.

Reinhart, Carmen, Vincent Reinhart, and Kenneth Rogoff. 2015. Dealing with Debt. *Journal of International Economics,* 96(1), 43-55.

Reinhart, Carmen, Vincent Reinhart, and Christoph Trebesch. 2016. Global Cycles: Capital Flows, Commodities, and Sovereign Defaults, 1815-2015. *American Economic Review Papers & Proceedings,* 106(5), 574-580.

Reinhart, Carmen, and Christoph Trebesch. 2017. The International Monetary Fund: 70 Years of Reinvention. *Journal of Economic Perspectives*, 30(1), 3-28.

Reinhart, Carmen. 2022. From Health Crisis to Financial Distress. *IMF Economic Review*, 70, 4-31.

Rodriguez, Arnulfo, and Pedro Rodriguez. 2006. Understanding and predicting sovereign debt rescheduling. *Journal of Forecasting,* 25(7), 459-479.

World Bank. 2023. *World Development Report 2023: Migrants, Refugees, and Societies.* Washington D.C.: World Bank.

# Online Appendix to “Predicting Debt Distress in Low-Income Countries”

Graf von Luckner, Horn, Kraay, Ramalho

**Overview:**



## Appendix A.  Data

This appendix provides detailed information on all variables used in our analysis. Appendix Section A1 provides information on the country and time coverage of our sample. Appendix Section A2 covers the variables used to construct our measure of external debt distress episodes and presents additional results. Appendix Section A3 gives detailed information on the definition and sources for the predictor variables. Finally, Appendix Section A4 discusses data on remittances which exhibits various peculiarities.

### Appendix A1. Sample coverage

Our analysis is based on the same sample of countries that the World Bank and IMF use to estimate the empirical model underlying the LIC DSF (IMF and World Bank 2017). This sample includes all countries that are eligible to receive IDA assistance, i.e., all countries with GNI per capita of less than 1255 USD and a small number of countries with per capita incomes above this threshold but with insufficient creditworthiness to borrow from the International Bank for Reconstruction and

Development.[24] Based on this definition, we identify a total of 80 countries and include in the sample all years with available data since the independence of each country. We exclude all country-year observations for colonies since debt distress under colonial dependence might arguably play out in different ways. Data on the first year of independence from colonial rule is from Reinhart and Rogoff (2009).

Table A1 lists all countries in our sample, the corresponding time coverage and the number of country-year observations with sufficient data to identify debt distress episodes. The regressions presented in Section 3 include only a subset of these country-year observations depending on the availability of predictor variables.

Table A1. Sample coverage

| **Country** | **First year** | **Last year** | **Number of observations** |
|---|---|---|---|
| Afghanistan | 2006 | 2021 | 16 |
| Albania | 1991 | 2021 | 31 |
| Angola | 1985 | 2021 | 37 |
| Armenia | 1993 | 2021 | 29 |
| Bangladesh | 1972 | 2021 | 50 |
| Benin | 1970 | 2021 | 52 |
| Bhutan | 1981 | 2021 | 41 |
| Bolivia | 1970 | 2021 | 52 |
| Burkina Faso | 1970 | 2021 | 52 |
| Burundi | 1970 | 2021 | 52 |
| Cabo Verde | 1981 | 2021 | 41 |
| Cambodia | 1981 | 2021 | 41 |
| Cameroon | 1970 | 2021 | 52 |
| Central African Republic | 1970 | 2021 | 52 |
| Chad | 1970 | 2021 | 52 |
| Comoros | 1970 | 2021 | 52 |
| Congo, Dem Rep | 1970 | 2021 | 52 |

[24] Over the past decades, several countries have graduated from and / or fallen back into IDA status. These countries are included in the sample if they have spent at least half of the sample years in IDA status (as of 2017 when the LIC DSF model was estimated).

| | | | |
|---|---|---|---|
| Congo, Rep | 1970 | 2021 | 52 |
| Cote d'Ivoire | 1970 | 2021 | 52 |
| Djibouti | 1970 | 2021 | 52 |
| Dominica | 1981 | 2021 | 41 |
| Equatorial Guinea | 1980 | 1985 | 2 |
| Eritrea | 1994 | 2021 | 28 |
| Ethiopia | 1970 | 2021 | 52 |
| Gambia, The | 1970 | 2021 | 52 |
| Georgia | 1992 | 2021 | 30 |
| Ghana | 1970 | 2021 | 52 |
| Grenada | 1970 | 2021 | 52 |
| Guinea | 1970 | 2021 | 52 |
| Guinea-Bissau | 1975 | 2021 | 47 |
| Guyana | 1970 | 2021 | 52 |
| Haiti | 1970 | 2021 | 52 |
| Honduras | 1970 | 2021 | 52 |
| Kenya | 1970 | 2021 | 52 |
| Kosovo | 2009 | 2021 | 13 |
| Kyrgyz Republic | 1992 | 2021 | 30 |
| Lao PDR | 1970 | 2021 | 52 |
| Lesotho | 1970 | 2021 | 52 |
| Liberia | 1970 | 2021 | 52 |
| Madagascar | 1970 | 2021 | 52 |
| Malawi | 1970 | 2021 | 52 |
| Maldives | 1978 | 2021 | 44 |
| Mali | 1970 | 2021 | 52 |
| Mauritania | 1970 | 2021 | 52 |
| Moldova | 1992 | 2021 | 30 |
| Mongolia | 1991 | 2021 | 31 |
| Montenegro | 2006 | 2021 | 16 |
| Mozambique | 1983 | 2021 | 39 |
| Myanmar | 1970 | 2021 | 52 |
| Nepal | 1970 | 2021 | 52 |
| Nicaragua | 1970 | 2021 | 52 |
| Niger | 1970 | 2021 | 52 |
| Nigeria | 1970 | 2021 | 52 |
| North Macedonia | 1992 | 2021 | 30 |
| Papua New Guinea | 1970 | 2021 | 52 |
| Rwanda | 1970 | 2021 | 52 |
| Samoa | 1970 | 2021 | 52 |
| Sao Tome and Principe | 1977 | 2021 | 45 |
| Senegal | 1970 | 2021 | 52 |
| Serbia | 1970 | 2021 | 52 |
| Sierra Leone | 1970 | 2021 | 52 |

| | | | |
|---|---|---|---|
| Solomon Islands | 1978 | 2021 | 44 |
| Somalia | 1970 | 2021 | 52 |
| Sri Lanka | 1970 | 2021 | 52 |
| St Kitts and Nevis | 1984 | 2012 | 29 |
| St Lucia | 1981 | 2021 | 41 |
| St Vincent and the Grenadines | 1970 | 2021 | 52 |
| Sudan | 1970 | 2021 | 52 |
| Tajikistan | 1992 | 2021 | 30 |
| Tanzania | 1970 | 2021 | 52 |
| Timor-Leste | 2012 | 2021 | 10 |
| Togo | 1970 | 2021 | 52 |
| Tonga | 1985 | 2021 | 37 |
| Uganda | 1970 | 2021 | 52 |
| Vanuatu | 1981 | 2021 | 41 |
| Vietnam | 1981 | 2021 | 41 |
| Yemen, Rep | 1971 | 2021 | 51 |
| Zambia | 1970 | 2021 | 52 |
| Zimbabwe | 1970 | 2021 | 52 |

## Appendix A2. Identifying debt distress episodes

As described in Section 2.1, we rely on multiple variables to define and identify external debt distress episodes. This subsection presents each of these variables in turn and discusses why we do *not* consider external debt restructurings and domestic default events as separate distress signals.

