FDL Publication Series:
Reforming Debt Sustainability Frameworks

Lifting the hood of the LIC-DSF to revamp its accuracy and transparency

Clemens Graf Von Luckner

Policy Note 18

October 2024

How to cite this paper: Graf Von Luckner C. : "Lifting the hood of the LIC-DSF to revamp its accuracy and transparency", Finance for Development Lab. Policy Note 18. Link
Keywords: Sovereign Debt, Debt Sustainability, Economic Development, Low-Income Countries
JEL Codes: F30; F34; G15, H63

About the Finance for Development Lab

The Finance for Development Lab is an independent non-profit, non-partisan think-tank dedicated to building a fairer and more effective architecture for international finance. Acting as a hub for policy discussions, the Lab collaborates with think tanks, researchers, and other key stakeholders across the Global South to generate constructive ideas, craft innovative proposals, and influence global policymakers, with a particular focus on G20 countries and Bretton Woods institutions.

The Lab is housed at CEPREMAP, a leading French research institute located within the Paris School of Economics, and is supported by the Bill & Melinda Gates Foundation. Its work is directed by the Steering Committee, a group of about fifteen experts and institutions in Africa, Latin America, and Asia. The founding members are individual experts and the following institutions.

While the steering committee members discuss and shape the research, reports from FDL only represent their authors' views.

Steering Committee Institutions

Table of Contents

Executive Summary

The Debt Sustainability Framework for Low-Income Countries (LIC-DSF), developed by the IMF and World Bank, serves as a vital instrument for assessing debt-related risks in vulnerable economies. With many low-income countries experiencing escalating debt pressures — particularly in the wake of the COVID-19 pandemic and higher global interest rates — the importance of the LIC DSF has grown significantly.

The LIC-DSF uses a statistical model to predict the risk of debt distress, influencing policy decisions and access to financing for low-income countries. While most of the attention of civil society has focused on macro-financial projections, a key component of the DSF is its risk model: when countries are at low, moderate or high risk of debt distress. This risk model relies on learning from past episodes of distress to be able to detect new ones. As a result, what is considered "risky" crucially depends on the way past information is incorporated.

In the current version of the LIC-DSF, the procedure follows steps that lose information, making it inefficient and its classification opaque. The paper documents a series of important limitations of the current framework. First, it groups together countries with vastly different economic conditions into the same risk categories. The central index, the "debt-carrying capacity", lacks transparency and allows undisclosed manipulations of risk assessments based on staff judgment. It also creates a number of thresholds that may lead to abrupt changes in risk classification, not mirrored by realities on the ground. Finally, a technical choice was made to emphasize the importance of reducing missed crises relative to "false alarms". As a result, developing countries can have tighter limits to their ability to invest than what would be warranted under a more balanced framework.

The paper explores three proposals to improve the framework: 1. allow automatic model selection instead of relying on human judgment to predict crises; 2. remove the grouping of countries; and 3. achieve better balance in risk weights. In this improved framework, the crisis detection model would be based on a smaller set of variables and be more resilient to structural changes in global financing conditions. Additionally, instead of grouping countries by "debt-carrying capacity" and losing information on each country's probability of default, each country would be assigned a risk assessment for distress based on a small set of structural and debt-related variables. Finally, policy decisions would be based on an assessment of what level of risk is acceptable: specifically, a probability threshold above which the risk of debt distress would be deemed "high".

These changes aim to create a more refined, transparent and accurate, framework that better serves the needs of low-income countries, all while maintaining fiscal responsibility. With the 2026 review of the LIC-DSF on the horizon, it is crucial for policymakers and civil society to critically assess the framework, ensuring it effectively strikes a balance between development goals and responsible debt management.

Introduction

The Debt Sustainability Framework for Low-Income Countries (LIC-DSF) is a crucial instrument in shaping the economic trajectory of the world's poorest nations. Jointly developed by the International Monetary Fund (IMF) and the World Bank, the framework serves as a cornerstone for assessing debt-related risks in the world's most vulnerable economies. The LIC-DSF's risk assessments have far-reaching implications, primarily influencing decisions by the Bretton Woods institutions and borrowing countries, as well as those of other official creditors and private creditors.

In recent years, the LIC-DSF has gained even greater prominence. Many low-income countries face mounting debt pressures, with several falling into financial distress. The COVID-19 pandemic and the subsequent global economic slowdown have exacerbated these challenges, straining government finances and highlighting the precarious debt positions of some nations. Simultaneously, there has been a shift in the debt landscape of low-income countries: increased financing from non-traditional official creditors and an ever-greater reliance on domestic and international private capital markets, with consequential exposure to new volatilities and risks. Interest rates have risen to levels not seen over the past two decades, bringing these vulnerabilities into sharp focus and underscoring the critical importance of robust debt sustainability analysis.

The LIC-DSF's role and impact

The LIC-DSF's debt distress signal plays a pivotal role in determining the borrowing capabilities of low-income countries' governments and shapes countries' fiscal space and, thus, their economic development. It influences policy dialogues, investment decisions, and fiscal strategies and even sets benchmarks in debt restructuring negotiations. Its implications are manifold and reach far beyond the IMF and World Bank's own policies:

- For countries rated as being at high risk of debt distress, access to non-concessional financing may be severely limited.
- The poorer countries in this group can access grants rather than (highly concessional) loans, limiting resources for other countries.
- The LIC-DSF's reach goes beyond the multilateral system: OECD countries use the LIC-DSF as a guide on whether to provide official export credit guarantees (through the "Agreement on Sustainable Lending Practices"), and there is anecdotal evidence that DSF ratings also influence private investors' decision-making.

But… what is this framework, exactly?

The LIC-DSF essentially consists of three fundamental building blocks:

(1) **It is a model of the future seeking to project the most likely evolution of debt variables, based on assumptions of growth, fiscal policies, external conditions, etc**. (This is often what gets the most attention: how projections are made; whether they are "over-optimistic", etc.)
(2) **Crucially, it is also a statistical model of risk.** Projecting debt variables (debt service, stock, etc.) is of course important, but whether or not those projections are signals of unsustainability requires an assessment informed by data from the past, packaged as a statistical model that predicts whether a given country in a given year is at risk of being in distress. The reason the

DSFs for Low-Income and Market-Access Countries (MACs) are so different is in large part due to the stark differences in their respective statistical models.

