INTERNATIONAL MONETARY FUND

SEPTEMBER 2026

REVIEW OF THE BANK-FUND DEBT SUSTAINABILITY FRAMEWORK FOR LOW-INCOME COUNTRIES—EXPLORING LIC-DSF PERFORMANCE AND SCOPE FOR IMPROVEMENTS

IMF staff regularly produces papers proposing new IMF policies, exploring options for reform, or reviewing existing IMF policies and operations. The following document(s) have been released and are included in this package:

- The **Staff Report**, prepared by IMF staff and completed on November 20, 2024.
- **A Staff Supplement: Background Paper** on the Review of the Bank-Fund Debt Sustainability Framework for Low-Income Countries—Exploring LIC-DSF Performance and Scope for Improvements, prepared by IMF staff and completed on November 20, 2024.

The report prepared by IMF staff has benefited from comments and suggestions by Executive Directors following the informal session on December 4, 2024. Such informal sessions are used to brief Executive Directors on policy issues and to receive feedback from them in preparation for a formal consideration at a future date. No decisions are taken at these informal sessions. The views expressed in this paper are those of the IMF staff and do not necessarily represent the views of the IMF's Executive Board.

The IMF's transparency policy allows for the deletion of market-sensitive information and premature disclosure of the authorities' policy intentions in published staff reports and other documents.

Electronic copies of IMF Policy Papers
are available to the public from
http://www.imf.org/external/pp/ppindex.aspx

International Monetary Fund
Washington, D.C.

November 20, 2024

REVIEW OF THE BANK-FUND DEBT SUSTAINABILITY FRAMEWORK FOR LOW INCOME COUNTRIES—EXPLORING LIC-DSF PERFORMANCE AND SCOPE FOR IMPROVEMENTS

EXECUTIVE SUMMARY

The LIC-DSF helps guide sustainable borrowing decisions of countries eligible for IMF and/or WB concessional financing and the lending decisions of the IMF and WB and other creditors. Since its launch in 2005, the framework has been periodically updated to reflect evolving debt landscape and analytical advancements. The current review is taking place amidst heightened debt vulnerabilities and increased heterogeneity in LICs borrowing behavior.

The first step in this review is to evaluate whether the current LIC-DSF remains "fit-for-purpose" and "future-proof." Following up on the April 2024 informal IMF and WB Board meetings on the launch of the current LIC-DSF review, this paper and accompanying background studies provide an analysis of the LIC-DSF predictive power and robustness to changes in the debt landscape. The predictive power of the framework is ultimately determined by the ability of the risk ratings and sustainability assessments to flag in advance the risks of debt distress and debt restructurings. A key question is whether this predictive power remains in line with what was expected at the time of the last revision of the framework in 2017.

The key findings from evaluating this question include:

- **Performance and robustness of the LIC-DSF.** Overall, the framework has proved successful in accurately flagging debt distress ahead of time, with a rate of false alarms in line with the target set in the last review. This suggests that the key design features of the framework, taken as a whole, continue to perform broadly as intended. However, re-estimating the model with revised and extended data suggests that the model's resilience to changes in the underlying debt landscape needs to be strengthened.
- **Domestic debt vulnerabilities.** The analysis also points to the need to strengthen the assessment of overall public debt distress, especially in light of the increasing importance of domestic debt in many LIC-DSF countries over the past decade.
- **Incorporation of climate-change risks and policies.** The DSAs coverage of climate change risks and policies has been evolving with the increasing recognition of their macro criticality, but the scope and depth of this coverage remains uneven, pointing to the need for developing additional tools to support the assessment.

- **Reflecting country heterogeneity.** The grouping of countries by debt-carrying capacity (DCC) continues to play a useful role in keeping the framework simple and easy to use. "Cliff effects" from changes in countries DCC have not had a significant impact on overall risk ratings and there is little evidence of instability in a country's DCC. Cliff-effects may nevertheless create some undesirable discontinuities in fiscal policy advice if not internalized in advance. The increased heterogeneity among countries using the LIC-DSF points to the need for further analysis of existing and potentially missing variables from the DCC classification and other DCC modalities.

- **Judgment-based assessments.** Whereas the use of judgment has generally improved the predictive power of the final risk ratings for countries that differ from the representative LIC, given the lack of a time differentiation of risk ratings, its use to reflect long-term considerations in the risk assessment has contributed to a higher rate of false alarms. This suggests a need for better modelling of long-term risks (e.g., from climate change) to help inform judgement. In addition, a consideration of recent debt restructuring cases suggests that the judgment-based sustainability assessments would also benefit from being informed by model-based signals.

- **Forecasting challenges.** Macroeconomic projections that underpin the debt sustainability assessment appear fairly accurate over the short- and medium-term horizons, but the drawing of definitive conclusions is complicated by the COVID-19 shock. Forecasts at longer horizons show some optimism bias in exports and revenue forecasts.

- **Debt coverage.** Comprehensive debt coverage remains key for properly capturing debt risks. State-owned enterprises (SOEs) remain a key source of fiscal risks in LICs and should ideally be fully captured in the debt sustainability assessment. However, cross-country differences in the completeness of debt coverage, including debt of SOEs, remain a challenge, potentially affecting the quality of some risk signals.

- **Continuity in assessment when countries graduate from LIC-DSF.** As LICs develop and increasingly rely on financing from international capital markets, the prospect of some "frontier markets" graduating from the LIC-DSF to the Sovereign Risk and Debt Sustainability Framework for market-access countries (MAC SRDSF) becomes more tangible. Preliminary analysis of a non-representative sample of Frontier LICs points to the possibility that some countries may experience a change in the medium-term assessments when graduating, despite no fundamental change in their underlying risk, calling for further analysis of the drivers.

These key findings suggest areas where the LIC-DSF could be further improved:

- **More resilient predictive accuracy.** A review of the underlying model will benefit from ongoing advancements in econometric modeling that can better capture the impact of macroeconomic fundamentals on the probability that debt distress materializes.

- **Better reflecting domestic debt vulnerabilities.** Increased relevance of domestic debt points to the need to further develop the model for risks to overall debt sustainability.

- **Better and more transparent treatment of long-term considerations, including related to climate change, in the risk assessment.** Introducing time differentiation in risk assessments, particularly for long-term risks, would help improve the interpretation of risk ratings. Development of a long-term climate change module would also help improve the consistency of the application of judgment to bring in these considerations in the risk assessment.

- **Better grounding of the debt sustainability assessment.** Developing a mechanical signal to better inform the application of judgment towards a final assessment of debt sustainability.

- **Better reflecting country heterogeneity.** There is scope to review and refine the classification of LICs by debt-carrying capacity to ensure it adequately captures cross-country differences across a more heterogenous group of Frontier and other LICs.

- **Strengthening forecast accuracy.** Further enhancing realism tools to help reduce optimism bias in long-term projections and incentivizing efforts to improve projections would enhance both the long-term risk assessment and the assessment of debt sustainability.

- **Addressing the challenges of incomplete debt coverage.** Exploring ways to bring all DSAs closer to the targeted debt perimeter coverage, including concerning debt of SOEs and other elements of the general government unless they pose limited fiscal risks taking due account of the potential operational burden on country teams and authorities capacity constraints. The review will also explore the feasibility of accounting for improvements in debt transparency in debt-carrying capacity classification.

- **Mitigating the risk of any discontinuity in the risk assessment when countries graduate from the LIC-DSF.** This includes the ongoing work to analyze potentially missing country characteristics from the DCC mapping with a view of minimizing any such discontinuities. As necessary, other solutions, such as allowing for a more continuous risk assessment will also be explored.

A key principle to guide the work on future-proofing the LIC-DSF will be to keep the framework transparent and easy to use. It will also aim to preserve backward-looking continuity by ensuring that changes in risk ratings and sustainability assessments relative to the existing LIC-DSF are well understood and economically sound.

Staff propose to provide a further update to the IMF and WB Boards on the LIC-DSF review, planned for Summer 2025. This work will be complemented by a robust program of engagement with key stakeholders.

Approved By
Guillaume Chabert, (IMF) and Manuela Francisco (WB)

Prepared by a team from the IMF's Strategy, Policy, and Review Department, led by Plamen Iossifov and coordinated by Karim Foda, consisting of Tamon Asonuma, Christopher Dielmann, Yukun Ding, João Capella-Ramos, Miguel Ricaurte, Moustapha Mbohou Mama, Jean-Claude Nachega, Sanan Mirzayev, Jasmin Sin, and Igor Zuccardi, under the overall guidance of Allison Holland (all SPR); and a team from the World Bank's Macroeconomics, Trade and Investment Global Practice, led by Luca Bandiera, consisting of Anna Corcuera, Mariela Caycho Arce, Wissam Harake, Juan Hernandez, Sebastian Horn, Muazu Ibrahim, David Mihalyi, Mellany Pintado Vasquez, Madi Sarsenbayev, Tessy Vasquez Baos, and Clemens Graf Von Luckner under the overall guidance of Frederico Gil Sander. Data analytics and visualization support was provided by Chen Chen, Zhuo Chen, Zhengyuan Xia, and Arthur Xie. Administrative assistance was provided by Eiman Afshar and Claudia Isern.

CONTENTS

Glossary

CI	Composite indicator
CPIA	World Bank's Country Policy and Institutional Assessment (CPIA)
DCC	Debt carrying capacity
DSA	Debt sustainability analysis
EM	Emerging market
IDA	International Development Association
LIC	Low-income country
LIC-DSF	Debt Sustainability Framework for Low-Income Countries
MAC	Market-access country
PPG	Public and publicly guaranteed
PRGT	IMF Poverty Reduction and Growth Trust
PV	Present value
RSF	IMF Resilience and Sustainability Facility
SCDI	State-contingent debt instruments
SOE	State-owned enterprise
SRDSF	Sovereign Risk and Debt Sustainability Framework
WB	World Bank
WEO	World Economic Outlook

INTRODUCTION

1. The joint IMF-World Bank Debt Sustainability Framework for Low Income Countries (LIC-DSF) helps guide sustainable borrowing decisions of countries eligible for IMF and/or WB concessional financing and the lending decisions of the IMF and WB and other creditors. It plays a central role in IMF and World Bank operations, providing key inputs for macroeconomic analyses and policy advice, as well as in application of internal policies. The framework is used to provide an early signal of the build-up of risks of public debt distress and to assess debt sustainability. This gives borrowers and creditors time to take pre-emptive actions to avoid debt distress and restructuring events. Box 1 provides an overview of the LIC-DSF and its analytical underpinnings.