**Outright default on private external creditors:** Our core data source is Farah-Yacoub et al. (2023) that track defaults on private external creditors in a cross-section of 200 countries and across 200 years. In comparison with similar data sources such as Reinhart and Rogoff (2009) or Asonuma and Trebesch (2016), this source has the advantage of offering comprehensive time and cross-sectional coverage of low-income countries. More specifically, this source tracks all low-income countries in our sample from their year of independence and therefore allows to build comprehensive default histories that we use as additional predictor variables.

**Arrears data:** Our arrears-based default measure is equal to one if total outstanding arrears exceed five percent of the total public and publicly guaranteed external debt stock and remain above this threshold for at least three consecutive years. All data for this signal is taken from the World Bank IDS. Specifically, we aggregate outstanding arrears to official and private creditors from both missed principal (IDS code: "DT.AXA.DLXF.CD") and missed interest payments ("DT.IXA.DLXF.CD") and divide by the total external public and publicly guaranteed debt stock of the country ("DT.DOD.DPPG.CD").

When constructing an arrears-based signal of debt distress, the HIPC Initiative requires specific attention. Despite receiving large-scale debt relief, several HIPC Initiative countries keep reporting sizeable legacy arrears to the World Bank International Debt Statistics. Consequently, our debt distress identification algorithm would treat these countries as in distress, despite their evident recovery after the HIPC Initiative. To avoid this issue and to allow all countries that successfully completed HIPC to reenter the estimation sample, we assume that all arrears to official creditors are deemed resolved once a country reaches its HIPC completion point.[25]

**Large IMF balance of payments assistance:** Our IMF distress signal is equal to one if a sovereign makes drawdowns of over 30 percent of its quota during the first six months of a non-concessional IMF lending program. To identify the first six months of non-concessional IMF lending programs, we rely on the data compiled by Reinhart and Trebesch (2016). We match this data with monthly IFS data on the net increase in fund credit from the general resource account ("HPG_XDR") and scale by the country's quota at the beginning of the program ("HPQ_XDR").

In addition to excluding disbursements under concessional lending programs, we also do not consider rapid disbursements under the IMF's Rapid Financing Instrument, which often addresses

[25] Data on HIPC completion points comes from Horn et al. (2022). The debt distress prediction model that underlies the existing World Bank and IMF LIC DSF uses a similar assumption to deal with HIPC legacy arrears (see IMF and World Bank 2017).

funding needs of a limited or transitory nature and which has primarily been used by countries with strong fundamentals and at low or moderate risks of debt distress. The IMF distress signal is activated only for the year in which the rapid disbursement takes place.

Table A2 and Figures A1 and A2 summarize the resulting debt distress episodes by country, year and signal. Table A2 gives an overview on the role of the different signals in identifying distress episodes, reporting the frequency of each of the three signals in the first year of the distress episode. Figure A1 shows the share of countries that are in debt distress in each year since 1970 and Figure A2 provides a granular perspective by plotting all episodes by country and year.

Table A2. Distress signals and episodes in LICs, 1970-2021

| Distress signal | Definition | Number of episodes |
|---|---|---|
| **Default** | | |
| Direct default measure | Missed payment on private creditors (Farah-Yacoub et al.) | 21 |
| Arrears | Arrears exceed 5% of ppg debt for three consecutive years | 27 |
| **Large BOP assistance** | | |
| Rapid disbursements under IMF program | Disbursements exceed 30 percent of quota in first six months of a non-concessional IMF program | 34 |
| **Multiple signals** | | 8 |
| **Total number of episodes** | | **90** |

*Note:* This table lists the different debt distress signals, their definitions, and the number of distress episodes that are triggered by each of the distress signals. For a distress episode to be triggered by a given signal in the first three rows of this table, the signal needs to be the only signal to indicate distress at the onset of the episode. Episodes triggered by multiple signals are indicated in the fourth row of this table.