(3) **Modules and stress tests allow us to "shock" the DSA and look at specific scenarios.** This is different from prediction exercises because it requires choices regarding which risks to pay attention to: in an uncertain world, putting the emphasis on a given set of shocks will influence both policy choices and what vulnerabilities are pre-eminent.

This note focuses on the second building block[1]. This block – the statistical prediction model – tends to receive a lot less attention in policy debates, despite it being the centrepiece of the DSA and an important effort in the ongoing IMF/World Bank review. It also features a fair degree of technical complexity, which is why we propose "opening the hood" of the LIC-DSF's prediction machinery with a deep dive into its mechanics. This paper presents the information with a pedagogical purpose, simplifying it slightly while preserving the essence of the methodology. It also highlights potential improvements and explains why they matter.

Amid the 2026 LIC-DSF's 2026 review, policymakers in developing countries, civil society at large, and the board members of the main International Financial Institutions (IFIs), must understand what the model predicts, where it may fall short, and how it can be improved. This understanding is essential for:

- Enabling more meaningful participation in the review process.
- Advocating for necessary improvements to the framework.
- Developing a more robust, inclusive, and effective system that better serves the needs of low-income countries and the global financial system as a whole.

This note is divided into two sections: the first section explains the estimation of the widely recognized debt- and debt service thresholds that determine the risk of debt distress. While the model design may seem technical, it has many direct policy implications. **The second section then explores areas for reforms, arguing that the current threshold approach is both inefficient and opaque.** It is inefficient because grouping countries with vastly different characteristics into three debt-carrying-capacity-levels ultimately results in a loss of critical information. It is also opaque because the methods used to create these groups can be, and as we will show at times are being manipulated. The review should seek improvements on both fronts.

This note further argues that the LIC-DSF needs to better reflect global policy priorities on the inevitable trade-off between enabling investments and reducing debt vulnerabilities. Countries' ultimate risk ratings must mirror policy priorities: balancing the risk of failing to predict crises on the horizon, with the risk of constraining countries' access to essential financing. The second section includes suggestions for procedures to define these priorities.

[1] Other FDL notes in this series will cover (1), looking at optimism in the context of the LIC-DSF, and 3, looking at risks stemming from climate change.

A. How does the Low-Income-Country Debt Sustainability Framework generate its short-term debt distress signal?

For simplicity purposes, this illustration only focuses on the key aspects of the LIC-DSF's machinery, omitting details such as multiple debt indicators and distinctions between low and moderate risk levels. This section thus follows the current DSF: in a way, it is a guide for the interested reader who would – like most – have skipped Appendix 3 very far down the 2017 review (you would be excused). It makes the case that this seemingly obscure appendix is, in fact, central: its role is to predict when debt distress will occur. Some critics contend that the IMF and the World Bank tend to ignore risks: while others, often in developing countries, decry its conservativeness. This section will show how, ultimately, this tension stems from a statistical model, and explains which decisions were made when it was developed.

Step 1: Predicting a probability of distress

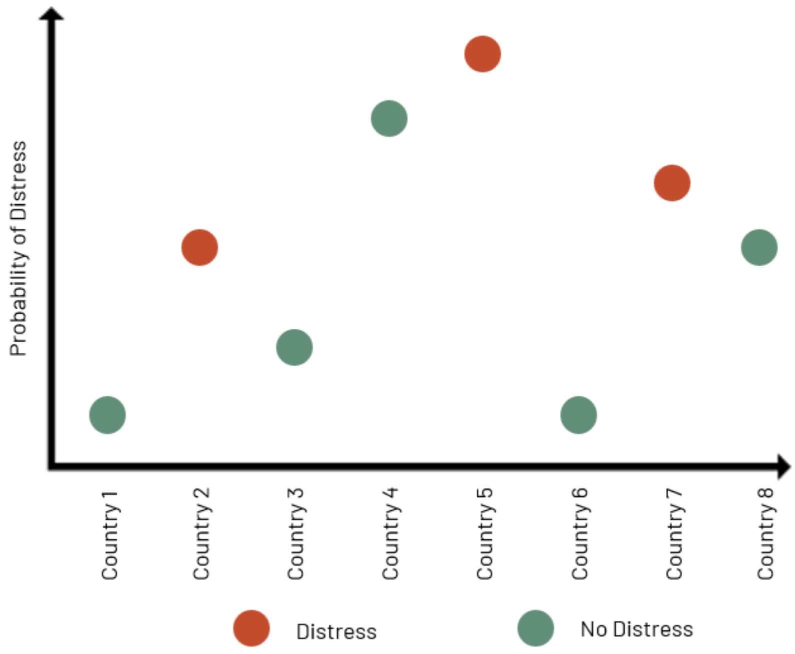
The LIC-DSF's short-term assessment relies on a "Probit model", whose objective is to provide a probability of a country experiencing distress in the near future, relying on its country characteristics and debt variables. The model takes into account that countries differ in how much debt they can safely take on: their *debt carrying capacity*. Since 2017, the IMF and the World Bank have assessed countries' debt-carrying capacities based on five indicators:

1. *Institutional strength (CPIA)*
2. *The country's growth rate*
3. *Foreign exchange reserves (scaled by imports, both as a ratio and as a quadratic term of that ratio to account for potential nonlinearities)*
4. *Remittances (scaled by nominal GDP)*
5. *World growth*

This is to account for the fact that the same debt stock, let's say 80% of GDP, represents more risk to a country with weak institutions, low growth rates, and low reserves; than for a country with strong institutions, robust economic growth and high FX reserves. Using these five indicators and a debt indicator (say, for example, private external debt to GDP), the model estimates the probability of distress for any country at a given time. The importance of each variable in determining the risk is thereby dependent on how debt distress in the past correlated with each indicator.

Figure 1 illustrates what a predicted probability of distress based on a probit model could look like. Factors like rising debt stocks, weakening institutions, or decreasing reserves increase the probability of distress, represented by dots moving upward in the graph. In some cases, these probabilities will be high, and distress will indeed occur (e.g. Country 5). In other cases, the probability might be high, yet no distress occurs (e.g. Country 4).