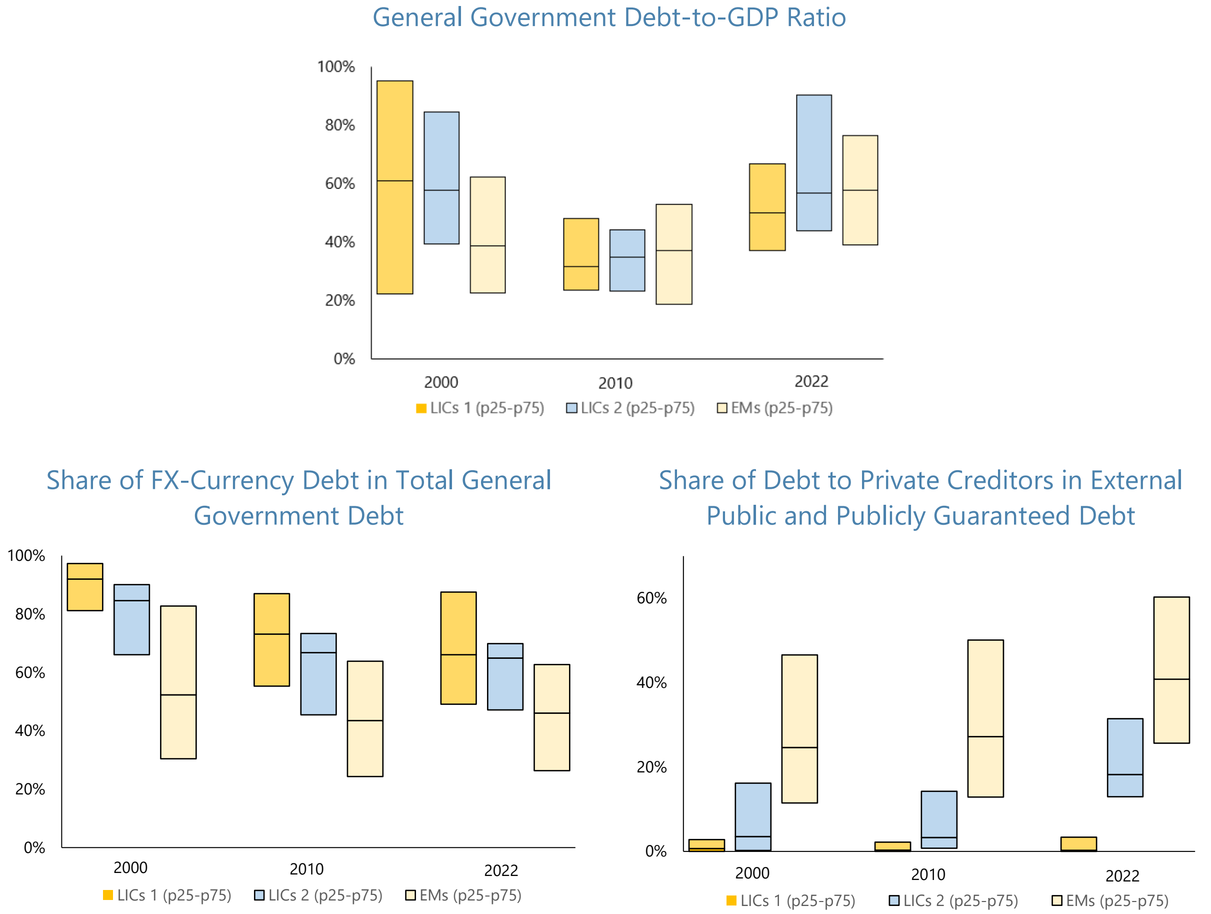
2. Since its launch in 2005, the LIC-DSF has been subject to periodic reviews to adapt it to the evolving debt landscape and keep it up-to-date with analytical advances in the field. The first reviews in 2006 and 2009 introduced technical and operational refinements, including new criteria to limit the frequency of changes of countries' CPIA score, which was then the sole measure for determining debt carrying capacity. The 2012 review introduced a new benchmark and risk rating for overall public debt and re-estimated the external debt indicator thresholds. The last review in 2017: 1) simplified the framework by reducing the number of thresholds and standardized stress tests; 2) expanded the measure of debt carrying capacity to include macroeconomic fundamentals; 3) sharpened the analysis of domestic debt vulnerabilities; 4) assigned transparent weights in the loss function used to derive the applicable thresholds – 2/3 for missed distress episodes and 1/3 for "false alarms"; 5) introduced new realism tools, tailored stress tests to capture specific risks, and a market financing module; and 6) enhanced the guidance on the application of judgment to deem away marginal/transitory breaches, and to justify an adjustment in the external risk rating to account for the severity of domestic-debt vulnerabilities, degree of exposure to external market-financing pressures, and other country-specific risks and mitigating factors.

3. The current review is taking place amidst heightened debt vulnerabilities and increased heterogeneity in LICs borrowing behavior:

- **Heightened debt vulnerabilities**
  - Over the last 10 to 15 years, overall public debt has increased markedly as a share of GDP across all LICs, mirroring developments in emerging markets (Figure 1). Across all LICs, the share of domestic in total public debt has also generally increased, with added impetus in the wake of the COVID-19 pandemic (Figure 1);
  - Since the COVID-19 pandemic, low-income countries have faced a liquidity squeeze, with diminished access to external financing and rising debt service costs, at a time of elevated development and climate-related needs; and

- **Increased heterogeneity in LICs borrowing behavior.** Since 2010, Frontier and other relatively wealthier LICs have decoupled from their peers to more closely resemble emerging markets (EMs), with an increased share of non-concessional debt (both external public debt owed to private creditors and domestic debt issued at market terms (Figure 1).

Figure 1. LICs and EMs: Changes in Debt Landscape, 2000-22
(Percent)



Source: IMF WEO, WB IDS and staff estimates.

Notes: The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at a given point of time, while the inside line indicates the median. EMs – emerging markets; Low-income countries are separated in two groups: LICs 2 countries include the Frontier LICs and other LICs with income > 300 percent of the GNI per capita cut-off for IDA. LICs 1 countries include the remaining LICs.

4. **Since the last LIC-DSF review in 2017, there have been some important analytical advances in the IMF and WB approaches to debt sustainability assessment.** The 2021 review of the DSF for market access countries (MAC SRDSF) has pushed the analytical frontier by more clearly delineating the risk of sovereign stress and the debt sustainability assessments, deriving mechanical signals for both, and introducing separate ratings for the risk of sovereign stress over the near-,

medium- and long-term tailored to countries' individual structural characteristics.[1] The 2023 evaluation of the LIC-DSF carried out by the World Bank Independent Evaluation Group (WB, 2023) highlighted the need for a more systematic treatment of the long-term impact of climate change in terms of both growth implications and the fiscal requirements of adaptation and mitigation; further strengthening of the realism tools guarding against optimism bias in projections; and closer scrutiny of differences in the coverage of SOE debt and related contingent liabilities.

5. Against this backdrop, an important aspect of the current review is to evaluate whether the LIC-DSF remains "fit-for-purpose" and "future-proof." The April 2024 informal IMF and WB Board meetings on the launch of the review highlighted some emerging challenges in the LIC-DSF implementation, flagging some potential focus areas for this review.[2] This paper and accompanying background studies provide a formal analysis of the degree to which the potential pressure points in the design of the framework have affected the predictive power of the current LIC-DSF and robustness to changes in the debt landscape. The predictive power of the framework is ultimately determined by the ability of the final risk and sustainability ratings to flag in advance the risks of debt distress and debt restructuring. This depends both on the out-of-sample results from the underlying econometric model and whether the key design features of the framework, taken as a whole, continue to perform broadly as intended.[3] The results will depend on the impact of changes in the debt landscape on the calibration and continued validity of each of the key design features of the LIC-DSF (see Box 1).

6. Key questions that will require consideration in this evaluation are outlined below. Considering these questions will help establish the relevance of the potential focus areas of the review anticipated in the April 2024 informal IMF and WB Board meetings:

[1] Sovereign stress refers to an event where market and/or fiscal pressures related to public debt become acute. It spans a broad spectrum, which in the extreme can manifest in debt distress (i.e., inability of the sovereign to meet its current financial obligations at the original contractual terms). Debt sustainability is a broader concept encompassing the ability of the sovereign to meet its current and future financial obligations at the original contractual terms, while also making adequate progress to meet development goals. Debt is considered unsustainable when there are no politically and economically feasible policies to bring about a reasonable prospect to resolve sovereign stress over the medium-term.

[2] These included: 1) Exploring ways to better differentiate between LICs; 2) Reviewing the adequacy of realism tools and stress tests; 3) Exploring how to strengthen the overall risk assessment; 4) Developing risk models by time-horizon to inform a proper time-differentiation of the risk signal, including a long-term, climate change module; 5) Developing mechanical signal for overall debt sustainability; and 6) Revisiting guidelines on debt coverage.

[3] Testing the out-of-sample model performance involves tallying its ability to correctly identify the onsets of external and overall debt distress episodes and classify country-year non-distress observations over one-to-two-year-ahead forecast horizon. Testing the ratings-based performance of the LIC-DSF entails determining in how many instances "In Distress" ratings have been preceded by a "High Risk" rating in any of the preceding two years. A separate exercise of quantifying the predictive power of the sustainability assessment in DSAs entails tallying how many "In Distress, Unsustainable" ratings have been flagged ahead of time by a "High Risk, Unsustainable" rating in a recently preceding DSA.

- Whether domestic debt vulnerabilities are adequately captured in the overall risk assessment, given the framework's focus on the external risk assessment, the rising share of domestic debt, and the limitations of residency-based debt data;

- Whether the reliance on the application of judgment for the debt sustainability assessment remains adequate given experience to date (including in debt restructurings), and heightened uncertainty;

- Whether the increased heterogeneity across LICs, driven by diverging trends in Frontier and other LICs, has strained the ability of the existing three debt-carrying capacity groups to adequately capture cross-country differences in macro fundamentals;

- Whether rising climate-related risks captured using judgment are clouding the near-term distress signal, given the lack of time differentiation of risk ratings;

- Whether the realism tools have proved effective in mitigating optimism bias in projections, in the face of increased frequency of shocks and associated heightened uncertainty around baseline forecasts;

- Whether the contingent liabilities stress test, which was redesigned and strengthened in the 2017 review, has sufficiently mitigated the impact of gaps in DSA debt coverage, including of SOE debt and other elements of the general government, on the quality of risk signals, against the backdrop of greater borrowing outside the central and general government.

7. **This paper is organized as follows.** Section II provides some brief background on the current LIC-DSF. Section III provides a more in-depth discussion on different aspects of the performance of the current framework. Section IV concludes by outlining some thoughts on the future direction of the work for this review.

BACKGROUND

8. **The LIC-DSF provides tools for assessing the risks of external and overall public debt distress, and debt sustainability.** External debt remains a critical concept for LICs. Most LICs exhibit twin current account and fiscal deficits,[4] underlining the importance of external public borrowing—which remains the most important source of financing across the majority of LICs (Figure 1)—in maintaining macroeconomic stability and its potential role as a source of risk. While retaining the separate consideration of external debt remains relevant, a more granular analysis of overall public debt vulnerabilities is critical as the relative importance of domestic debt increases.

[4] This reflects their generally lower stage of economic and financial sector development, characterized by persistent negative national saving-investment balance and limited capacity of the private sector to earn or attract foreign currency from abroad.

9. The design of the LIC-DSF aims to balance analytical rigor with the need for transparency and ease of use. The risk assessments draw from analyses of risks under the baseline and their distribution in stress-test scenarios, informed by mechanical signals from estimated near-term debt distress models combined with expert judgment (Figure 2 and Box 1). The analysis involves comparing a set of solvency and liquidity debt burden indicators — four for external public debt and an additional solvency one for overall public debt — to threshold values chosen for their ability to predict debt distress in the past. The solvency risk indicators compare the present value of debt to nominal GDP or exports of goods and services.[5] The liquidity risk indicators compare the external debt service to exports of goods and services or fiscal revenues net of grants.

10. The LIC-DSF risk ratings and sustainability assessments are determined by both the underlying econometric model, and the way it is operationalized in the LIC-DSF (Figure 2 and Box 1). Key features of the framework are: 1) country heterogeneity is captured by classifying countries in three groups by debt-carrying capacity: weak, medium, and strong; 2) the same models are used for near-, medium- and long-term risk signals that are then conflated into a single risk rating; 3) structured use of judgment is applied to capture country-specific considerations in the final ratings, given the model calibration to a representative LIC; and 4) the sustainability assessment is determined by judgment, based on the size and persistence of the mechanical risk signals.

11. The current framework has successfully flagged debt distress ahead-of-time. Since its rollout in 2018, all cases of new "In Debt Distress" ratings have been preceded by a "High Risk" rating. However, over a third of LIC-DSF users have been rated at "High Risk" of debt distress (Figure 3) without a materialization of the risk in many cases—the up-to-two-years ahead probability of debt distress[6] conditional on a "High Risk" rating has only been around 5 percent.[7] This is not necessarily a weakness per se, as the framework was put in place to flag risks so countries can take pre-emptive actions to avoid a crisis, such as seeking an IMF adjustment program or additional WB concessional financing.[8] However, the large, undifferentiated number of countries at "High Risk" raises questions about the information content of the risk signal and could unduly constrain access to financing for some LICs, to the extent lenders use the LIC-DSF risk ratings in their lending decisions.