Figure A1. Share of LICs in external debt distress, 1970-2021

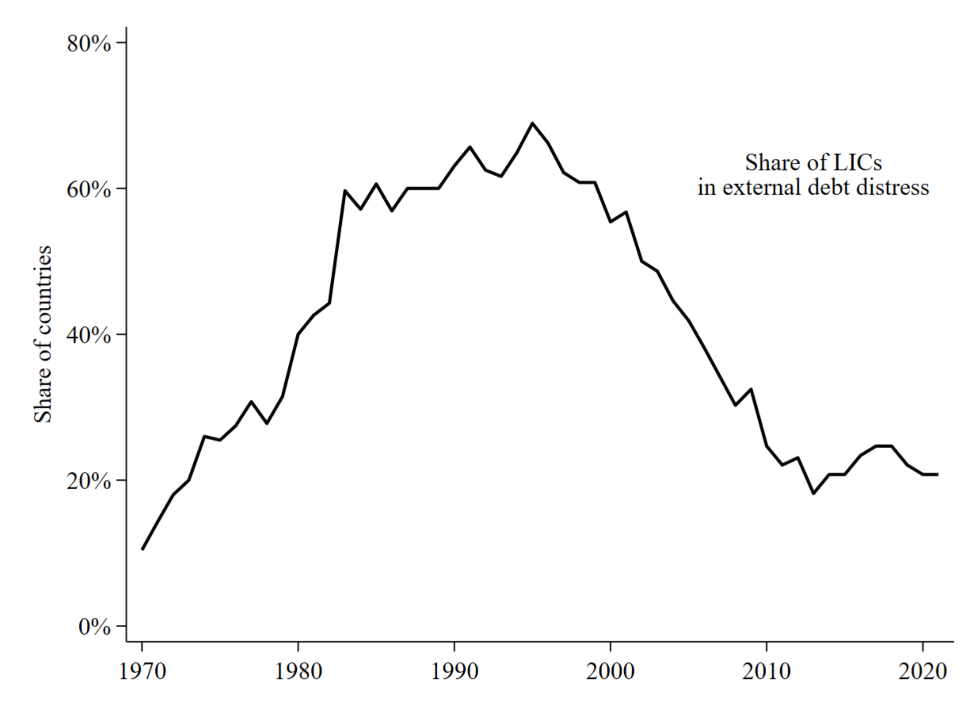
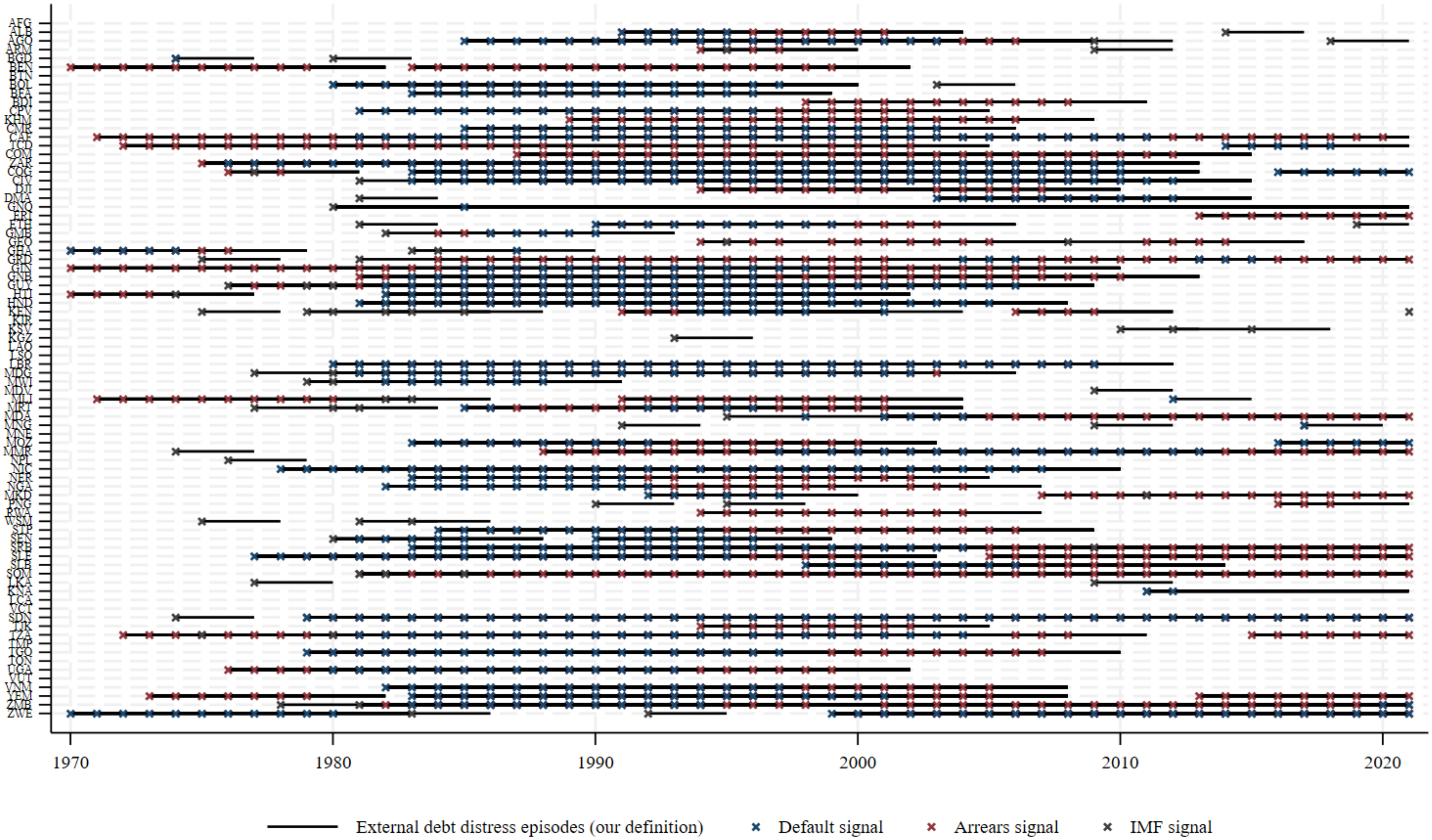
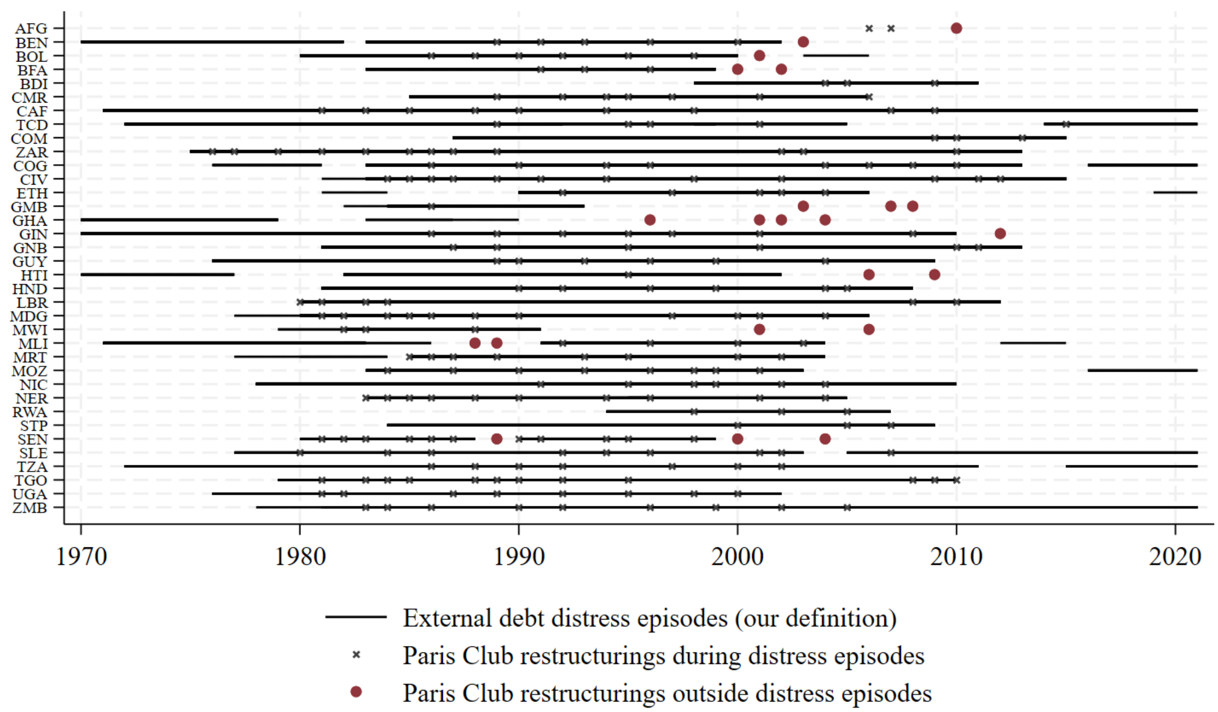


Figure A2. External debt distress episodes by country, signal and year, 1970 – 2021



**Debt restructurings with the Paris Club:** As discussed in the main text, our measure of external debt distress differs from much of the existing literature by not considering debt restructurings as independent signals of debt distress episodes. Here we show that the exclusion of Paris Club restructurings leads to better timing of distress signals without causing the loss of valuable information on distress episodes. For this purpose, we compare our debt distress measure based on payment defaults and large-scale IMF balance of payments assistance with data on 295 Paris Club debt restructurings in low-income countries from Horn et al. (2022).