Figure 1: Predicted probabilities



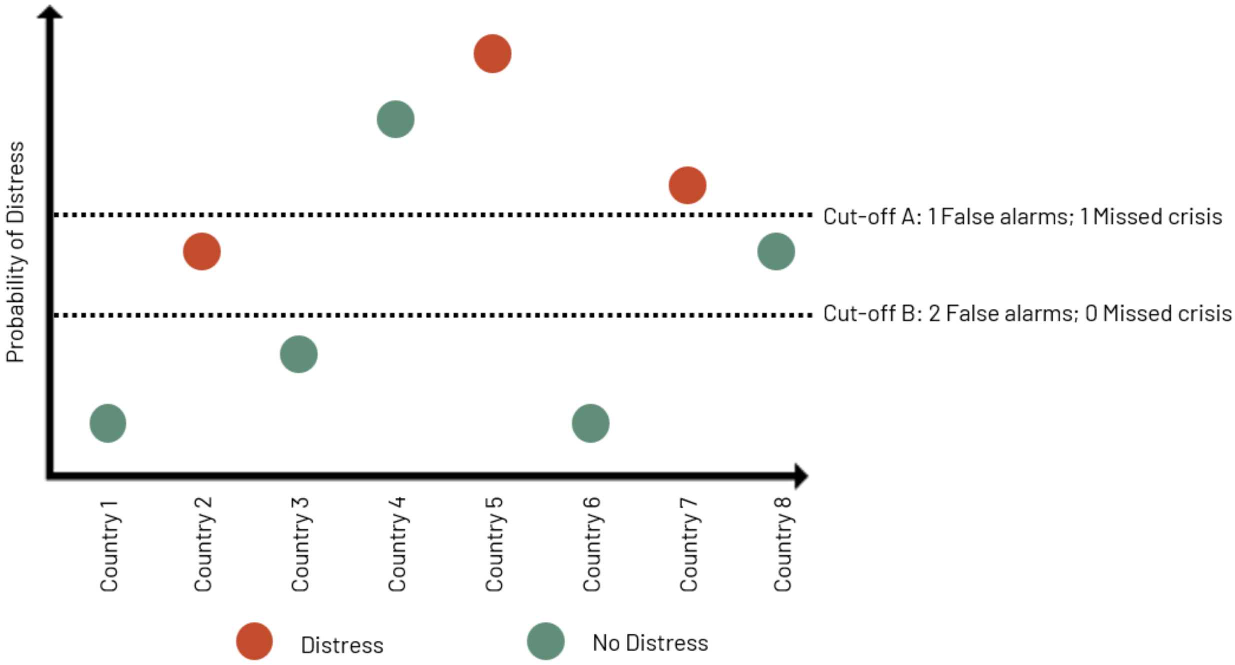
Step 2: Setting thresholds to create buckets of risks and guide policy decisions

One could have stopped there: every country is associated with a given probability of distress, that could be taken into account when lending to this specific country. Policy decisions, however, require clearer distinctions: *when to lend, or not to lend*. This is where classifications and cut-offs come into play, and are used to group countries with similar probability of default.

How these cut-off levels are set depends on the policymakers' preference over missed crises (red dots below the cut-off) and false alarms (green dots above the cut-off), and hence the policymakers' risk tolerance: a risk-averse policymaker would set lower thresholds, potentially limiting productive financing for many countries, but reducing the chance of exacerbating debt crises through continued lending. The ramifications of this seemingly technical detail are vast and discussed below.

To illustrate the trade-offs when setting the cut-off levels, Figure 2 shows two possible cut-off levels and their associated number of missed crises and false alarms.

Figure 2: Predicted probabilities and classification cut-offs



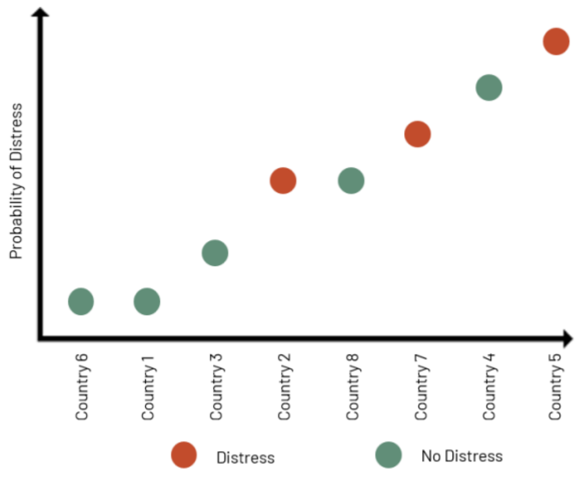
Step 3: Moving debt to center-stage

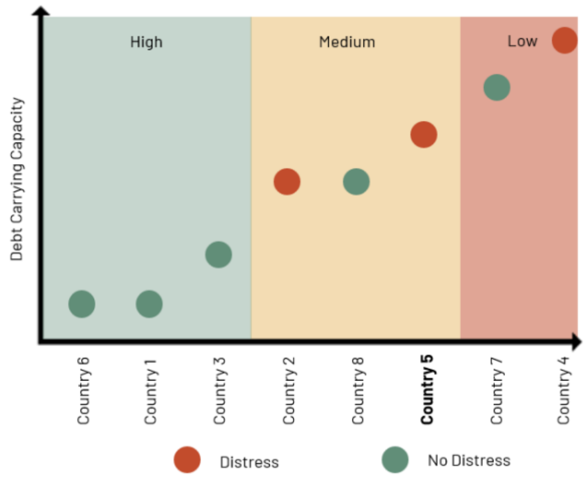
Once these classification cut-offs are in place, the definition of policy rules becomes intuitive. Countries are simply treated according to their position relative to the cut-off. However, a new problem arises: The LIC **Debt** Sustainability Framework is aimed at directing **borrowing** decisions. However, when using probability cut-offs driven by all predictors, debt is not necessarily at the center of the discussion. Indeed, the dots travel north, not only as debt increases, but also, when (world)-growth, remittances or reserves decline, or institutions become weaker. To improve its risk classification, a country might focus on strengthening institutions, promoting growth, or attracting remittances, rather than addressing fiscal policy through spending cuts or revenue increases. To bring the debt variable to the limelight, the LIC-DSF imposes a twofold approach:

1. First, it smooths the other variables by taking their 10-year moving average, thus turning debt-carrying capacity into a slow-moving, *structural* characteristic.
2. The DSF then inverts the probit regression to turn probability thresholds into debt thresholds. To achieve this, the DSF groups the countries by their debt-carrying capacity (DCC), which one can also think of as the general probability of default, ignoring the debt variable. Meaning countries within one group now differ primarily with respect to their debt stock (or flow).