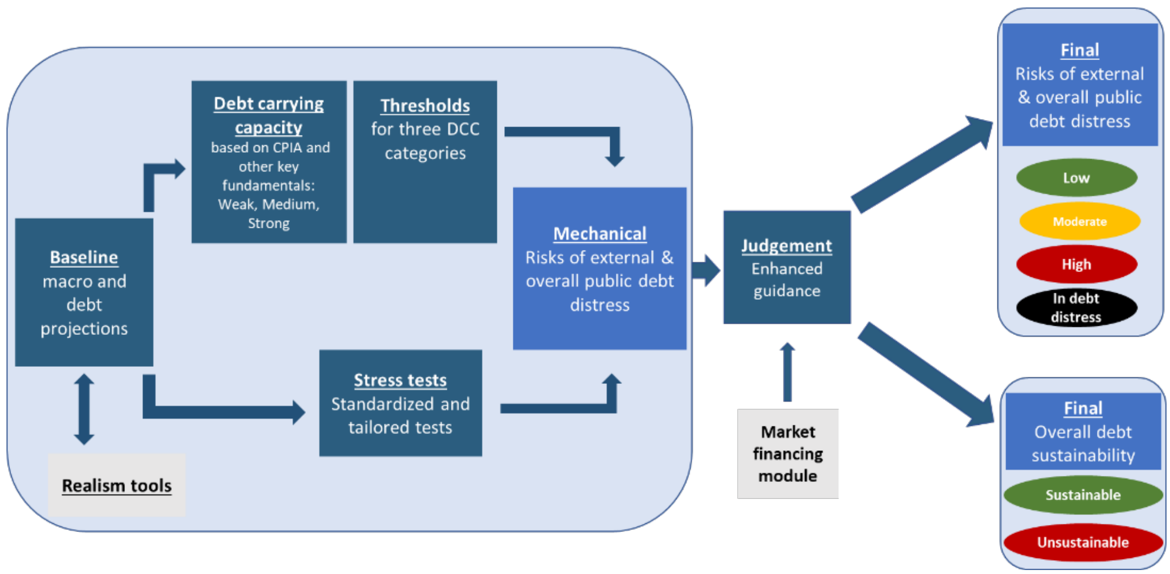
[5] When a large share of the debt is concessional, the present value of debt is a better measure of its economic burden than its face value.

[6] While the LIC-DSF model was not calibrated with respect to the up-to-two-years ahead probability, this time span is used in the analysis as an operational proxy for the short-term.

[7] Conversely, 95 percent of country-year instances of "High Risk" ratings were not followed by an "In Debt Distress, Unsustainable" rating in the following two years, a very high rate.

[8] The COVID-19 shock complicates drawing definitive conclusions about the magnitude of the endogenous policy response to a DSA ratings downgrade, as authorities' efforts to save lives and livelihoods generally resulted in higher primary deficits across LICs. Prior to 2020, only two countries (Kenya and Papua New Guinea) experienced a downgrade to High Risk of debt distress without falling in debt distress and in both cases primary deficits in the four years following the downgrade were on average lower than in the preceding eight years (by 2.6 and 0.7 percentage points of GDP in Kenya and Papua New Guinea, respectively), indicating significant fiscal adjustments to reduce debt risks.

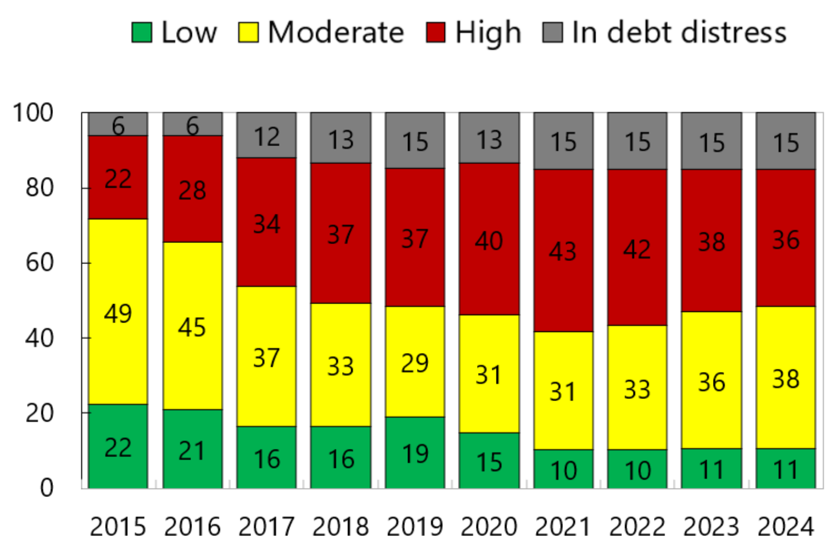
Figure 2. Main Features of LIC-DSF



Source: Staff analysis of 2017 LIC-DSF econometric model and of DSA ratings.

Figure 3. Countries Using LIC-DSF: Evolution of External Risk Ratings, 2015-24

(Share of countries)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: As of end-June 2024, based on countries with recent DSA (issued within the last three years). The 2015-2017 ratings are based on an earlier calibration of risk thresholds.

Box 1. Summary of Underlying Model and Implementation Modalities in LIC-DSF

Underlying Econometric Model

Continuous Model. The targeted debt perimeter is public and publicly guaranteed (PPG) debt. Four PROBIT regressions are estimated separately with continuous dependent variables that include one debt burden indicator per specification and a common set of explanatory macroeconomic variables, to detect risk of debt distress one-year ahead. It includes four fitted PROBITs for external public debt distress and a separately calibrated PROBIT, which is used alongside the combined risk signal for external debt distress, for overall public debt distress. Distress probability cut-offs (above which the model predicts a distress event) are then derived by minimizing a "loss function," in which the proportion of missed distress episodes and "false alarms" by the system of PROBITs enter with weights of 2/3 and 1/3, respectively.

Under the main Threshold Approach, the algorithm for deriving debt thresholds uses an aggregation rule to combine information from the four PROBIT regressions and selects the debt thresholds to minimize a prediction loss function. A composite indicator (CI) is calculated reflecting the contributions of all non-debt explanatory variables to the risk of debt distress in the four regressions. Given a cut-off probability for each debt burden indicator, the estimated probit regressions are inverted to obtain debt thresholds as a function of the CI score. Evaluating these debt thresholds at the 75th, 50th, and 25th percentiles of the distribution of the CI over the 2005–14 period results in debt thresholds for the strong/medium/weak debt carrying capacity (DCC) for each of the debt burden indicators.[1] Each country is assigned to one of the three DCC groups and a prediction rule is defined for predicting debt distress if any of the thresholds is breached, with the thresholds defined according to the DCC each country is assigned to. Finally, the cut-off probabilities for the four debt burden indicators are selected to minimize a prediction loss function that penalizes missed crisis and false alarms with a double weight (i.e., 0.67) assigned to missed crisis.[2]

Supplementary Probability Approach is included for auxiliary use when debt indicators are close to the thresholds. Under it, empirical distress probabilities are calculated using the coefficients from the Continuous Model and the values of the explanatory variables, fixed at their long-term average as of the time of each DSA, except for debt burden indicators which enter with their untransformed forecasted values. The empirical probabilities are then compared to the probability cut-offs derived from the Continuous Model.

Implementation modalities in LIC-DSF

Targeted debt coverage is the same as in the model, namely PPG debt. Public sector is defined as central, state, and local governments, social security funds and extra budgetary funds, the central bank, and all state-owned non-financial enterprises (SOEs). Central bank debt is included, if contracted on behalf of the government or taken for purposes other than to support the implementation of monetary policies or reserve management. Under clearly defined circumstances, specific SOEs may be excluded, if the enterprise poses limited fiscal risk. Publicly guaranteed private sector debt is also covered.

Mechanical risk signal is generated under the main threshold approach when a debt burden indicator breaches its threshold for <u>more than one year</u> at <u>any time in the first 10-years</u> of projections under the baseline (resulting in a High Risk signal) or in standard, tailored, or customized stress-test scenarios (Moderate Risk signal). The analysis relies on forecasts of debt burden indicators and macroeconomic fundamentals (reduced to the estimated values of their quartiles as of the last review as in the model).

The signals for external and overall public debt distress are combined with judgment to arrive at final ratings of the risk of overall and external debt distress over the entire forecast horizon, which may reach up to 20 years to take into account specific vulnerabilities, such as those related to climate change.[3] The sustainability of overall PPG debt is determined by judgment, based on the pattern of threshold breaches.

1/ The country CI-scores are estimated as a weighted average of the 10-year averages of the macroeconomic fundamentals that enter the four external PROBITs, with weights given by their average coefficients in the four PROBITs.

2/ For a full description of the process see IMF and WB (2017).

3/ Additional debt burden indicators with no set thresholds are used to inform the overall risk assessment.

Source: Staff synopsis of IMF and WB (2017 and 2018).

ANALYSIS OF LIC-DSF PERFORMANCE

A. Predictive Power and Model Stability

Out-of-Sample Model- and Ratings-Based Performance

Out-of-sample performance of both underlying model and final risk ratings for external and overall public debt are in line with the model in-sample calibration. This suggests that the key design features of the framework, taken as a whole, including the calibration of debt-burden thresholds and countries grouping by debt carrying capacity (DCC), continue to perform broadly as intended. The predictive power of the sustainability assessment is more difficult to judge. However, analysis of recent debt restructuring cases suggests that the sustainability assessments would benefit from being informed by model-based signals.

12. The predictive power of both the LIC-DSF model under the main Threshold Approach and the final debt risk ratings can be analyzed by considering the rate at which the onsets of debt distress are correctly predicted ("True Positive" rate) relative to the rate at which non-distress episodes are mis-classified ("False Positive" rate). This is summarized in the Receiver Operating Characteristics (ROC) charts (Figure 4).[9] The bigger the vertical distance between a data point summarizing a model predictive power and the diagonal line is, the better its performance for a given rate of false alarms.[10]

13. The out-of-sample performance of the LIC-DSF model under the main Threshold Approach and final risk ratings for external public debt are in line with the model in-sample calibration (Figure 4, left panel):

- **Under the Threshold Approach, the LIC-DSF model exhibits a perfect out-of-sample True Positive rate and a similar False Positive one to its in-sample calibration over 2018-22.**[11]

[9] Following the estimation of the system of PROBITs, different cut-offs for the predicted probability of a debt distress will classify differently the sample observations into distress and non-distress ones. The model can then be graphically represented as a curve on the ROC chart, in which the probability cut-offs are ordered from the highest to the lowest starting at the axes' origin. At the axes origin (corresponding to zero False Alarms and zero True Positive Rate), the cut-off for the predicted probability is 1.0, at which point the model will correctly identify all non-distress occurrences but will mis-classify all distress events as non-distress ones. As the probability cut-off is lowered, the model is able to identify progressively more distress occurrences at the expense of misclassifying an increasing share of non-distress ones.

[10] The 45-degree diagonal line corresponds to the ROC curve of an uninformative model corresponding to flipping a coin for each sample observation. The vertical distance between a data point summarizing a model predictive performance and the diagonal line is equal to the difference between the True Positive and False Positive rates.