Figure A3. External debt distress episodes and Paris Club restructurings



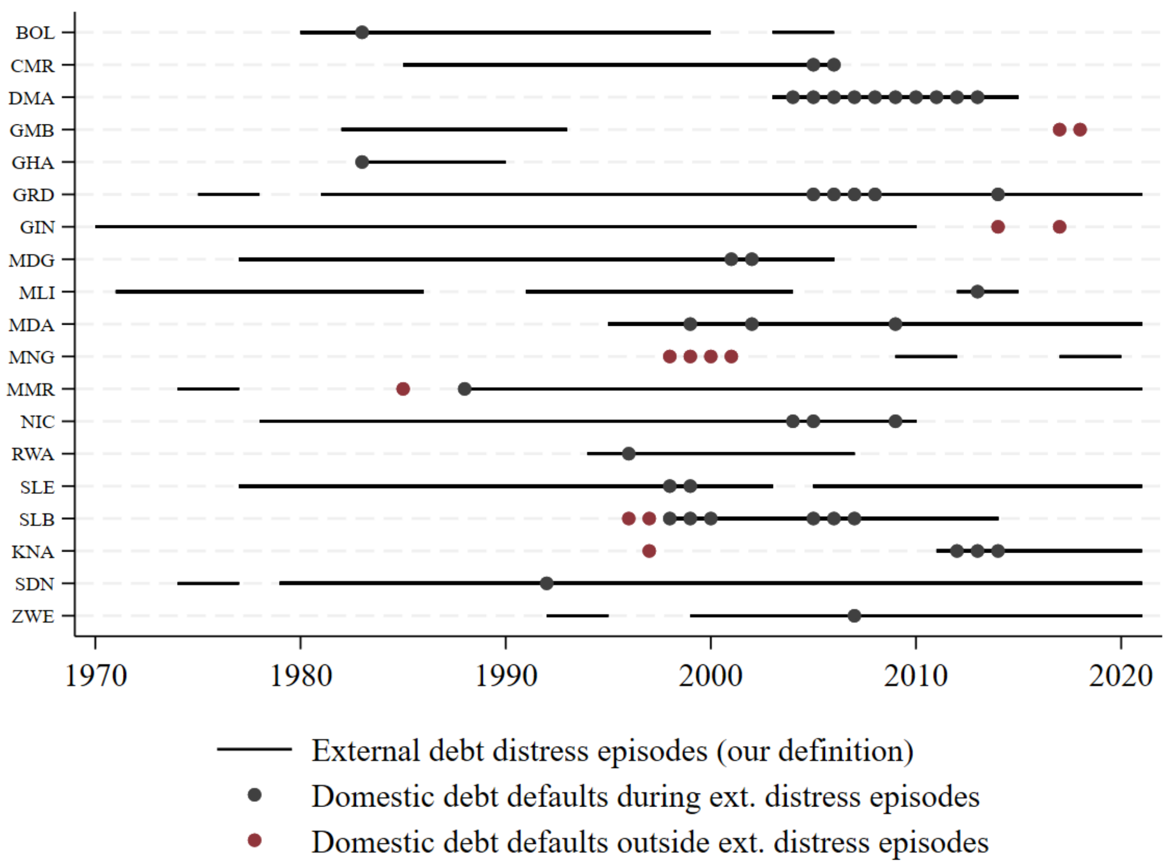

Our comparison reveals that only 25 or 9 percent of the 295 Paris Club debt restructurings occurred outside of our external debt distress episodes. Figure A3 plots this comparison for the subset of HIPC countries, in which the large majority of Paris Club restructurings have occurred. The Figure shows that Paris Club restructurings outside of external debt distress episodes seem to lag rather than lead our distress episodes. This is expected given that restructurings, as discussed in the main text, tend to mark the end of a distress episode rather than its onset. An inspection of Paris Club restructurings outside of our distress episodes confirms this point. The large majority of these restructurings treated arrears and debts that had been in default for multiple years and in countries that had already managed to clear arrears with other creditors and even regain market access. We conclude that the exclusion of Paris Club restructurings increases the timing accuracy of our distress measure without risking to lose valuable information on debt distress episodes.

**Debt restructurings with private creditors:** Debt restructuring events with external private creditors fall into our external debt distress episodes by construction. In the dataset of Farah-Yacoub et al. (2023), on which we rely to construct distress signals for defaults on private creditors, countries enter default by missing payments and exit default by concluding a comprehensive debt restructuring with their creditors. As discussed in the main text, restructurings with private creditors therefore mark the conclusion rather than the onset of debt distress episodes in our data.

**Domestic debt distress events:** To measure domestic debt distress episodes, we build on an IMF dataset of 67 domestic debt restructurings in low-income countries from 1970 to 2020 (IMF 2021a). Since there is no systematic data on domestic *defaults* for our full sample of low-income countries, we map these restructuring events into defaults by imposing a timing assumption. In line with recent empirical research by Erce, Mallucci and Picarelli (2022), we assume that domestic defaults occur one year prior to the restructuring.

Figure A4 shows how our measure of debt distress episodes would change if we considered domestic debt defaults as additional signals of debt distress. The Figure documents a remarkable overlap between the external debt distress episodes identified by us and the domestic debt defaults. Out of 67 domestic debt defaults, only 12 cases or 18 percent fall outside of our external debt distress episodes. In Myanmar and the Solomon Islands, domestic defaults simply extend or precede existing external distress episodes, without creating new episodes. Overall, the inclusion of domestic debt defaults leads to only four additional distress episodes (one each in The Gambia, Guinea, Mongolia and St. Kitts and Nevis). As a result, and for all practical purposes, the composite measure of total public debt distress episodes strongly resembles our existing measure of external debt distress episodes.

Figure A4. External debt distress episodes and domestic debt defaults



## Appendix A3. Predictors of debt distress

This appendix section gives detailed descriptions of the sources and variable transformations for each of the 28 predictor variables that enter our model selection algorithm. Before each of the 28 variables is discussed in turn, we describe two general data collection and cleaning procedures that we apply to multiple predictor variables.

**Filling data gaps from old vintages:** To maximize the country and time coverage of our analysis, we do not just collect data from the most recent releases of our sources but also make use of data points published in earlier vintages. In the World Bank's International Debt Statistics, for example, data for countries that graduate from DRS reporting obligation is removed from subsequent vintages. To the extent that these countries fall under our definition of a low-income country (see Appendix A1), these data points remain relevant to our analysis. To fill data gaps with information from old vintages, we use the data collected by IMF and World Bank (2017) and from Horn et al. (2024).

**Winsorizing:** Macroeconomic and financial indicators for low-income countries are notoriously volatile (see e.g., Aguiar and Gopinath 2007). To limit the influence of extreme outlier observations and measurement error on our analysis, we winsorize all predictor variables at the 1$^{st}$ and the 99$^{th}$ percentile unless their scale is naturally bounded. Specifically, we winsorize as follows:

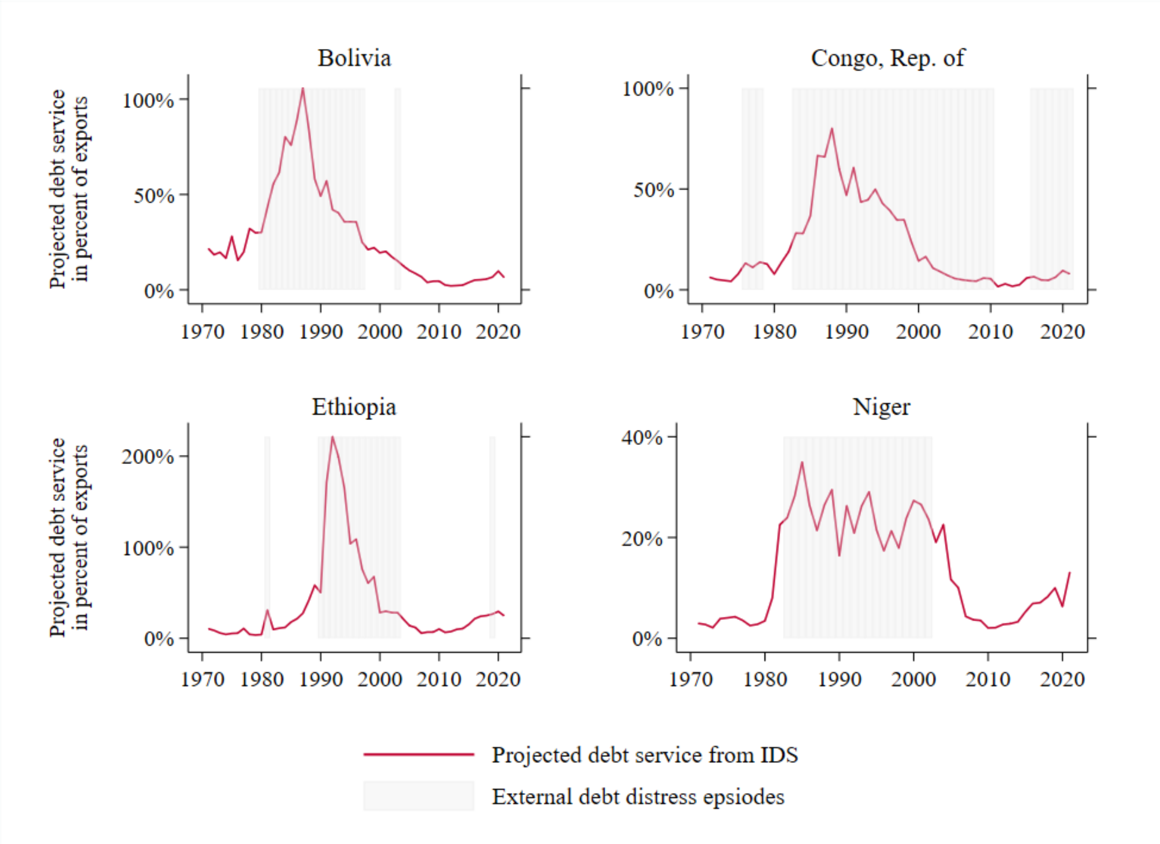
- CPIA, GDP per capita, world growth, the US 10 year yield and the credit history variables are not winsorized.
- Remittances, nominal GDP, as well as the debt stock and flow variables have a natural lower bound at zero, and are hence only winsorized at the 99$^{th}$ percentile.
- All other variables are winsorized at the 1$^{st}$ and 99$^{th}$ percentile.

## **Debt indicators**

**External debt stocks:** This variable measures the economy's external public and publicly guaranteed debt stock. Public and publicly guaranteed debt includes all long-term obligations of public debtors, including the national government, public corporations, state-owned enterprises, development banks and other mixed enterprises, political subdivisions (or an agency of either), autonomous public bodies, and external obligations of private debtors that are guaranteed for repayment by a public entity. Long-term debt is defined as debt with an initial maturity of more than 1 year. External debt is defined on a residency basis. This data is from the World Bank's International Debt Statistics (series code "DT.DOD.DPPG.DT"). We scale the debt stock by nominal GDP (from the World Bank WDI database "ny_gdp_mktp_cd" and the IMF WEO database "NGDPD") and by the country's exports of goods and services (from the World Bank WDI database "ne_exp_gnfs_cd" and the IMF WEO database "BXS_BP6").

**Debt service due:** Our measure for the debt service is defined as the sum of all interest and principal payments on public and publicly guaranteed external debt that the country is required to make in a given year. To construct this variable, we use an internal World Bank IDS variable that projects debt service due in year t+1 on the basis of the disbursed and outstanding debt stock in year t. Crucially, this variable only uses information available in year t to project debt service payments to be made in year t+1 and therefore enters our debt distress prediction exercise with a lead. We scale this measure with exports of goods and services in year t from the World Bank WDI (series code "ne_exp_gnfs_cd") and the IMF World Economic Outlook database (series code "BXGS_BP6") and with a measure of total fiscal revenue in year t. The measure of fiscal revenue is defined as total general government revenue excluding revenue from grants (from the IMF WEO, series codes "GGR" and "GGRG"). Figure A5 illustrates this variable and its association with external debt distress episodes for selected countries.

Figure A5. Projected debt service payments due – selected country cases




**Net present value of debt:** This variable measures the net present value of all future interest and principal repayments related to the country's disbursed and outstanding, public and publicly guaranteed external debt stock. This measure is derived by the World Bank IDS team by applying a constant discount rate of five percent. We scale the net present value by nominal GDP (from the World Bank WDI database "ny_gdp_mktp_cd" and the IMF WEO database "NGDPD") and by the country's exports of goods and services (from the World Bank WDI database "ne_exp_gnfs_cd" and the IMF WEO database "BXS_BP6").

**Total public debt:** To build a data series on total public (domestic plus external) debt, we combine data from multiple sources. Our starting point is total general government debt to GDP from the IMF's WEO database (series code "GGXWDG_NGDP"). This series consists of all general government liabilities that require payment of interest and/or principal by the debtor to the creditor at a date in the future. This includes debt liabilities in the form of SDRs, currency and deposits, debt securities, loans, insurance, pensions and standardized guarantee schemes, and other accounts payable. Since this data is only available starting from the 1990s for most LICs, we supplement this series with data from Abbas et al. (2010) and from Reinhart and Rogoff (2011) for earlier periods.[26] We scale this variable with nominal GDP (from the World Bank WDI database “ny_gdp_mktp_cd” and the IMF WEO database "NGDPD").

**Public debt interest payments:** This variable measures interest payments on total public (domestic plus external) debt. Our starting point for this variable is the 2023 update to the Public Finances in Modern History Database compiled by Mauro et al. (2015). To expand coverage to all low-income countries in our sample, we fill existing coverage gaps with data from the IMF WEO database (using series “dsi”). As with external debt service payments, we scale variable by both exports and government revenue (see details above).

**Measures of institutional quality and other country characteristics**

**CPIA:** Our primary measure of a country’s institutional quality is the World Bank's Country Policy and Institutional Assessment, which is conducted annually for all of countries borrowing from the World Bank. Assessment criteria are grouped in four clusters: (a) economic management; (b) structural policies; (c) policies for social inclusion and equity; and (d) public sector management and institutions. Ratings for each of the criteria reflect a variety of indicators, observations, and judgments that focus on the quality of each country's current policies and institutions. This data is publicly available only since 2005, and only for countries eligible for concessional lending from the World Bank.

[26] A potential drawback of splicing data from different sources is that the consistency of institutional coverage cannot always be ensured for early years of the sample (see Abbas et al. 2010 for details).

**GDP per capita:** This variable is defined as the log of nominal gross domestic product divided by midyear population and is taken from the World Bank's World Development Indicator database (series code "NY.GDP.PCAP.CD").

**Openness:** Our measure for the openness of the economy is defined as the sum of exports and imports of total goods and services as a share of gross domestic product. Data on imports and exports is taken from the World Bank WDI database (series codes "NE.IMP.GNFS.CD" and "NE.EXP.GNFS.CD") and from the IMF WEO database (series codes "BMGS_BP6" and "BXGS_BP6"). Nominal GDP data is from the WDI and the WEO (series codes "NY_GDP_MKTP_CD" and "NGDPD").