Figures 3 and 4 below regroup our imaginary set of countries to illustrate how a higher probability of distress usually means a lower debt-carrying capacity.

Figure 3 and 4: Predicted Probabilities and Debt Carrying Capacity



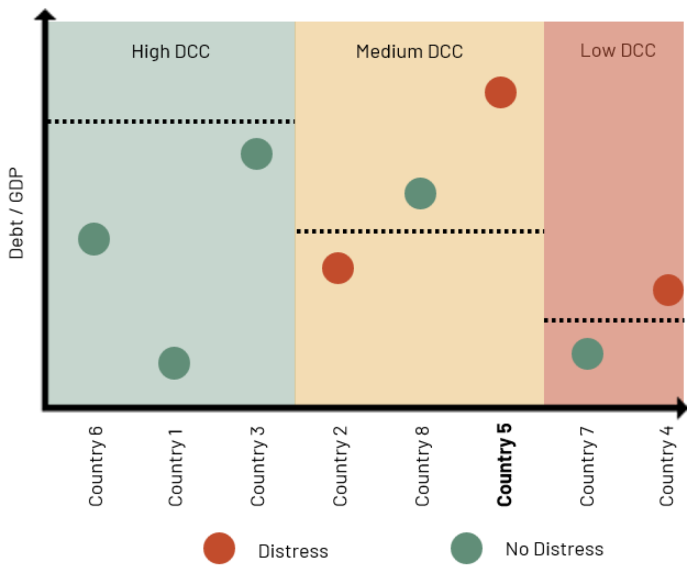


Of course, there are also cases with relatively high debt-carrying capacity and yet a high probability of distress, because of a **very large** debt stock. We shall think of the illustrative country 5 (in bold) as such a case. When ordering the countries based on their predicted probabilities of distress, it falls to the right tail (Figure 3). However once one subtracts the impact of the debt stock on the probability of distress, the country appears around the middle of the distribution (Figure 4).

After grouping countries by debt-carrying capacity to account for structural differences, the level of indebtedness becomes the key variable distinguishing their probability of distress. This means that one can now impose debt thresholds for each subgroup rather than an overall probability cut-off level. Again, these cut-off levels are set within each subgroup such that misclassifications are minimized.

Voilà: the distress prediction *machine* as applied within the LICDSF.

Figure 5: Classification Threshold by Debt Carrying Capacity Subgroup



B. Towards an improved DSF: more informative, more transparent

The methodological choices made during the design of the LIC-DSF significantly influence policy outcomes. While they were made for good reasons, they also have costs. This section proposes three innovations to reform the DSF's prediction machinery: **simplify the framework**, **enhance accuracy, and reduce manipulability.**

The proposed changes are designed to address the fundamental tension inherent in the DSF's dual objectives: supporting LICs in achieving their development goals while minimizing the risk of debt distress. This balancing act is hard to get right, since tighter fiscal policies may reduce debt distress risks, but financial constraints may hinder progress towards the Sustainable Development Goals (SDGs). In fact, this balance should be transparent, in the sense that while adjustments are justified, they should be made explicitly; and it should reflect both risk preferences expressed by the board of the IFIs and the broader evolution of the Global Financial Safety Net. Debt is riskier during times of shrinking aid and high interest rates. Our proposals seek to make this trade-off explicit, and highlight areas of the model most immediately linked to it.

Three proposed improvements to the DSF are:

- Revisit and refine the underlying model used in the DSF.
- Replace group-specific classification cut-offs with a continuous risk index applicable across all countries to reduce manipulability and increase accuracy.
- Since a binary classification is necessary for policy purposes, balance the weights between false alarms and missed crises when determining the optimal threshold.

These changes aim to create a more nuanced, accurate, and flexible framework that better serves the needs of LICs while maintaining fiscal responsibility.

a. A more straightforward model selection

The first section of this note explained the rationale behind the choice of a specific set of variables' used to predict crises, with a probability model. **The focus is now on whether these models and variable selection they rely on can be improved.**

First, to avoid relying too much on the specific set of examples in the data, machine learning can help to select a model which does not "overfit". Graf von Luckner, Horn, Kraay and Ramalho (forthcoming) have investigated the choice of variables used to predict debt distress in LICs and found that the variables used in the current LIC-DSF's probit model represent a seemingly arbitrary variable selection, with many -- in fact, a majority of alternative combinations of variables performing much better at predicting debt distress. . Given how these variables were chosen, this is actually little surprising: in past reviews, the LIC-DSF's prediction model was never tested out-of-sample, although basic statistical theory has long taught us about the risk of "overfitting", that is, choosing a model and a set of predictors that perfectly predict "in-sample" distress events, but are then unable to predict "out-of-sample" events, when presented with new data (Mullhainathan and Spiess 2017). The solution Graf von Luckner et al. (forthcoming) propose is an agnostic machine-learning approach to the model-selection, which build over 500,000 models with every possible linear variable combinations, and compares these models not by asking how well they fit the data, but by asking how well they can predict events "out-of-sample". That is in data the prediction model has not seen before. As a result, the best models are less affected by human bias, and selected on the basis of their predictive performance.