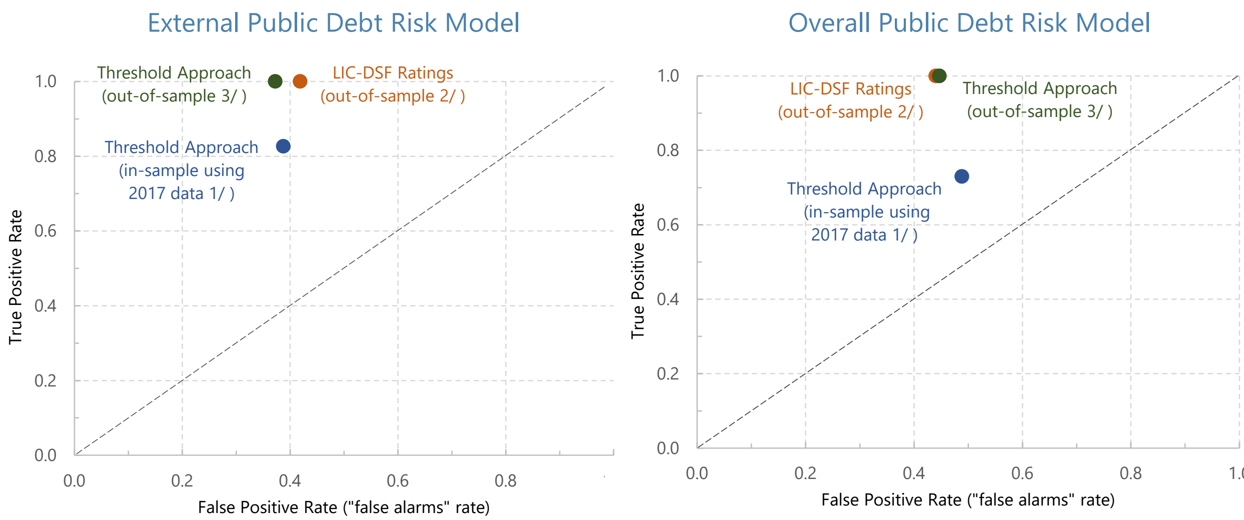
[11] The out-of-sample performance of the Thresholds Approach and LIC-DSF ratings are evaluated from 2018 onwards, as even though the LIC-DSF model was estimated with data through 2014, it effectively became operational with the launch of the new LIC-DSF starting in 2018. In the 2015-17 interim period between the end of the sample used in the model calibration in the 2017 LIC-DSF Review and the launch of the current LIC-DSF, which is not covered by the presented out-of-sample analysis, the model under the Thresholds Approach would have misclassified three distress episodes (as defined in the model database) as non-distress ones. One of these episodes is related to an instance of "hidden SOE debt" that subsequently gave rise to a misreporting case under a Fund-supported program (Box 1 in Background Paper); given the underlying data quality issues, missing this episode of debt distress is unsurprising.

- **The results based on the final risk ratings (LIC-DSF Ratings) have the same out-of-sample success rate, albeit with a higher rate of false alarms compared to both the in-sample calibration and the model-based out-of-sample one.**

14. The out-of-sample performance of the Threshold Approach and LIC-DSF Ratings in identifying overall debt distress is better than the model in-sample calibration (Figure 4, right panel).[12] They both feature perfect True Positive rates and lower rates of false alarms compared to the in-sample calibration.

Figure 4. LIC-DSF: Comparison of In- and Out-of-Sample Predictive Power of Model and Final Risk Ratings

(Index)



Source: Staff analysis of 2017 LIC-DSF econometric model and of DSA ratings.

Notes: The chart shows the in- and out-of-sample performance of DCC threshold-based approaches (see Box 1). The "in-sample" discriminatory power is based on 1-year-ahead forecast horizon and considers only onsets of distress and non-distress episodes (to match published results). The "out-of-sample" predictive performance of both the Thresholds Approach and LIC-DSF ratings is evaluated from 2018 onwards over the 1-to-2-year ahead forecast horizon. The model-based results use the same definition of observations as the reported "in sample" ones, which classify only the onsets of distress and non-distress episodes, whereas the ratings-based results consider country-year non-distress observations (to match how the DSA is used in practice) alongside the onsets of distress episodes.

1/ Over the full sample of 80 current and former LIC-DSF users in 1970-2014. Data from Tables 2 and 4 of IMF and WB (2017).

2/ Over the 69 current LIC-DSF users in 2018-23 (period of current framework implementation).

3/ Over the 69 current LIC-DSF users in 2018-22 (2022 is the year of latest available data).

15. The predictive power of the sustainability assessment is more difficult to judge, but consideration of recent comprehensive debt restructuring cases suggests that the assessment would benefit from being informed by model-based signals. The challenge in evaluating the power of the sustainability assessment is due both to the fact it is judgment-based, without an "in-sample" calibration, and that debt restructuring events often lack a clear starting point and are

[12] The in-sample predictive power of the model for the risk of overall debt distress is weaker than that for external debt distress (see the model robustness sub-section for details).

shaped by decisions taken by authorities.[13] In principle, episodes where a comprehensive debt restructuring was needed to restore debt sustainability would have been preceded by an "unsustainable" assessment. In practice, if the most recent DSA was not sufficiently timely, it may be *de facto* coincident with the restructuring process. For instance, of the six recent cases requiring a comprehensive debt restructuring,[14] Ethiopia and Somalia would be in this group. Of the remaining four cases with a DSA prepared around one year to 18 months prior to the formal request for a comprehensive debt restructuring or formal default, both Malawi (2021) and Zambia (2019) were rated as "High Risk, Unsustainable". In the case of Ghana, which formally defaulted in December 2023, the prior 2021 DSA rated debt as "High Risk, Sustainable". While debt sustainability risks were highlighted in the analysis, other considerations—including the authorities' commitment to medium-term fiscal consolidation—underpinned the risk assessment. The case of Chad was somewhat idiosyncratic as the need for a comprehensive debt restructuring (formally requested under the Common Framework in January 2021) arose in the context of a request for Fund-supported program involving PRGT exceptional access.

Contribution of Debt Burden Indicators to Aggregate Risk Signal

High-Risk ratings are assigned for a variety of reasons, informed both by mechanical risk signals and judgment. All external debt burden indicators contribute to the aggregate risk signal, though their marginal contribution differ by time horizon.

16. In the LIC-DSF implementation, High-Risk ratings are assigned for a variety of reasons, informed both by mechanical risk signals and judgment. In the period 2018-24, the share of LIC-DSF users rated at High Risk of external debt distress has been around one-third (Figure 3). Analysis of the pattern of first breaches of debt burden thresholds over time and by type of risk (liquidity, solvency, or a combination of the two) demonstrates that the mechanical signals underpinning the High Risk ratings are of a diverse nature (Figures 5 and 6).[15] Threshold breaches occur first in the near-term (within the first two forecast years) in close to two-thirds of the DSAs for countries rated at High Risk (Figure 5).[16] Most first-time risk signals involve breaches of liquidity thresholds that in a number of cases are also accompanied by solvency ones. Around a fifth of all DSAs with high external debt risk ratings do not feature any breaches of debt indicator thresholds over the 10-year period used for generating a mechanical risk signal, which is indicative of the use of judgment to override the mechanical signal in the final risk rating (see section below).

[13] A debt restructuring may be initiated even in the absence of unequivocal signs of debt distress, for example if the country seeks financing under a Fund-supported program involving PRGT exceptional access or high combined GRA-PRGT credit exposure. In such cases, the risk of debt distress would need to be brought down to "Moderate" within a specified period of time, which may necessitate a debt treatment.

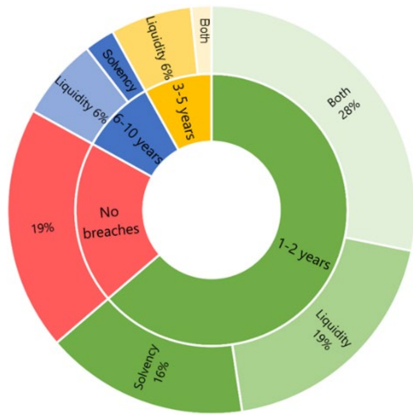
[14] All were rated as "In debt distress, Unsustainable" in the DSA that underpinned the negotiations with creditors.

[15] The mechanical risk signals are generally persistent – once triggered they tend to continue to signal distress for some time.

[16] This is in line with the framework being based on a near-term econometric model of the probability of debt distress and further reflects the fact that near-term projections are more reliable.

Figure 5. LIC-DSAs with High External Debt Risk Rating: Pattern of First-Time Breaches of Debt Indicator Thresholds over Time and by Type of Risk, 2018-23

(Share, in Percent of Total)



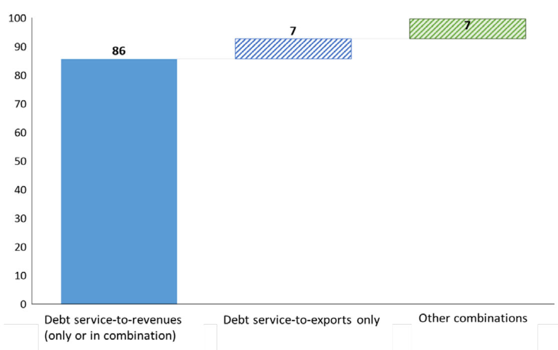
Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: The 2-level pie chart illustrates the forecast horizon over which the first breaches of debt burden thresholds occur within the 10-year forecast period of the risk assessment and within that the types of risks being flagged. The High-Risk rating corresponds to the final rating, determined after application of judgment.

17. **Focusing on the latest DSAs with High-Risk ratings, most of the first breaches of risk thresholds that occur in the near to medium term are driven by liquidity indicators.** In particular, the external public debt service-to-revenue is the indicator with the largest contribution to first-time breaches, both individually and in combination with other indicators in the near term (one-to-two years) (Figure 6). Where first-time breaches of a threshold occur over the medium-term (three-to-five years), they most frequently involve the other liquidity indicator—the debt service-to-exports ratio. Solvency indicators, on the other hand, dominate the risk signals that first occur in the long- and very long-term (beyond 5 years).

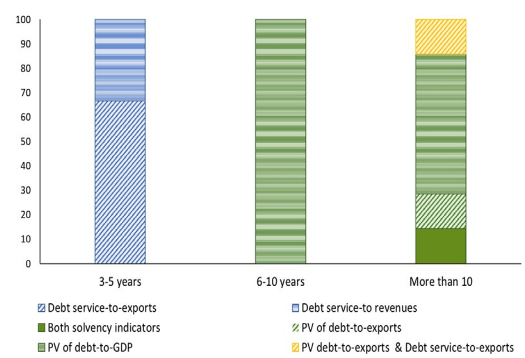
Figure 6. Latest LIC-DSAs with High External Risk Rating: Pattern of First-Time Breaches by External Debt Burden Indicator and Time Horizon

(Share in Percent of Total)

Near term (1 to 2 years)



From 3 to 20 years



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Latest DSAs refer to the last available DSA for LIC-DSF users as of December, 2023. DSAs with no breaches of debt indicator thresholds over the 10-year period used for generating a mechanical risk signal are excluded from the analysis.

Model Robustness to Data Revisions and Sample Extension

Re-estimation of the external debt risk model with updated data suggests that, while it continues signal distress with a similar performance, there are signs that the increased heterogeneity across LICs is raising some modelling challenges. Despite an improvement in the performance of the re-estimated model for the risk of overall debt distress relative to the one reported in the 2017 review, its predictive power remains weaker than the external one, indicating that there is room to improve how domestic debt vulnerabilities are captured. The main Threshold Approach performs significantly better than the supplementary Probability Approach, which has been rarely used in practice.