**Credit history I:** Our first proxy for the credit history of the country is the number of years since the last external debt distress event (defined as described above). We build this variable for years 1960 to 2021 so that the first value of this variable in 1970 reflects the country's credit history since 1960 (or since the year of independence if the country was a colony until after 1960). Since no data on arrears is available prior to 1970, the distress variable in the 1960s only reflects outright payment defaults on private creditors as identified by Farah-Yacoub et al. (2023) and instances of large and rapidly disbursed IMF balance of payments assistance as defined above.

**Credit history II:** Our second proxy for the credit history of the country follows IMF (2021b) and is obtained as follows: A unit impulse is generated when an external debt distress event (as defined above) occurs and then decays geometrically at a rate of 10 percent. We generate this variable for years 1960 to 2021 and again start from the year of independence if the country was a colony after 1960. Since no data on arrears is available for the 1960s, the distress variable in the 1960s only reflects outright payment defaults on private creditors as identified by Farah-Yacoub et al. (2023) and instances of large and rapidly disbursed IMF balance of payments assistance as defined above.

**Measures of the domestic business cycle**

**GDP growth:** This variable is defined as the annual percentage growth rate of GDP based on constant local currency. Data is from the IMF's World Economic Outlook database (series code "NGDP_RPCH") and the World Bank's World Development Indicator database (series code "ny_gdp_mktp_kd").

**Inflation:** Our inflation variable measures the annual percentage change in the consumer price index and is taken from the Global Database of Inflation constructed by Ha et al. (2021) using series code "hcpi_a". Our variable is a measure of headline inflation that includes changes in the prices of all goods and services that enter the basket of a representative consumer.

**Measures of the external environment**

**World growth:** This variable is defined as the annual real GDP growth rate for country aggregate "World" from the IMF WEO database (series code "NGDP_RPCH")

**US interest rates:** This variable is defined as the market yield on US Treasury Securities at 10-year constant maturity. Data is from the St. Louis Fed (series code "DGS10").

**Remittances:** This variable is defined as all personal remittances received in percent of recipient country GDP. Personal remittances comprise both personal transfers and the compensation of employees. Data is from the World Bank World Development Indicators (series code "BX.TRF.PWKR.CD.DT") and is scaled by nominal GDP (also taken from the WDI database, series code "NY.GDP.MKTP.CD"). In practice, remittances are challenging to measure and Appendix Section A4 discusses some of the related challenges in greater detail.

**Foreign exchange reserves:** This variable measures total foreign reserves in percent of imports. Total reserves include holdings of gold, special drawing rights, and holdings of foreign exchange that are under the full control of the central bank. Data is from the World Bank World Development Indicators (series code "FI.RES.TOTL.CD"). The series is scaled with imports of goods and services from the World Bank WDI (series code “NE.IMP.GNFS.CD”) and the IMF World Economic Outlook database (series code “BMGS_BP6”).

**Foreign exchange rate:** This variable is defined as the annual percentage change in the country’s nominal foreign exchange rate (defined as national currency units per U.S. dollar, period average). Data is from the IMF’s WEO database (series code “ENDA”).

**Foreign direct investment:** This variable measures net foreign direct investment inflows to the reporting economy divided by nominal GDP. FDI flows include investments by non-residents in the reporting economy that acquire a lasting management interest in an enterprise and comprise equity investments, reinvestment of earnings and other long and short-term capital. Net inflows refer to new investment inflows minus disinvestment. Data is from the WDI database (series code “BX.KLT.DINV.WD.GD.ZS”) and the IMF WEO database (series code “BFD_GDP”).

**Current account balance:** This variable measures the current account balance in percent of GDP. The current account balance is defined as the sum of net exports of goods and services, net primary income, and net secondary income. Data is from the World Bank WDI database (series code “BN.CAB.XOKA.GD.ZS”).

**Measures of the political cycle**

**Years in office:** This variable measures the number of years that the Chief Executive of the country has been in office. Data is from the Database of Political Institutions (Cruz et al. 2021), using series code “yrsoffc”). This data covers years 1975 to 2020. To extend coverage to the early years of our sample, we linearly extrapolate this variable for all chief executives which had been in office for multiple years.

**Years left in current term:** This variable measure the number of years that are left in the Chief Executive’s current term. Data is from the Database of Political Institutions (Cruz et al. 2021), using series code “yrscurnt”). This data covers years 1975 to 2020. To extend coverage to the early years of our sample, we make the following two assumptions: (i) We assume that the duration of legislative periods remained the same in each country. (ii) We assume that there were no unexpected and early elections during years 1970 to 1975.

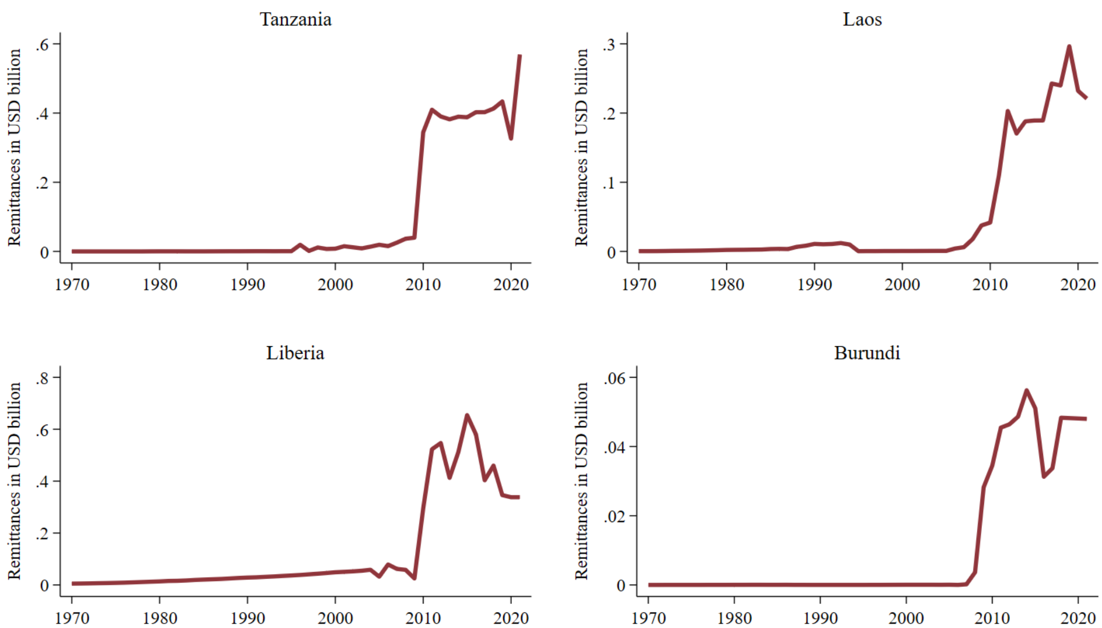
## Appendix A4. Discussion of remittances data

The inclusion of remittance flows as a predictor variable poses specific challenges that we discuss in this appendix subsection. The existing IMF and World Bank debt sustainability framework uses remittance inflows as a predictor variable and our own empirical analysis presented in Section 3 confirms that remittance data has strong predictive power for debt distress episodes in low-income countries.