Given the economic costs of misclassifications, the LIC-DSF must rely on the best possible model, not the one closest to the existing one – a result of two decades of retrospectively opaque piece meal adjustments. . Three broad principles should be drawn from that research:

1. **Parsimony vs. Complexity:** Simpler models with fewer variables can achieve predictive power comparable to much more complex models. A model with just three variables (CPIA, external debt service to exports, and reserves to imports) is one that performs remarkably well, given its parsimony, and always outperforms the current model.
2. **Out-of-sample testing:** Any decision made in the model-design phase should be tested for possible effects on out-of-sample prediction of distress episodes.
3. **Data availability and quality trade-offs:** The study highlights the importance of balancing predictive power with data availability. Models that perform well in-sample but require data that is not widely available across low-income countries may be of limited practical use. Policymakers should favor models that use readily available data to ensure broad applicability. Moreover, data quality matters. For instance: The current LICDSF includes remittances as a predictor. Due to changes in regulation driven by the war on terror post 9/11, many cross-border transactions became newly classified as remittances, leading to jumps in the series on remittance data post-2001. Regressions, that do not account for this jump produce meaningless coefficients for prediction purposes: this is the case of the LIC-DSF probit model.

These findings suggest that there is significant room for performance improvement through recalibration of statistical predictions of defaults. Policymakers should consider adopting simpler, more robust models that focus on a few key indicators, especially when adding more complexity does not improve the performance. The trade-offs between model complexity, data availability, and predictive power highlighted in this new research should inform ongoing efforts to refine the LICDSF's probit model.

b. Retire group-specific classifications

As shown in the first section, applying debt thresholds requires inversing the Probit predictions and consequently grouping countries by their debt-carrying capacity to arrive at debt-carrying-capacity-grouping specific debt thresholds. Although the final thresholds might seem to allow for parsimonious policy rules (e.g. country X needs to achieve a debt stock below Y % of GDP), the inversion and consequential grouping has three negative consequences: distorted incentives, information loss, and instability in restructuring cases.

1. Distorted incentives

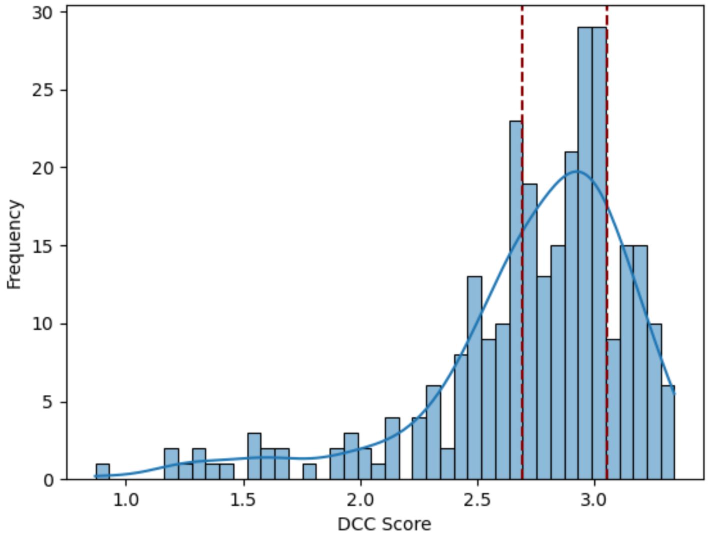
"Why did God create economists? To make weather forecasters look good."

(A time-honored joke often told among economists. Origin unknown.)

When a tiny difference in the Debt Carrying Capacity (DCC) score makes for a large difference in the final risk rating, IMF or World Bank staff might be tempted to "meddle" with data to opaquely adjust the "mechanical rating" to the staff's judgment. The careful reader will remember how the first part of this note introduced the DCC index, based on coefficients multiplied with 10-year moving averages (five years from the past, **five years into the future**) of variables that are notoriously hard to predict. Making small and certainly defendable changes to a five-year prediction that anyways is about as good as a guess might hence seem tempting. Especially when a final risk rating might otherwise seem too optimistic or pessimistic in the eyes of the staff performing the analysis.

And indeed, the data suggests such meddling likely takes place: if the "mechanical" setting of debt carrying capacity scores were indeed mechanical, then the DCC scores produced by it should be smoothly distributed, given the continuous distributions in its components. And where the cut-off thresholds lie, should have no impact on the distribution of DCC scores whatsoever. The data, however, tells a different story. Figure 8 shows significant bumps in the distribution right around the thresholds.

Figure 8 – Debt Carrying Capacity Score Histogram, 2017-2023, with DCC cutoffs



Note: The number of bins is set to 42, to arrive at equal-sized-bin-splits close to the thresholds. The appendix shows other bin-splits for illustration.

This is not as surprising as it seems, and indeed follows a convincing body of literature. There are such incentives for manipulation inside the IFIs, for a broad range of reasons, which numerous scholars have explored (see Lang and Presbitero, 2018, and the works cited therein). What is new is that even after the previous review aimed at reducing the space for discretion, it does not seem to have disappeared. The data presented here suggests that meddling with the DSA is still commonplace, even after the last review, whose goal had been to make staff judgment more explicit.

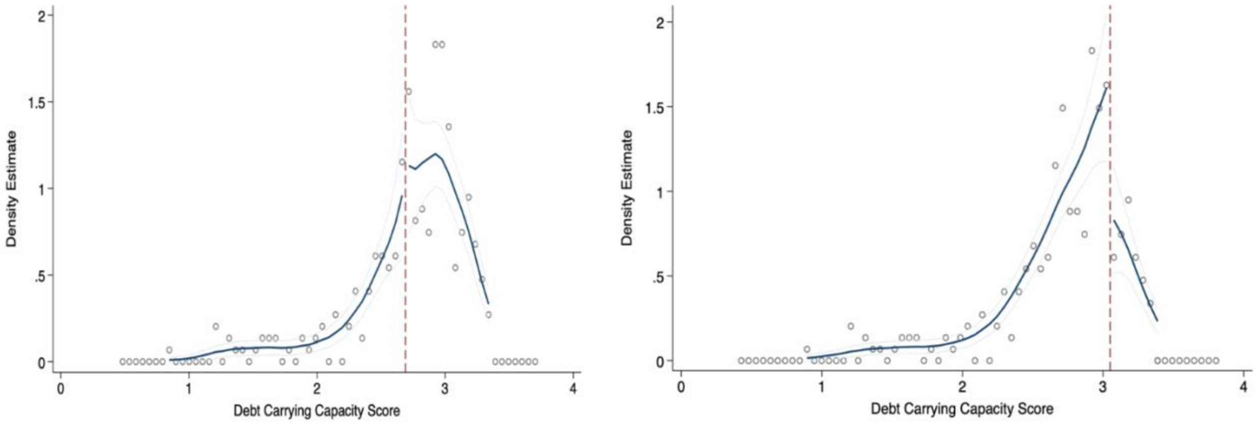
Note that it is of course broadly acknowledged that "use of judgment" by staff is at times adequate. The DSF encourages judgment in several decisions that staff has to make. However, the use judgment is precisely delimited, and its use to be made explicit and kept and communicated separate from "mechanical" results. Adjusting DCC scores to affect risk ratings is neither transparent, nor in any way separate from the mechanical results.