18. Re-estimation of the external debt risk model suggests that, while it continues to signal distress with a similar performance, the increased heterogeneity across LICs has weakened the explanatory power of macro fundamentals and their mapping into DCC groups. Results from re-estimating the Continuous Model (see Box 1) using updated and extended data for 1970-2022 show that the estimated distress probability cut-offs are lower for two of the PROBITs (Figure 7, top left panel).[17] This is partly a reflection of the reduction in the support being provided by countries DCC for their ability to carry debt burden, driven by the interplay between the generally weaker macroeconomic fundamentals and estimated coefficients in the updated and extended sample. There is also further evidence of the opening of a wedge relative to the predictive performance of the Continuous Model and under the Threshold Approach (Figure 7, top left panel).[18] Nevertheless, the main Threshold Approach continues to perform significantly better than the supplementary Probability Approach, which is rarely used in practice (Figure 7, top right panel). At the same time, the global nature of the COVID-19 shock and the coordinated policy response to it may diminish the information content of post-2019 data for the evolution of debt vulnerabilities in the future. This will need careful consideration when updating the models and drawing inferences.

19. The results from re-estimation of the overall debt risk model suggest improved performance in the extended sample (Figure 7, bottom panel). Nevertheless, the in-sample predictive power of the model for the risk of overall debt distress remains weaker than that for external debt distress. This general result is likely driven by the more limited coverage of domestic debt vulnerabilities. Whereas externally driven risk of overall debt distress is modelled using the system of four external debt risk PROBITs, domestic debt vulnerabilities are reflected only indirectly in a risk signal derived from the dynamics of the PV of overall debt-to-GDP ratio through an ad-hoc application of the “signal-to-noise” method.[19]

[17] In the PROBIT regressions, higher values and coefficients of variables that enter in the calculation of a country’s debt-carrying capacity translate into higher cut-offs for the probability of debt distress. Regression findings are also affected by data revisions over the entire sample, including changes in the timing and frequency of distress episodes due to the use of updated and newly available data sources and the impact of revisions stemming from statistical methodology improvements, such as rebasing and expansion of coverage of GDP series.

[18] As the Threshold Approach synthetizes available information in a finite number of groups (see Box 1 for details), it is expected to underperform the unconstrained Continuous Model as part of the tradeoff between precision and simplicity. The difference in performance has, however, substantially increased compared to the original 2017 Model.

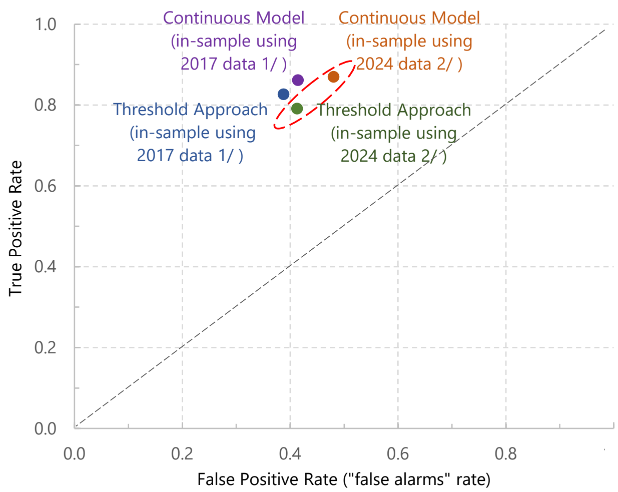
[19] In the model, overall public debt distress is defined by combining external and domestic debt distress episodes.

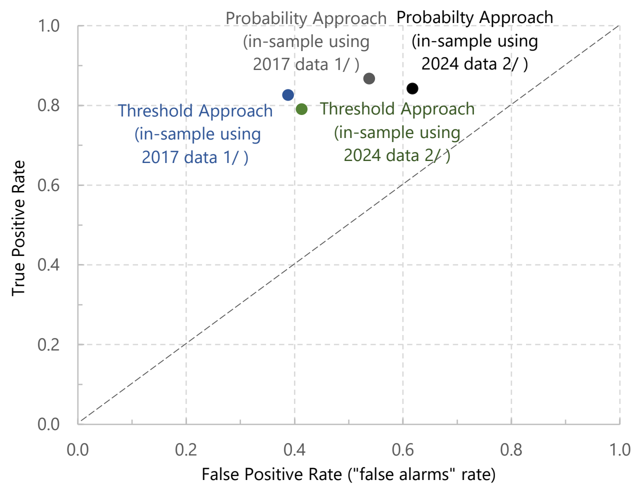
Figure 7. LIC-DSF: Comparison of In-Sample Predictive Power of Continuous Model, Probability and Threshold Approaches in 2017 Review and Updated Samples

(Index)

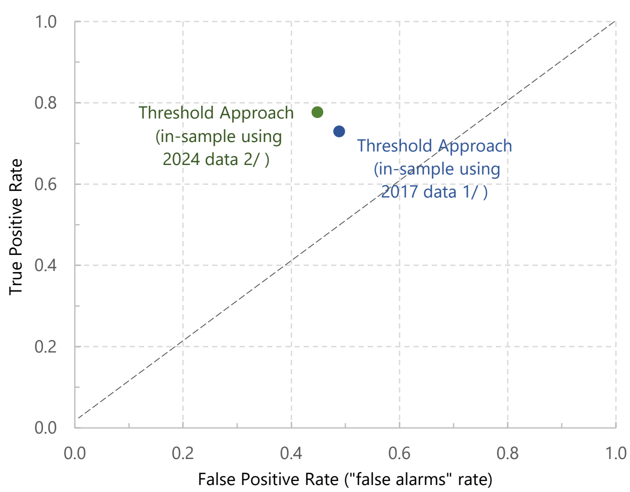
External Public Debt Risk Model

Sample / Prob. Cut-offs	PV of debt-to-GDP	PV of debt-to-exports	Debt service-to-revenues	Debt service-to-exports
2017 data	0.16	0.16	0.15	0.14
2024 data	0.12	0.16	0.13	0.14





Overall Public Debt Risk Model



Source: Staff analysis of 2017 LIC-DSF econometric model.
Notes: The Probability Approach is implemented using 10-year moving historical averages of explanatory variables, except for debt burden indicators.

1/ Over the full sample of 80 current and former LIC-DSF users in 1970-2014.

2/ Over the full sample of 80 current and former LIC-DSF users in 1970-2022.

B. Analysis of Key Design Features of LIC-DSF

Use of Debt-Carrying Capacity to Capture Heterogeneity

The grouping of countries by debt-carrying capacity (DCC) continues to play a useful role in keeping the framework simple and easy to use. "Cliff effects" from changes in countries DCC have not had a significant impact on overall risk ratings and there is little evidence of instability in countries' DCC. Cliff-effects may, nevertheless, create some undesirable discontinuities in fiscal policy advice if not internalized in advance. Frontier LICs do not appear to be disproportionately affected by potential gaps in the DCC definition, as their DCC generally aligns well with market signals of credit risk, and they are subject to a similar degree of "false alarms" generated by the model as other LICs. However, the increased heterogeneity across LICs suggests there may be merit in further analysis of existing and potentially missing variables from the DCC classification.

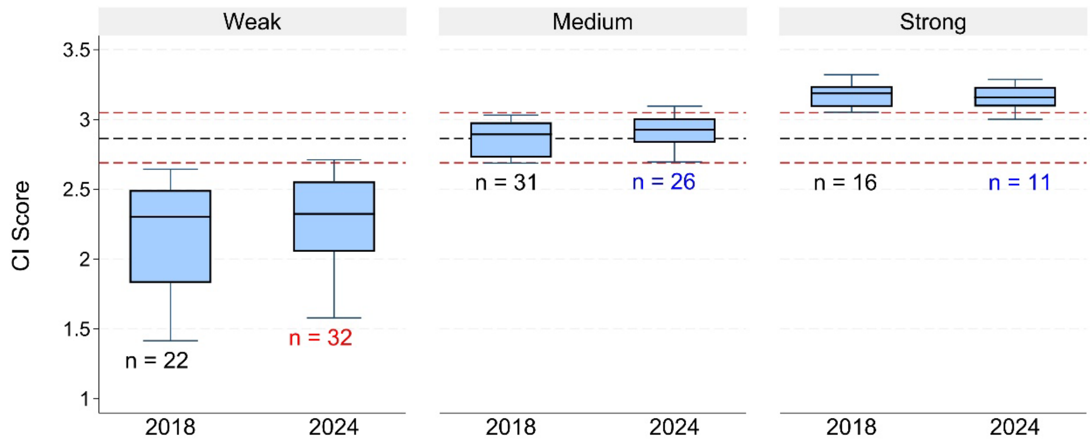
20. While an inherent feature of the framework, "cliff effects" from changes in countries DCC have not had a significant impact on overall risk ratings in LIC DSAs.[20] Since 2018, there has been a general weakening of countries' CI-scores. As a result, about a quarter of LICs have experienced a change in their DCC, nearly all of which were downgrades, in the majority of cases to Weak DCC (Figure 8). This downgrade in the DCC triggered a downgrade in the final risk ratings in only three instances, all of which occurred around the pandemic and were corroborated by a deterioration in debt vulnerabilities.

21. Cliff-effects may nevertheless create some undesirable discontinuities in fiscal policy advice if not internalized in advance. As seen in Figure 9, for countries that were already rated at High Risk of debt distress in the final risk assessments of LIC DSAs, the downgrades have resulted in greater positive threshold breaches, while for countries that were rated as Moderate Risk, the buffers between the risk indicator and its respective threshold have diminished. Whereas the implied intensification of risk signals in these cases has been in-line with the weakening of macro fundamentals, the resulting change in the policy-relevant threshold may create some discontinuity in fiscal policy advice if not internalized in advance. This underscores the importance of tailoring fiscal policy advice to reflect the likelihood of pending or expected DCC changes ahead of time. Cliff-effects may be particularly relevant in instances where the authorities are undertaking a substantive and transformational reform program in response to a crisis. In that context, the centered average of historical institutional factors and macroeconomic outturns and forecasts may not adequately reflect the expected structural improvement in DCC and overly restrict fiscal policy in the post-crisis reform period.

[20] "Cliff effect" occurs when a country's CI score changes from being just above a given percentile cut-off to just below it, giving rise to a signal for change in the DCC grouping. If the latter occurs, the distance between debt burden indicators and their thresholds can change abruptly from one DSA to the next, due to the step difference between thresholds for countries in different DCC groupings.

Figure 8. Distribution of CI Scores by Debt-Carrying Capacity, 2018 and 2024

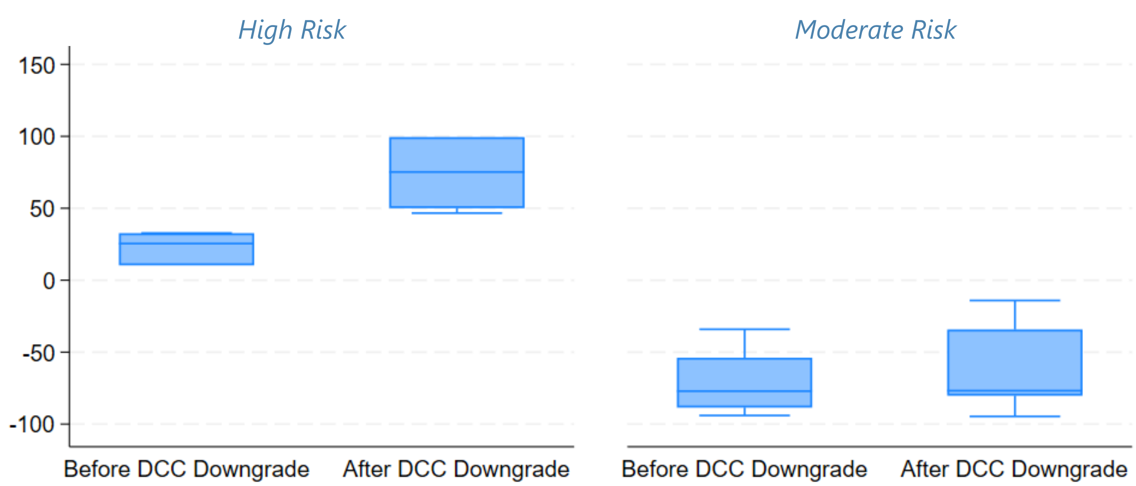
(Index number)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Boxes represent interquartile ranges and black lines denote medians. Dashed red lines indicate the 25th (2.69) and 75th (3.05) percentiles, and the dashed black line represents the median (2.86) of the distribution of average CI-scores in 2005-14, assigned to countries under existing DCC methodology. The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at the given forecast horizon, while the horizontal line indicates the median.