At the same time, there is growing recognition that remittance data is plagued by severe measurement error (Clemens and McKenzie 2018, World Bank 2023). Over the past two decades, reported remittance flows to developing countries have increased at a pace that is hard to reconcile with fundamental economic trends. As shown in World Bank (2023), inflation-adjusted recorded remittance inflows have increased by around 200 percent since 2000, whereas the total number of migrant workers increased by only 50 percent and world per capita GDP by only 30 percent. Strong discrepancies between fundamental trends and trends in remittance flows point to a potential role for conceptual changes in how remittance flows are recorded and estimated over time and highlight the existence of possibly severe measurement error. Strong discrepancies are also evident when comparing source data on remittance flows (remittance outflows) to destination data (remittance inflows) or when comparing trends in balance of payments data with trends observed in survey data (Clemens and McKenzie 2018).

Discrepancies across sources and over time are likely linked to how countries record and classify cross-border capital transactions. As shown in Clemens and McKenzie (2018), discontinuities in recording practices can often be traced back to the wide-spread roll-out of anti-money laundering (AML) and combating the financing of terrorism (CFT) regulation that most countries started to enforce after the September 11, 2001 terror attacks on the United States. As developing countries started to adhere to stricter standards of cross-border capital flow monitoring, the recorded (and estimated) volumes of remittance flows rose sharply, often showing discrete upward jumps in the year of reform. These patterns are also visible in the remittance flow data from the World Bank Word Development Indicators that we use in our binary debt distress prediction models. In around 20 low-income countries, recorded remittance flows show discrete upward jumps after 2001. Figure A6 illustrates this issue for four selected countries.

Figure A6. Reported remittance flows in selected low-income countries




A specific concern with these measurement issues is that the upward jump in remittance flows coincides with the completion of the HIPC initiative and therefore introduces spurious correlation between high recorded remittance flows and a period of relative tranquility and relative low levels of debt distress.

To account for this measurement issue, we include our measure of remittance flows jointly with a post-2001 time dummy and an interaction term between remittance inflows and the time dummy. Specifically, these three variables (remittances, dummy, and interaction) always enter together as a group in models in which remittances appear. This approach helps to absorb the fundamental shift in how remittances are recorded and the possibly different dynamics that the newly measured variable has on predicting debt distress. While this trio of variables appears in some of our top-performing models, it is an open question whether the quality of this data is sufficient to rely on it for policy purposes such as the LIC DSF.

## Appendix B. Additional model results

Table B1. Predictive performance by predictor variable

| | Coverage ratio since 2000 | Perverse incentive filter | Coefficient size filter | Mean LF | Median LF |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Ext. debt service / exports | 0.96 | 1.00 | 1.00 | 0.33 | 0.33 |
| Ext. debt service / revenue | 0.95 | 1.00 | 0.89 | 0.35 | 0.35 |
| Inflation | 0.97 | - | 0.52 | 0.36 | 0.35 |
| Remittances / GDP | 0.93 | - | 1.00 | 0.35 | 0.35 |
| US 10 year yield | 1.00 | - | 0.99 | 0.35 | 0.36 |
| Reserves / imports | 0.97 | 1.00 | 1.00 | 0.36 | 0.36 |
| GDP p.c. | 0.98 | - | 1.00 | 0.37 | 0.36 |
| Credit History 2 | 1.00 | - | 0.97 | 0.37 | 0.37 |
| Public debt / exports | 0.98 | 0.73 | 0.55 | 0.37 | 0.37 |
| Public debt / GDP | 0.98 | 0.88 | 0.51 | 0.37 | 0.37 |
| Commodities terms of trade | 0.97 | - | 0.84 | 0.37 | 0.37 |
| CPIA | 0.99 | 1.00 | 0.81 | 0.37 | 0.37 |
| Openness | 1.00 | - | 0.52 | 0.38 | 0.37 |
| NPV of ext. debt / exports | 0.96 | 0.75 | 0.63 | 0.38 | 0.38 |
| CA balance / GDP | 0.99 | - | 0.41 | 0.38 | 0.38 |
| FDI / GDP | 0.99 | - | 0.48 | 0.39 | 0.38 |
| Nominal ext. debt / exports | 0.96 | 0.63 | 0.53 | 0.38 | 0.38 |
| Change in FX rate | 1.00 | - | 0.18 | 0.39 | 0.38 |
| Years until next election | 0.96 | - | 1.00 | 0.39 | 0.38 |
| Nominal ext. debt / GDP | 0.96 | 0.61 | 0.50 | 0.38 | 0.38 |
| GDP growth | 1.00 | - | 0.93 | 0.38 | 0.38 |
| NPV of ext. debt / GDP | 0.96 | 0.56 | 0.46 | 0.39 | 0.39 |
| World growth | 1.00 | - | 0.08 | 0.39 | 0.39 |
| Credit History 1 | 1.00 | - | 0.15 | 0.40 | 0.40 |
| Years in office | 0.89 | - | 0.18 | - | - |
| Public debt interest / exports | 0.90 | 1.00 | 1.00 | - | - |

*Note:* This table provides additional details on the predictive performance of the different predictor variables included in our model selection exercise. The first column shows the share of country-year observations since 2000 for which data is available (see Section 3.4.2). The second column shows the share of models with a given variable that do not provide perverse policy incentives (see Section 3.4.1). The third column shows the share of models with a given variable that pass the meaningful effect size filter (see Section 3.4.3). Columns 4 and 5 give the mean and median loss function of all models that include a given variable and that satisfy the model selection criteria. Remittances / GDP in all models is included with a post-2001 dummy and interaction term (see Appendix Section A4).

Table B2. Top performing prediction models for five-year ahead debt distress prediction

| | Dep. variable: Incidence of external sovereign debt distress within next five years | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| CPIA | -0.41*** | -0.41*** | -0.41*** | -0.35*** | -0.17*** | -0.19*** | -0.13*** |
| Ext. debt service / exports | 0.66*** | 0.72*** | 0.89*** | 0.90*** | 0.36*** | 0.47*** | |
| GDP p.c. | | 0.59*** | 0.60*** | 0.49*** | 0.43*** | 0.43*** | 0.44*** |
| Nominal ext. debt / GDP | | | -0.36*** | | | -0.22*** | |
| Nominal ext. debt / exports | | | | -0.38*** | | | 0.23*** |
| Inflation | | | | 0.20*** | | | |
| Remittances / GDP | | | | | -5.99*** | -5.92*** | -6.34*** |
| Post-2001 dummy | | | | | 5.82*** | 5.75*** | 6.12*** |
| Remittances / GDP x post-2001 | | | | | -0.40*** | -0.38*** | -0.43*** |
| CA balance / GDP | | | | | | | -0.17*** |
| Credit History | | | | | | | -0.29*** |
| Number of variables | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Loss function | 0.40 | 0.30 | 0.28 | 0.28 | 0.25 | 0.23 | 0.22 |
| False positive rate | 0.41 | 0.30 | 0.29 | 0.28 | 0.28 | 0.23 | 0.23 |
| False negative rate | 0.40 | 0.29 | 0.27 | 0.28 | 0.22 | 0.23 | 0.20 |
| Data coverage since 2000 | 0.96 | 0.93 | 0.93 | 0.91 | 0.88 | 0.88 | 0.89 |
| Number of observations | 899 | 899 | 899 | 899 | 899 | 899 | 899 |

*Note:* This table shows standardized beta coefficients for the top performing prediction models. Each of the models shown in this table minimizes the loss function for a given number of predictor variables. No additional constraints are imposed. In contrast to the tables shown in Section 3, the dependent variable in this table measures the incident of debt distress in any of the next five years. Stars refer to statistical significance of the underlying probit coefficient estimates: * p<0.1, ** p<0.05, *** p<0.01. All regressions include a constant not reported here.