We formally test whether there is a discontinuity in the density of observations around the known cutoffs using a more precise test, known as the McCrary Test (2008). Confirming the histogram's intuition, we find evidence for a statistically significant discontinuity in the distribution around the moderate-to-high threshold, but not for the low-to-moderate threshold. Figure 9 shows the discontinuity graphically. A set of pseudo-tests (where we run the test for DCC values, which do not represent thresholds) can be found in the appendix.

Figure 9 – McCrary Discontinuity Test around the 2.69 and 3.05 DCC threshold

Panel A: 2.69 threshold

Panel B: 3.05 threshold



This analysis suggests that in some cases, staff-adjusted five-year forecasts fed into the model to adjust DCC scores downward, at least when they fall just above the medium-to-strong threshold.[2] This is no proof of any mal-intent of course: it might very well be, that the observed manipulation and the consequent more conservative outcomes are welfare improving. They could for example guarantee a country's access to IDA grants. Or help set more ambitious debt-relief targets in debt restructuring cases, to avoid the all-to-common story of sovereigns relapsing into default soon after a restructuring (Graf von Luckner et al., 2021).

2. Inefficiency of information loss

By design, bunching countries into subgroups inevitably means ignoring valuable available information. Those large buckets group together countries with very significant differences. For example, the "weak" debt-carrying capacity category groups together countries as different as Somalia (with a Gross Domestic Product per capita of 644 US$) and Cameroon (1654 US$ GDP p.c.), while the medium DCC category, imposes the exact same thresholds on countries as different as Madagascar (528US$ GDP p.c.) and Kenya (2,070 GDP p.c.). These countries' GDP per capita illustrates how different they are in terms of their stages of economic development. Indeed, their vastly different DCC scores reflect this. Yet all the differences are ignored when grouping countries over vastly different DCCs in order to[3]. While countries at the extremes of one group can vary greatly, those just above or below a threshold can be quite similar in many regards, yet face drastically different situations – dictated by the LICDSF's group-specific debt limits.

[2] Interestingly, the systematic downgrading seems to go against a practice recorded for the 2005 to 2017 period, when internal documents showed staff more likely to override the mechanical signal in order to upgrade ratings (Land and Presbitero, 2018).

[3] For comparability, all stats are dated back to 2021. GDP per capita is expressed in current USD. Debt to GDP, in line with the LICDSF methodology, measures the net present value of a government's external debt, as a share of its GDP.

This loss of information applies not only to the DCC but also to the final risk rating. Countries rated at "high risk of debt distress" include places where the actual probability of distress is high, but also countries where it could be much lower. The procedure described in the first section prevents from affecting a precise minimum probability, but it is clearly low.

This is why the seemingly inconsistent claims of the LICDSF unduly imposing credit constraints; and claims that the DSF allows for too much debt can in fact both be true, depending on whether a country falls right above or right below a given threshold.

3. Instability in restructuring cases

The existence of thresholds suggests that the debt ceilings used for determining necessary debt relief are non-linear, unstable, and subject to sudden shifts, depending on the difficult-to-foresee recovery trajectory of the defaulted country. Indeed, the DSA, as is now well-known, sets targets for debt relief. The exact thresholds will matter. They thus cause yet another problem that has gained attraction on Wall Street in particular, but more recently, well beyond. **Proximity to thresholds means targets in debt restructurings set on the basis of these thresholds risk to be unstable and subject to sudden jumps.**

It is precisely this discontinuity that motivated the contingent instrument in Zambia's recent restructuring, which dictates that Zambia's interest bill will rise significantly if the country's debt-carrying capacity reaches the medium threshold by 2026. This, as we illustrated above, could happen not only because the country's recovery is accelerating (which would affect growth, reserves, etc.) but, in theory, could be driven by world growth, entirely independent of the country's economic well-being. So far, Zambia was an outlier, since large private external debt restructurings typically happen, as the name suggests, in market-access countries, subject to the MACDSF. This new use case may recur though, given that many low-income countries first accessed capital markets in the past decade.

In summary, the current framework based on country groups (1) opens the gate for non-transparent manipulation of countries' ratings; (2) imposes the exact same treatment for drastically different sovereigns or drastically different debt limits for very similar sovereigns; and (3) is poorly suited for debt-relief envelope-setting in restructuring cases.

An alternative approach that would not rely on debt thresholds, but continuous predicted probit probability thresholds (one may want to call them risk indices). Such risk indices would allow the framework to account for the rich heterogeneity among countries rather than ignore available information by construction. And rather than impose opaque country-specific debt-thresholds, it could feature universal risk-index thresholds. Moreover, deriving debt limits, if and where necessary, could still always be achieved by means of inverting the risk-index function.