Figure 9. High and Moderate Risk LIC DSAs: Maximum Average Threshold Deviations of External Debt Indicators over First Five Forecast Years Before and After DCC Change

(Percentage Points)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: The deviations are calculated relative to thresholds corresponding to countries' debt carrying capacity for the four external debt burden indicators. They are averaged over the first five forecast years, using only positive deviations for countries at High Risk and negative ones for countries at Moderate Risk. The maximum of the average values for each country is then used to construct the boxplots.

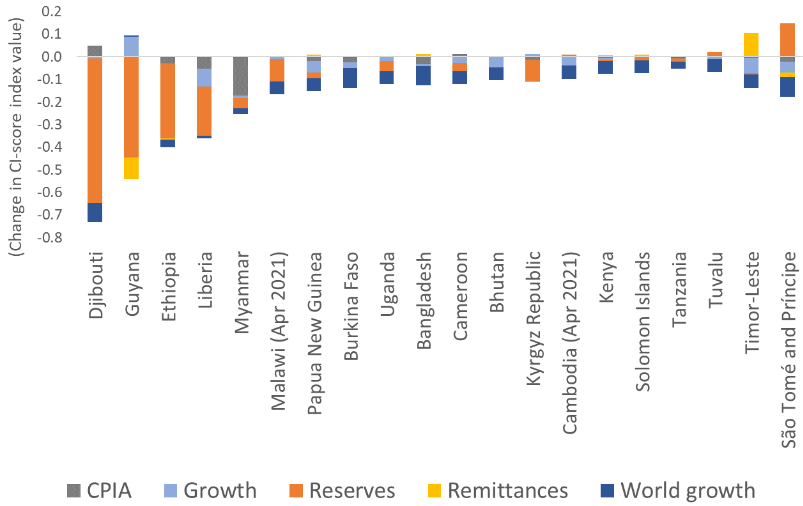
22. There is little evidence of instability of DCC. Out of 23 DCC changes since 2018, 20 were downgrades. Figure 10 shows the contributions of macro fundamentals to the change in the CI-score in the year a country was downgraded to a lower DCC group. Results show that in many of these cases, the lead cause has been the downward revision of the 10-year centered average of global growth in the wake of the COVID-19 pandemic (Figure 10). The new DCC classifications have since been largely maintained notwithstanding the post-pandemic improvement in global growth prospects, reflecting subsequent weakening of country-specific indicators[21] Beyond the immediate aftermath of the pandemic, DCC downgrades have been less frequent and mostly led by declining external buffers, as measured by the 10-year centered average of reserve coverage of imports.

23. The increased country heterogeneity within existing DCC groups points to the need for further analysis of potentially missing country characteristics that could lead to DCC misclassification and unduly limit Frontier and other LICs' access to financing and contribute to the rate of "false alarms." Potential gaps in the DCC definition are more likely to manifest in the classification of Frontier LICs, which have progressively decoupled from their peers in the share of non-concessional in total public debt (Figure 1). However, considering the relationship between the DCC of Frontier LICs and market signals of credit risk suggests that the DCC classification is generally well aligned. Issuers with stronger DCCs have generally benefitted from lower spreads (Figure 11). At the same time, some Frontier LICs with Medium DCC have issued at spreads comparable to those of Strong DCC countries, pointing to the need to explore additional country characteristics potentially missing from the current measure of debt-carrying capacity. Analysis of the common features of Frontier LICs suggests that in addition to strong policy frameworks, which are already captured in the DCC, other characteristics such as the size of the economy and level of domestic financial development may matter for their capacity to carry debt and for market access (Figure 12). The impact of any missing country characteristics from the DCC classification is also likely to be similarly relevant for other LICs, as there is no significant difference between the two groups in the rate of "false alarms" in the out-of-sample predictions of the LIC-DSF model under the Threshold Approach.

[21] The largest concentration of DCC downgrades occurred in October 2020 and April 2021, following post-pandemic updates of CI-scores. The CI-scores are updated twice-a-year reflecting forecast revisions in the April and October World Economic Outlook, in addition to the once-a-year update of countries CPIA scores. No CI-scores update took place at the time of the April 2020 WEO update, which took place at the height of the pandemic.

Figure 10. LIC-DSF Users: Contributions to CI-Score Changes Resulting in DCC Downgrades, 2018-24

(Level change in index)

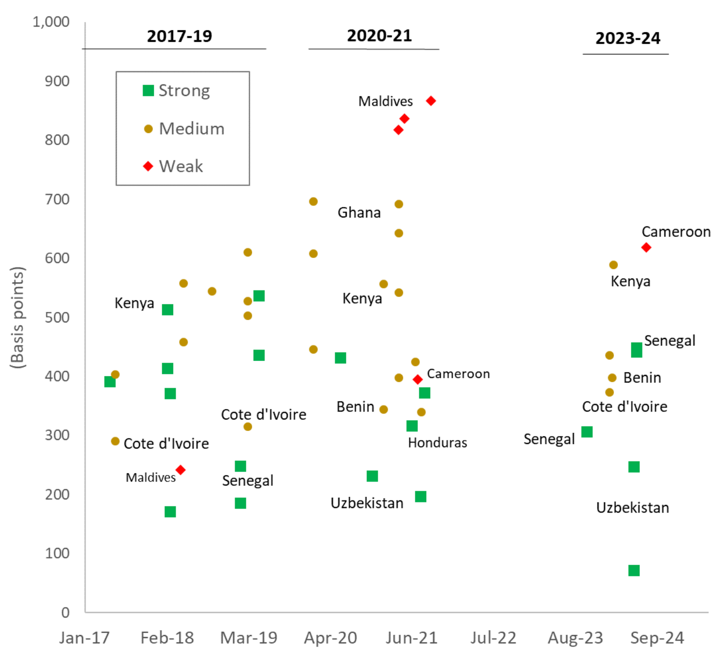


Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

Notes: Shown are level contributions to changes in CI scores that resulted in DCC downgrades. A country CI-score is estimated as a weighted average of the centered average of five macroeconomic fundamentals based on five years of historical data and five years of projections, using the fixed weights set in the 2017 LIC-DSF review.

Figure 11. Frontier LICs: Sovereign Bond Spreads by DCC at Issuance, 2017-24

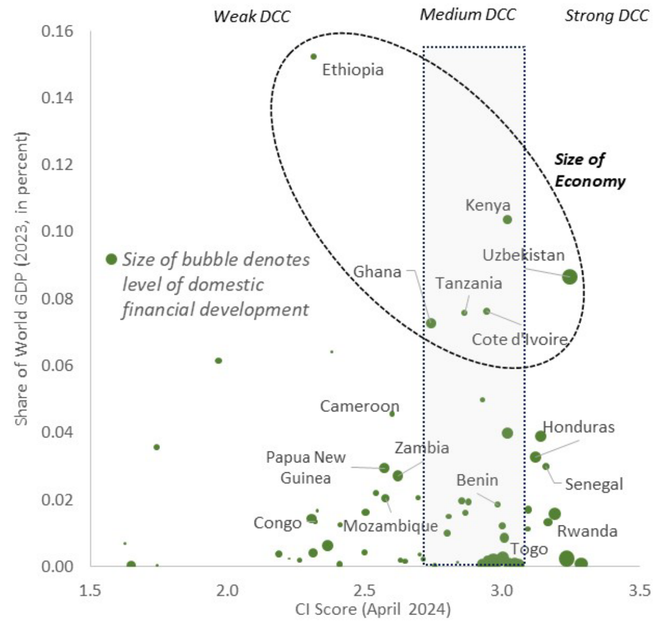
(Basis points)



Sources: Staff analysis of Bond Radar data and joint IMF-WB LIC-DSF country analyses.

Notes: As of August 2024. 2017 DCC is as of Oct. 2018. Mozambique, Congo, Rep. of and Togo are missing data.

Figure 12. LICs: Distribution of Frontier and Other LICs by Financial Development, Economic Size, and Debt-Carrying Capacity



Sources: Staff analysis of joint IMF-WB LIC-DSF country analyses; IMF WEO and Financial Development databases.
Notes: GDP shares based on 2023 nominal GDP measured in US dollars. CI scores as of April 2024. Labeled countries are Frontier LICs. Size of bubbles proportional to values of the IMF Financial Development index.

Use of Judgment and Lack of Time Differentiation of Risk Ratings

Whereas the use of judgment has generally *improved outcomes for countries that differ from the representative LIC, its use to reflect long-term considerations in the risk assessment, against the backdrop of lack of time differentiation of risk ratings, contributes to a higher rate of false alarms.*

24. The structured use of judgment to capture country-specific considerations in the final risk ratings and sustainability assessments is an integral part of the LIC-DSF. Over 2018-23, judgment was applied in around one fifth of all DSAs for 31 of the 69 LIC-DSF users. In all cases where judgment was applied, the final risk ratings differed from the ones implied by the mechanical risk signals by one notch. Judgment was most frequently used in countries with medium debt-carrying capacity (DCC), equally split between cases of an upgrade and downgrade of the risk rating and was least frequently applied in countries with strong DCC.

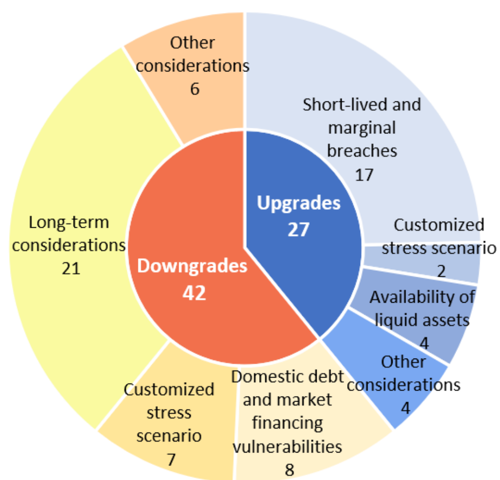
25. The use of judgment to upgrade risk ratings by itself has helped reduce the incidence of false alarms, without resulting in any missed onsets of debt distress.[22] In around 40 percent of the cases, the application of judgment improved the final risk rating (Figure 13). In half of these cases, the mechanical signal for either the overall or external public debt was "High Risk", while the

[22] The framework allows for the use of judgment to complement the model-based risk signals with country-specific considerations that might not be well captured by the model. Judgment is captured in the final risk rating and sustainability assessment.

final risk rating after the application of judgment was upgraded to "Moderate Risk". This has helped reduce the rate of false alarms, as in none of these cases, the final "Moderate Risk" rating was then followed by a change in the risk rating to "In Distress" in subsequent DSAs.[23] In most cases, the use of judgment was applied to discount short-lived and marginal breaches of the thresholds, in line with the given guidance (Figure 13). In over half of these cases, the debt indicators returned to below their respective thresholds in subsequent DSAs within two years re-affirming the adjusted final rating, with the use of judgment helping to promote continuity in policy advice. In other cases, the application of judgment has provided a temporary backstop of previously assigned risk ratings, while awaiting further evidence for a change in the assessment. Other commonly used justifications to discount mechanical risk signals include the availability of liquid assets, which is not part of the model as it plays a minor role for the representative LIC but can provide a significant buffer in some cases, and large remittances relative to the size of the economy.