## Appendix C. Background on the IMF and World Bank LIC DSF model

This appendix section provides a primer on the IMF and World Bank LIC DSF model (Section C1) and explains in detail how it differs from our "best parsimonious model" (Section C2).

### Appendix C1. A brief overview on the LIC DSF model

Debt distress predictions in the existing LIC DSF Model are based on several steps. First, probit regressions are estimated to predict the occurrence of external sovereign debt distress in a historic sample of low-income countries. After estimating the regressions, the framework groups countries based on their assessed debt-carrying capacity and compares their projected debt burdens to threshold levels that are derived from the underlying regression. This subsection discusses each of these steps in turn.

The LIC DSF Model relies on four distinct probit regression, each including a different debt burden indicator: the present value of external debt in percent of GDP and of exports, as well as external debt service in percent of fiscal revenue and exports. In addition, each of the four regressions includes the same set of additional predictor variables: the strength of institutions and policies – measured by the CPIA – and measures of the country's growth, reserves, remittance inflows, and global GDP growth. Table C1 shows results for these four probit regressions as reported in IMF and World Bank (2017).

Based on these regressions results, each country's debt carrying capacity is assessed through a composite index (CI). The CI is the weighted sum of a country's 10-year average CPIA, remittances over GDP, reserves over imports and the national and global growth rate, with the estimated coefficients from the above regressions as weights, averaging across the four equations. The CI is used to classify countries into three groups with weak, medium or strong debt carrying capacity, which correspond to the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentile of the index. Once debt carrying capacity is determined, projections of the country's debt burden indicators are compared to group-specific debt thresholds. These group-specific debt thresholds in turn are also obtained from the probit regressions by inverting each regression for a given cut-off probability of p* and the group-specific threshold level for the CI (see IMF and World Bank 2017 for details).

Table C1. Probit Regressions underlying the existing LIC DSF Model

| | PV of debt as a percent of | | Debt service as a percent of | |
|---|---|---|---|---|
| | GDP | Exports | Revenue | Exports |
| Debt burden indicator | 1.541*** | 0.359*** | 3.745*** | 3.541*** |
| CPIA | -0.400*** | -0.381*** | -0.362*** | -0.395*** |
| Domestic growth | -3.081* | -2.853* | -3.001 | -1.942 |
| Reserves | -4.223*** | -4.591*** | -3.696** | -3.699*** |
| Reserves ^ 2 | 3.953* | 4.582** | 3.743 | 3.683* |
| Remittances | -2.235** | -2.282*** | -1.635* | -1.934** |
| World growth | -12.40*** | -14.09*** | -13.84*** | -13.75*** |
| Constant | 1.310*** | 1.331*** | 0.979* | 1.148** |
| Observations | 409 | 403 | 343 | 380 |
| Pseudo R-sqaured | 0.169 | 0.174 | 0.190 | 0.184 |
| Log-likelihood | -150.1 | -145.7 | -116.8 | -135.2 |
| BIC | 348.4 | 339.5 | 280.2 | 317.9 |

*Note:* This table shows results for four different probit regressions, estimated for years 1970 to 2014. The dependent variable in all four regressions is a binary indicator for external sovereign debt distress episodes. Each regression uses a different debt burden indicator. See text and IMF and World Bank (2017) for details.

A country is classified as being at high risk of external debt distress if any of the four thresholds for the external debt burden indicators are breached in the baseline scenario. A country is classified as being at moderate risk of external debt distress if any of thresholds for the external debt burden indicators are breached in stress test scenarios, but not in the baseline scenario.[27] Finally, a country

[27] The LIC DSF framework includes a total of seven standardized stress tests in which baseline projections are subjected to different macroeconomic and financial shocks, e.g., a reduction in real GDP growth or exports. See IMF and World Bank (2017) for details.

is classified as being at low risk of external debt distress if none of the thresholds for the external debt burden indicators are breached in any of the stress test scenarios, or in the baseline scenario.

**Appendix C2. Differences between BPM and existing LIC DSF framework**

Before providing a systematic comparison between the predictive performance of the best parsimonious model derived in Section 3 and the existing LIC DSF framework, it is useful to summarize the key differences between both frameworks.

First, the existing LIC DSF is based on four different probit regressions obtained from an ad-hoc model selection process. In contrast, this paper has introduced a systematic model selection algorithm based on a comparison of out-of-sample predictive performance to choose a single probit model with high predictive accuracy and high amenability for policy use. And while the LIC DSF relied on a linear loss function with particular emphasis on avoiding false negatives, our model selection is based on a quadratic loss function aimed at generating balanced false negative and false positive rates (see Section 3).

Second, this paper has relied on a refined and simplified definition of external sovereign debt distress episodes. As explained in Section 2.1, we focus on actual payment defaults and large IMF programs, whereas the existing LIC DSF framework combines data on defaults, restructuring events and IMF programs with ad-hoc timing assumptions. Table C2 summarizes the different definitions underlying the dependent variable.

Finally, debt distress prediction in the existing LIC DSF relies on a complex multi-step threshold-based approach, in which countries are grouped according to their debt-carrying capacity and then assessed by comparing debt burden projections to group-specific debt thresholds derived from the probit regressions. In contrast, we will see below that a simple single probit regression like the BPM generates predictions of debt distress that are at least as accurate as the LIC DSF model.

This means that assessments of the risk of debt distress can be based very simply and transparently on the predicted probabilities from this single regression.[28]

Table C2. Definition of external debt distress episodes in LIC DSF and this paper

| **LIC-DSF Model** | **This paper:** |
|---|---|
| **Payment defaults on private creditors:**<br>Data from S&P and Catao & Milesi-Ferreti, whenever available | **Payment defaults on private creditors:**<br>Comprehensive data for all low-income countries from Farah-Yacoub et al. (2023) |
| **Arrears signal:**<br>Arrears > 5 percent of external ppg debt stock for at least three consecutive years | **Arrears signal:**<br>Arrears > 5 percent of external ppg debt stock for at least three consecutive years |
| **Large IMF lending:**<br>Rapid disbursements > 30 percent of quota, all types of programs considered | **Large IMF lending:**<br>Rapid disbursements > 30 percent of quota, only GRA-financed programs considered |
| **Restructurings with private external creditors**<br>Data from Cruces & Trebesch (2013) episode assumed to start one year prior | |
| **Restructurings with Paris Club creditors**<br>Data from Das et al. (2011) episode assumed to start one year prior | |

Note: This table shows the definition of external sovereign debt distress episodes in the existing LIC DSF model (IMF and World Bank 2017) and our refined definition in this paper.

[28] Of course, the BPM could also be applied through a threshold-based approach. In this case, a country index of debt carrying capacity could be constructed based on the weighted sum of the CPIA and reserves over import ratio, with weights equal to the estimated slope coefficients. A grouping of countries with strong, medium and weak debt carrying capacity could be derived based on the 75th, 50th and 25th percentile of the country index. Threshold levels for debt service over exports would then be derived by inverting the probit regression for the chosen level of p*.