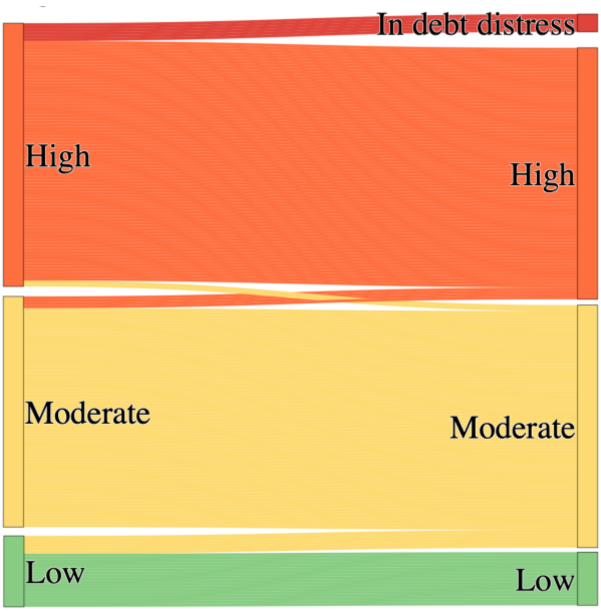
c. Setting classification cut-offs

Whether there exist debt variable-related thresholds or classification based on continuous probabilities, the need for transparent policy rules might eventually impose some classification thresholds. As is illustrated above, when setting such cut-offs, there always exists a zero-sum trade-off between false alarms (false positives) and missed crises (false negatives). As one lowers a threshold, more countries will have debt levels above it and will be considered at high risk, with all its ramifications. Such lower thresholds inevitably mean fewer missed crises but more false alarms. An extreme case helps to illustrate this: with a threshold so low that all countries are above it, all crises would happen in a country that is classified as high-risk. However, all other countries will be false alarms and not fall into distress despite being above the threshold.

To optimize the cut-off levels, the current LIC-DSF minimizes the number of misclassifications of both types. However, as in any such statistical classification system, weights must be applied between two types of errors. The weights applied in the parametrization of the LIC-DSF are so well hidden, many experts are unaware what they are. In fact, they are unmentioned in the LIC-DSF Guidance Note, despite their importance. It is these weights that show how IFIs have clearly been more concerned about "missing crises" than about "false alarms", assigning 2:1 weights on the two error types respectively.[4] Naturally, this leads to thresholds such that false alarms are **very common:** of any given number of countries at high risk of distress, only a small fraction will ever face distress.

Figure 10 below illustrates this. Only a tiny fraction of countries rated to be at high risk of falling into distress were found in distress one DSA out. Of course, distress can be avoided with emergency support such as an IMF program, potentially contingent on a high-risk rating. But consider an alternative scenario where fewer countries would have been deemed at high risk. Some of them would have had more fiscal space and the ability to borrow from multilaterals. This, in turn, could allow more lending and growth. It is likely that a lower threshold, as a result of more balanced weights on the different types of misclassifications, would reduce high-risk ratings and their restrictive consequences without necessarily increasing the number of missed crises. The point is not that one should be a lot more relaxed about the risks of debt crises, but that the balance of risks needs to consider the benefits and costs of over-predicting crises compared to missing some crises.

Figure 10 – Only a small fraction of high-risk cases, face distress by the time of the next DSF



[4] International Monetary Fund—The World Bank, 2017, page. 30

Those considerations need to take into account three factors: the costs of reduced investment, the cost of crises (Farah-Yacoub et al., 2024), and the ability of the global financial safety net to contain them. With low levels of liquidity, and slowly growing MDBs, high tolerance to risk might be optimal. However, if and when the "bigger, better, bolder" agenda for MDBs takes off, an overly conservative DSF would limit the ability of countries to borrow for their growth. Ultimately, balancing those risks is a political decision that should be transparently taken by the boards of Bretton Woods Institutions.

In a world without the negative consequences of a high-risk rating, the risk-averse weight-setting used in the current LIC-DSF would be reasonable, as it ensures that all potential distress cases receive attention. Yet there are also hidden costs of such excessive cautiousness. This is especially the case if this provides incentives for countries to turn from institutions that would take the DSF into account and restrict their credit (such as multilateral institutions and OECD creditors) towards lenders that would lend at higher rates for higher risks. While it is not the case today, IDA could decide to back to its former practice of reducing its envelopes for high-risk countries. In addition, there is a global externality to having too many countries eligible for grants, as it must reduce envelopes for others. Today, with evermore outside creditors using the LIC-DSF as a reference point and given that high-risk ratings cut sovereigns off from crucial non-concessional multilateral financing, setting weights in such a way has damaging consequences. Holding the economic development of sovereigns back, as they are unreasonably often cut off from important sources of financing.

d. A modest proposal: country-specific thresholds and opening a debate on acceptable risks

Here is how the new generation of the LIC-DSF could address the dual criticism of allowing excessive and insufficient debt. Let us consider how a restructuring with continuous probabilities and well-defined thresholds would work. For simplicity, we assume a debt distress prediction model that considers only debt service and institutional strength (measured by CPIA):

1. For Zambia, given its CPIA score of 3.2, the framework – through the threshold setting mechanism described below – might determine that reaching a debt service to exports ratio of 20% is acceptable.
2. For Ghana, with a CPIA score of 3.4, the acceptable debt service/exports ratio might be higher, say 25%.
3. In both cases, the IMF would prescribe setting the probability of redefault (i.e., the risk index) at 15% post-restructuring, which the board considers an acceptable level for "moderate risk" after restructuring.

This approach can also be applied in program design. For instance, the IMF could set a policy that it can only lend to a country with a default probability exceeding 20%, with this probability clearly expressed. This defines the program parameters. Importantly, improving various economic indicators will change things, but without abrupt changes. For example, improvements in CPIA, or foreign exchange reserves, might provide slightly more fiscal space, yet the relationship remains continuous - there are no sudden jumps in risk assessment or lending capacity.

This continuous probability approach offers several advantages: it improves transparency and flexibility by enabling nuanced assessments that reflect the unique circumstances of each country.

It also makes debt thresholds much more stable, with debt variable ceilings changing smoothly without any sudden jumps caused by changes in the group classification. Finally, it increases fairness, ensuring that countries with similar risk profiles are treated equitably, regardless of whether they fall above or below an arbitrary threshold.

As explored in the last section of this chapter, those "acceptable probabilities" would be transparent. Importantly, this choice between more credit and risk is complex and depends on each country's investment quality, fiscal management, and other factors. The ability of institutions to model their impact on growth is limited, often due to data uncertainty and challenges in assessing the quality of investments and their growth effects. Given the additional investment needs for climate resilience and nature protection, where IFIs are bound to play an important role, it may be reasonable to shift the focus from "missed crises" to "over-predicted crises".

By adopting this approach, the LIC-DSF could provide more accurate, stable, and useful assessments for both countries and creditors, ultimately leading to more effective debt management and sustainable economic development.