Figure 13. LIC-DSAs: Frequency, Direction and Reason for Application of Judgment, 2018-23

(Number of DSAs)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

Notes: The 2-level pie chart illustrates the impact of the use of judgment on the final rating (upgrade or downgrade) and within that the main considerations for it. A downgrade occurs when the final rating indicates a higher risk than implied by the mechanical signal.

26. **Conversely, the use of judgment to incorporate long-term considerations in the risk assessment has contributed to a higher rate of false alarms.** In close to 60 percent of the cases, the use of judgment has resulted in higher risk ratings (in most cases High Risk), potentially contributing to a higher rate of false alarms. Half of these rating downgrades—spread across seven countries that are either small islands or fragile and conflict-affected (or both)—were driven by long-term considerations, mostly related to climate risks (Figure 13). In all these cases, there were no threshold breaches over the 10-year forecast period used to derive mechanical risk signals. As a

[23] Have they not been upgraded, the "High Risk" ratings implied by the mechanical risk signals would have resulted in false alarms.

result, around a fifth of all DSAs with a final high external debt risk rating (covering 10 out of the 30-35 countries rated High Risk over 2018-23) do not feature any breaches of debt indicator thresholds over the 10-year period used for generating a mechanical risk signal (Figure 5). This potentially complicates the information content of a "High Risk" rating in these instances.

Reliance on Forecast Accuracy

Projections in the short- and medium-term appear fairly accurate, but the drawing of definitive conclusions is complicated by the COVID-19 shock. Forecasts at longer horizons show some optimism bias in exports and revenue forecasts.

27. The COVID-19 shock complicates drawing definitive conclusions about the quality of the medium-term forecasts of debt indicators underpinning the risk assessment. Results from the analysis of forecasts of debt burden indicators and main macroeconomic variables that enter in their calculations do not reveal signs of persistent or widespread optimism bias in short-term projections, up to the 3-year forecast horizon, where 'forecast errors'[24] for the median LIC DSF country remain near or below zero for all debt burden indicators (Figure 14).[25] Over the medium-term, forecast errors tend to show signs of optimism for the median country and a wider range of countries around the median. However, these results are skewed by the impact of COVID-19—a "black-swan" event that occurred midway through the implementation of the current LIC-DSF—as forecast errors for horizons longer than three years entirely reflect projections made pre-COVID for outcomes that occurred post-shock.

28. The 'forecast errors' for the PV of the overall public debt-to-GDP ratio show more clearly the fiscal cost of the COVID-19 shock and subsequent loss of market access that forced LICs to ramp up domestic borrowing (Figure 14). Given the disproportionally large impact of the COVID-19 pandemic on medium-term forecasts errors, this is not an indication of optimism bias, but highlights the importance of closer scrutiny of domestic debt developments and financing terms.

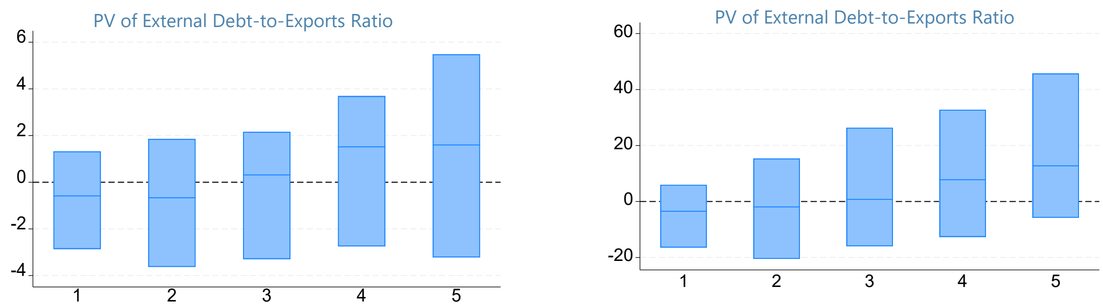
[24] Differences between data outturns and earlier projections are not necessarily an indication of issues with the quality of forecasts. Forecasts aim at capturing the central tendency of economic indicators and, as such, are not able to capture the impact of exogenous shocks beyond the mechanical extrapolation of their average impact on historical realizations. As a result, global exogenous shocks that impact all LICs in a similar manner will drive a wedge between sample measures of the difference between data outturns and their earlier forecasts irrespective of the quality of the underlying projection tools. Differences between data outturns and their earlier forecasts can also be due to revisions of historical data series made after a forecast has been made, as such revisions will have a carry-over effect on the level of future realizations of the projected variable.

[25] The findings for the debt burden ratios are representative for both their numerators and denominators, with little sign of offsetting biases in their forecasts.

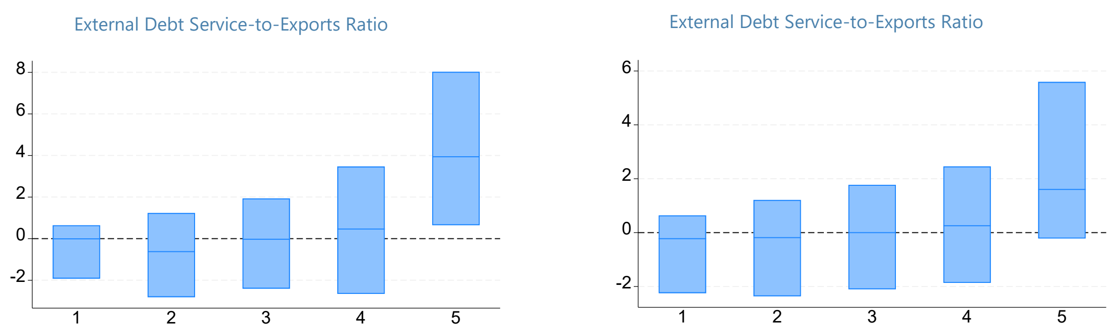
Figure 14. LIC-DSAs: Forecast Errors for External Debt Burden Indicators and PV of Overall Public Debt-to-GDP Ratio, 2018-23

(Difference between Outturns and Projections, Percentage points)

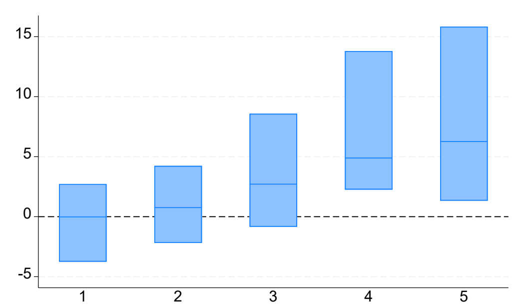
External Solvency Debt Indicators



External Liquidity Debt Indicators



Present Value of Overall Public Debt-to-GDP Ratio



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

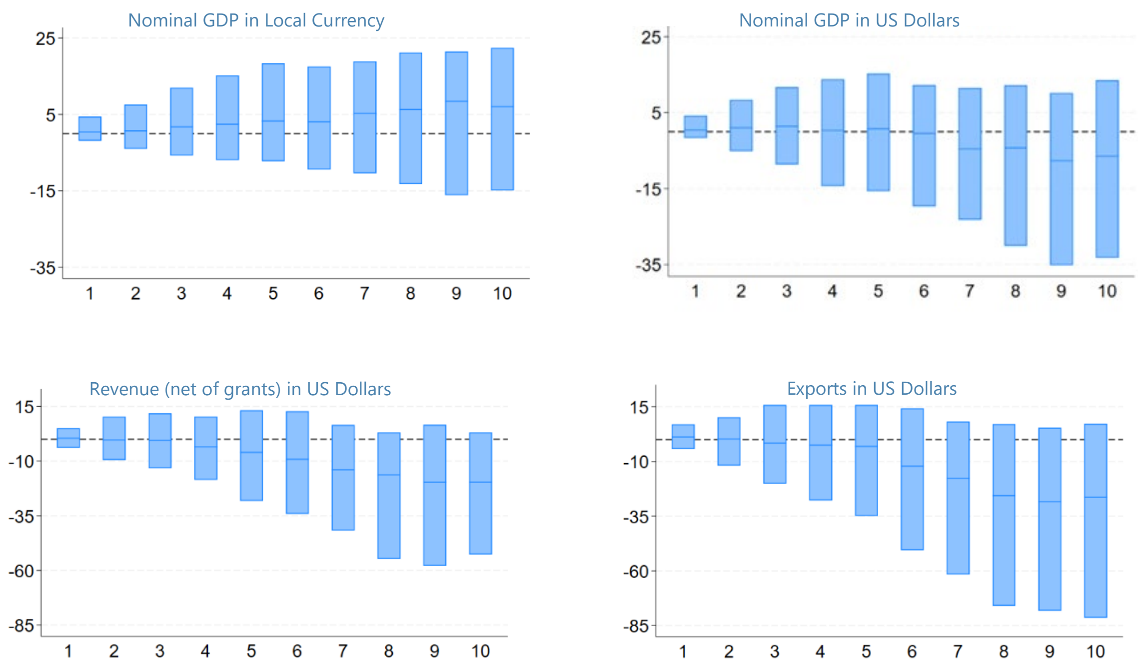
Notes: Forecasts errors in this figure measure the difference between outturns and the forecasts of their values made in earlier DSA vintages. Positive values signal the presence of Optimism Bias, as projections underestimate the debt burden outturns. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported. The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at the given forecast horizon, while the horizontal line indicates the median.

29. Analysis of forecast accuracy spanning a longer time period suggests that forecasts used in the DSA are fairly accurate over the medium term, with some optimism bias in exports

and revenue forecasts at longer horizons. Figure 15 presents analysis of optimism bias in an extended dataset over 2006-23 that spans the period of implementation of the LIC-DSF, as modified in successive reviews of the framework. While this analysis is less impacted by the COVID-19 shock, it does include forecasts made before other major external shocks, such as the 2008 Global Financial Crisis, underscoring the inherent difficulty in evaluating forecast quality in the presence of major global external shocks. Results of the analysis of forecasts of the main macroeconomic variables that enter in the calculation of the debt burden indicators do not reveal signs of persistent or widespread optimism bias over the medium term (up to 5-year forecast horizon), where 'forecast errors' for the median LIC DSF country remain near zero in all cases.[26] There are, however, signs of optimism bias for the median country in longer term forecasts of exports and fiscal revenues (excluding grants),[27] and to a lesser extent in forecasts of GDP levels in USD.

Figure 15. LIC-DSAs: Forecast Errors for Macroeconomic Series Entering Debt Burden Ratios, 2006-23

(Deviation of Outturn from Projections, Percent)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Forecasts errors in this figure measure the percentage deviation of outturns from the forecasts of their values made in earlier DSA vintages. Negative values signal the presence of Optimism Bias, as the projections overestimate the outturn of variables that enter as denominators of the debt burden indicators. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

---

[26] In the case of nominal GDP in levels, this is in part driven by statistical methodology improvements, such as GDP rebasing and expansion of coverage that result in higher nominal levels of GDP, occurring in the period between when a projection was made and when the data outturn was first reported. In fact, in more than half of the cases, the direction of forecast errors in the level of GDP in local currency is toward under-projection, reflecting in part upward revisions of historical values that affect the entire data series.

[27] While forecast errors of revenues and exports are more pronounced in commodity exporters, they are also present across other LICs.

Targeting of a Comprehensive Debt Perimeter

Comprehensive debt coverage remains key for properly capturing debt risks. SOEs remain a source of heightened fiscal risks in LICs. Differences in debt coverage across countries can, however, raise questions about the quality and consistency across countries of some risk signals, as countries with less comprehensive debt coverage tend to exhibit smaller average threshold breaches, potentially implying a tendency to under-estimate underlying risks. Such concerns are in part mitigated by the more conservative calibration of the contingent liabilities stress test for some of the affected countries.

30. Comprehensive debt coverage is key for properly capturing public debt risks. The existing framework targets a near-complete coverage of both external and domestic, public, and publicly guaranteed debt. The institutional coverage includes both the general government (or its individual components, such as central, state, and local governments, social security funds and extra budgetary funds, if consolidated data is not available), the central bank,[28] and all state-owned non-financial enterprises (or SOEs). Under clearly defined circumstances, specific SOEs may be excluded, if the enterprise poses limited fiscal risks.