Conclusion: towards a more effective LIC-DSF

The Debt Sustainability Framework for Low-Income Countries is a critical tool in shaping economic development paths. However, its current implementation raises important questions about methodology and policy implications. As we approach the 2026 review, policymakers and civil society must engage critically with the framework's mechanics and outcomes. Key areas for improvement include:

1. Updating model selection based on recent research to enhance predictive power and reliability
2. Exploring continuous risk assessments to capture country heterogeneity more accurately and restrict opaque and potentially arbitrary manipulation of risk ratings.
3. Reassessing threshold-setting trade-offs to balance risk aversion with development needs

By addressing these issues, we can work towards a more nuanced, effective, and equitable LIC-DSF that would better serve low-income countries' development goals while maintaining prudent debt management practices. The responsibility to balance out the dual objectives of the LIC-DSF will eventually fall jointly to the IMF Executive Board and low-income borrowing countries, who must recognize that as global economic challenges evolve, so too must the tools for assessing and managing debt sustainability.

References

Barro, R.J. and Lee, J.W., 2005. IMF programs: Who is chosen and what are the effects?. Journal of monetary Economics, 52(7), pp.1245-1269.

Berg, M.A., Berkes, M.E., Pattillo, M.C.A., Presbitero, A. and Yakhshilikov, M.Y., 2014. Assessing bias and accuracy in the World Bank-IMF's Debt Sustainability Framework for Low-income Countries. International Monetary Fund.

Farah-Yacoub, J.P., Graf von Luckner, C. M. and Reinhart, C.M., 2024. The social costs of sovereign default (No. w32600). National Bureau of Economic Research.

Graf von Luckner, C., Horn, S., Kraay, A. and Ramalho, R., 2024. Predicting Debt Distress in Low-income Countries. Forthcoming.

International Monetary Fund—The World Bank, 2017, "Review on the Debt Sustainability Framework in Low-income Countries: Proposed Reforms" (September)(Washington).

Lang, V.F. and Presbitero, A.F., 2018. Room for discretion? Biased decision-making in international financial institutions. Journal of Development Economics, 130, pp.1-16.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics, 142(2), pp.698-714.

Appendix

From a simplified illustration to the 2017 WB IMF Low-income-country Debt Sustainability Frameworks' short term risk signal

In practice, the LICDSF is more complex: it adds several aspects not covered so far: in its application, the LIC DSF employs four indicators of debt; and it uses alternative scenarios ("shocks") to stress test the assumptions/projections in cases where risk is found to be low.

1. **Four Debt Indicators**
   Rather than one debt indicator, the framework applies four:
   a) Private External Debt to GDP
   b) Private External Debt to Revenues
   c) Debt Service on Debt to private external creditors to Exports
   d) Debt Service on Debt to private external creditors to Revenues

   The additional indicators allow to account for the fact that especially in the case of low income countries, with large stocks of concessional debt with long maturities, the debt stock might not be as useful as a signal for distress. In such cases, naturally, the debt service becomes a much more telling indicator about the financial pressures at play.
   In terms of the machinery little changes. The thresholds are set and optimized by indicator. If at least one of the four thresholds that result is breached, the country is considered in high risk of distress.

2. **Low and Moderate Risk:**
   The LICDSF, in addition, classifies countries into three categories (low, moderate and high risk), rather than two. Whereas high risk is assigned, when at least one threshold is breached by one of the four debt variables, as is described above, the moderate risk signal is assigned when there are not threshold breaches in the baseline, but thresholds are breached under additional stress tests. Such stress tests include, for example, the simulation of a bankruptcy in a state-owned entity that requires the government to assume a before contingent liability as its own debt.

To summarize, the LICDSF hence operates applying the following steps:

1. Estimation of four probit regressions. All four apply the same to set the debt carrying capacity, but different debt indicators.
2. Division of countries into three groups, based on their debt carrying capacity, estimated using the Country Composite Index (CCI), essentially the predicted probability net of the debt stock. The countries are grouped along the 25th and 75th percentile of the resulting distribution.
3. Definition of thresholds for each of the four debt indicator and each debt carrying capacity. Resulting in 12 different thresholds, each set such that misclassification is minimized.
4. Classification of countries as low, medium, or high risk of debt distress:
   i. Low: No threshold breaches
   ii. Medium: Threshold breach only under stress test
   iii. High: Any threshold breach

Annex Figures

Figure A1 – McCrary Discontinuity Test around the pseudo DCC thresholds of 2.8 and 3.2

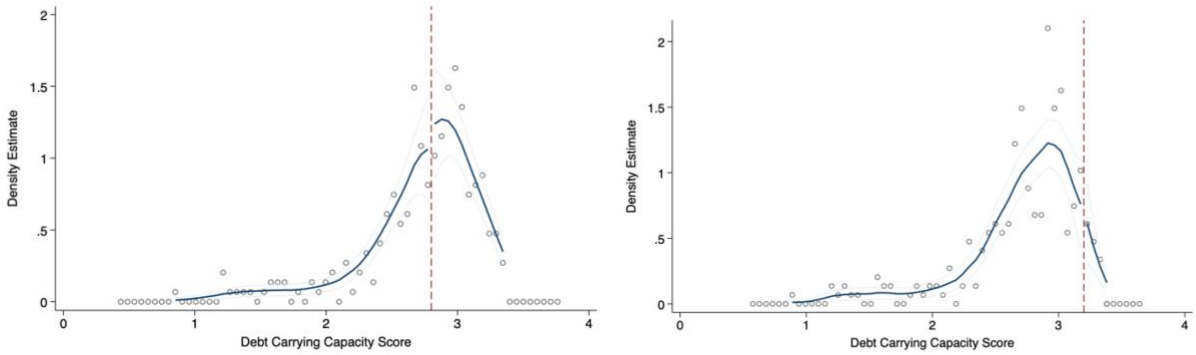


Figure A2 - NPV PPG Debt/Exports, by Debt Carrying Capacity around DCC thresholds (vertical) and risk classification cut-offs (horizontal)

Contact

Finance for Development Lab

Communications@findevlab.org
www.findevlab.org