31. The rationale for including in the debt perimeter all SOE debt unless the SOE poses limited fiscal risks, including debt not guaranteed by the government, is even stronger now that LIC SOEs have stepped up their borrowing to match that across EMs, even while their profitability remains weaker. Since the LIC-DSF was introduced in 2005, all SOE debt has been included by presumption in the debt perimeter.[29] This reflects the fact that LICs often lack the institutional framework and capacity to ensure the commercial viability of SOEs that often carry out quasi-fiscal activities. Over the intervening period, the SOEs' debt-to-GDP ratio in LICs has increased to the level observed in EMs, mainly driven by greater local currency borrowing. At the same time, SOEs in LICs have lower average and median net profit-to-GDP ratios and in a number of LICs the SOE sector as a whole is loss-making (Figure 16). Fiscal risks from public debt channeled through SOEs are aptly illustrated by the 'hidden SOE debt' in Mozambique which triggered an 8 percent of GDP increase in the stock of public and publicly guaranteed debt (IMF, 2024).

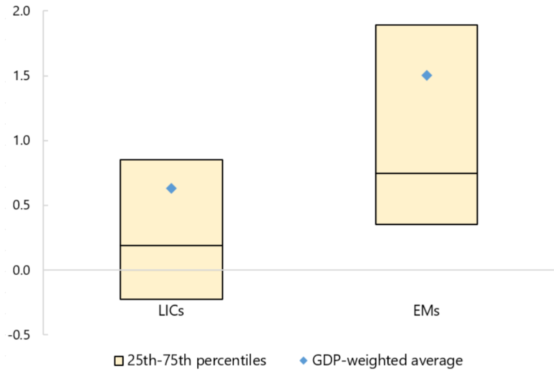
32. There has been a welcome improvement in institutional debt coverage in DSAs since the last review, though progress has proved slow and the depth of coverage remains uneven, reflecting in part capacity constraints. Since 2018, the share of DSAs that cover debt beyond the central government has increased from 45 to 52 percent of the total (Figure 17). Improvements were observed in the inclusion of central bank debt, state and local government debt, SOE debt, and social security fund liabilities. At the same time, less than a third of all DSAs include non-guaranteed SOE debt. Coverage of other elements in the general government, such as social security and extrabudgetary funds is also uneven. These data gaps can affect the quality and consistency of the risk and sustainability assessments.

[28] For the central bank, only debt contracted on behalf of the government or taken for purposes other than to support the implementation of monetary policies or reserve management is counted as public debt.

[29] Under clearly defined circumstances, specific SOEs may be excluded, if the enterprise poses limited fiscal risk.

Figure 16. LICs and EMs: Fiscal Risks from SOEs
(Percent)

SOE Net Profit/Loss as Share of GDP

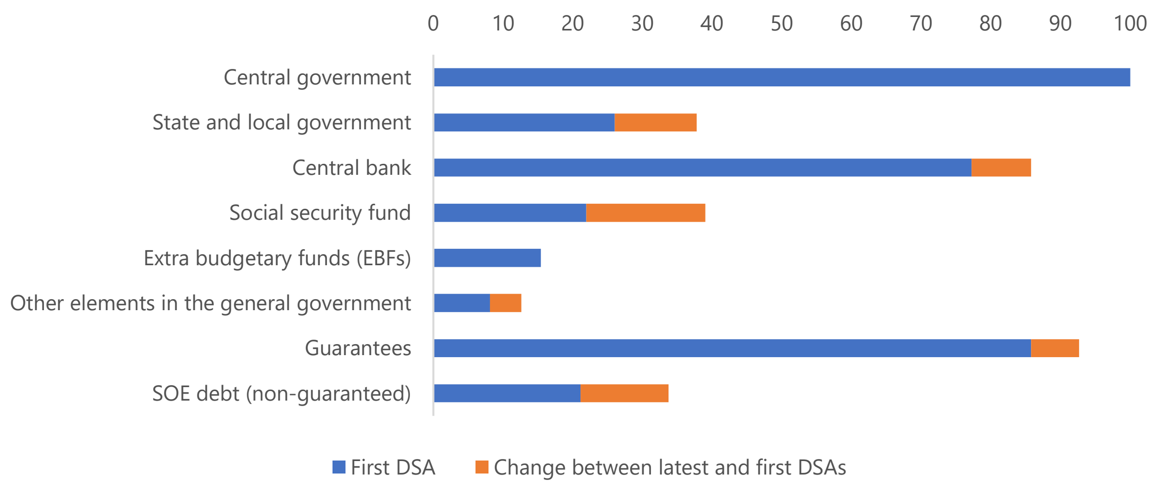


Source: World Bank Business of the State database.
Note: Excluding high income. Database tracks enterprises with any state ownership rather than majority state ownership.

Ratio of Loans to SOEs to GDP for Median LIC and EM (in Local and Foreign Currency)

Sources: IMF's MFS and WEO databases.
Note: Outlier observations for Venezuela are excluded. Data coverage varies by year and is available on average for 71 percent of LICs and EMs. LICs are LIC DSF user countries and EMs are SRDSF user emerging markets.

Figure 17. LIC DSF Users: Institutional Coverage of Debt in First and Latest DSAs Using Current LIC-DSF
(Percent)



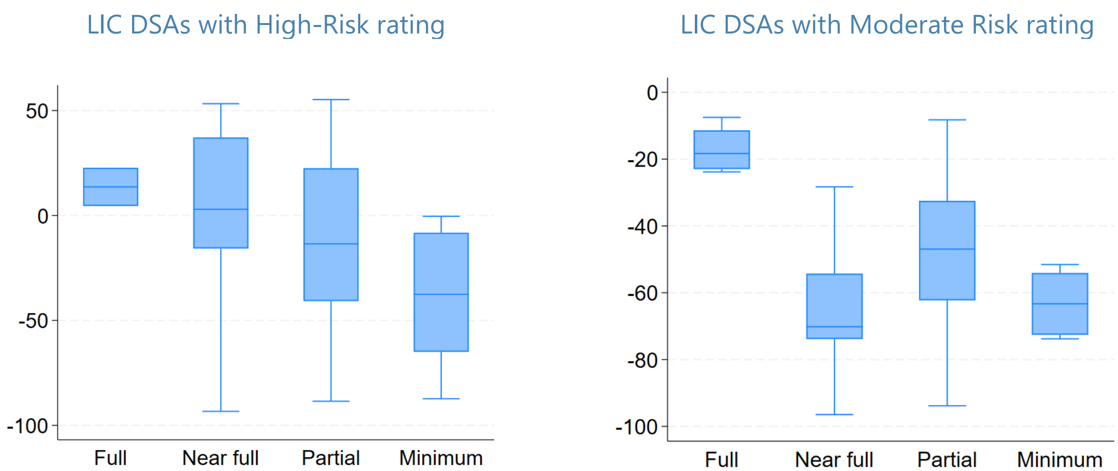
Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Shown is the institutional coverage in the first and latest DSAs using the current framework. Lack of coverage of elements of general government can reflect different institutional arrangements rather than data gaps.

33. **Differences in data availability, however, raise concerns about the quality and consistency of some risk signals across countries, notwithstanding the more conservative calibration of the contingent liabilities stress test in some of the affected countries.** This issue

arises from the fact that countries are judged against uniform thresholds, irrespective of the breadth of actual debt coverage. Analysis of the magnitude of the average deviations of debt indicators from their thresholds over the medium-term suggests that countries with less comprehensive debt coverage tend to exhibit smaller threshold breaches in High Risk cases and are farther from the thresholds from below in Moderate Risk ones, with corresponding effect on fiscal space assessments, potentially affecting the quality of some risk signals (Figure 18).[30] This implies a relatively weaker risk signal for countries with less comprehensive coverage relative to a similar country with better coverage, and may miss an important element of fiscal risk. As a way of illustrating the magnitude of the problem, including the estimated contingent liabilities in the debt perimeter could impact both the overall and external risk ratings of countries with incomplete data coverage.[31] At the same time, the contingent liabilities stress tests in DSAs for some of the countries with partial SOE coverage have generally incorporated bigger shocks than the median for other LICs, helping to bring into the analysis the associated risks (Figure 19).[32]

Figure 18. High and Moderate Risk LIC DSAs: Average Threshold Deviations of External Debt Service-to-Exports Ratio Over First Five Forecast Years by Debt Coverage

(Percent)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

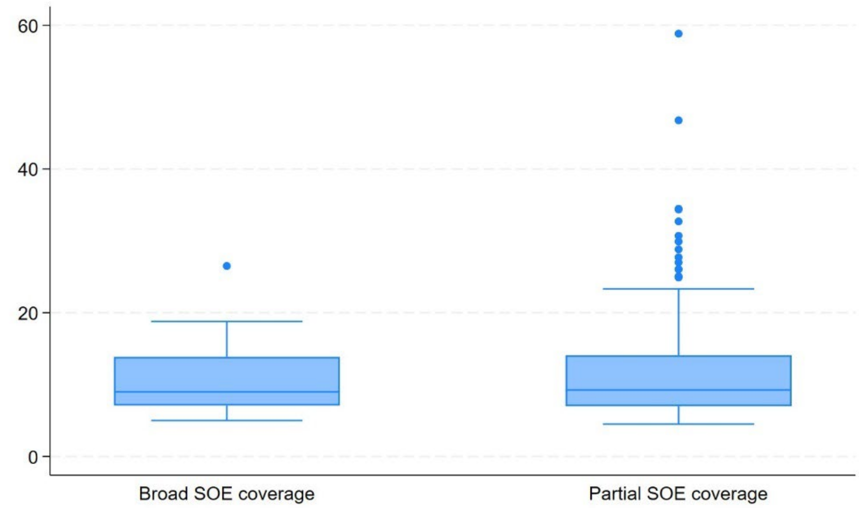
Notes: The deviations are calculated relative to the thresholds corresponding to countries' debt carrying capacity and averaged over the first five forecast years. Debt coverage groups: Full (general government, guarantees and non-guaranteed SOE debt); Near full (narrower budgetary government, guarantees and non-guaranteed SOE debt); Partial (either general or narrower budgetary government and guarantees); Minimum (either general or narrower budgetary government). The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at the given forecast horizon, while the horizontal line indicates the median.

[30] Analysis is carried out only for debt indicators, for which there is a sufficient number of countries in different debt coverage groups in the High and Moderate Risk groups.

[31] If the estimated CLs are brought into the DSA baseline for Low and Moderate external risk cases with no breaches in the baseline, this will generate a mechanical High-Risk signal in about one-quarter of them. Among those with CLs above the median value across LIC-DSF users, about half would face such mechanical risk signals.

[32] The calibration of the contingent liabilities stress test can generate a separate risk signal only for countries with no other threshold breaches under the baseline or in the other stress test scenarios.

Figure 19. LIC DSAs: Size of the Contingent Liabilities Stress Test by Coverage of SOE Debt
(In percent of GDP)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Debt coverage groups: Broad SOE coverage – guaranteed and non-guaranteed SOE debt; Partial SOE coverage – guaranteed debt only. Coverage of the universe of SOEs varies across countries.

Incorporation of Climate-Change Risks and Policies

The DSAs coverage of climate change risks and policies has been evolving with the increasing recognition of their macro-criticality, but both the scope and depth of the coverage remains uneven. The 2024 Supplementary Guidance Note provided more structured guidance on their incorporation in the baseline/ alternative scenarios, and in the distribution of outcomes around the central projections. But, given the challenges in long-term macroeconomic projections and the use of judgment, a module for climate change to aid the risk assessment and ensure cross-country consistency.

34. The coverage of climate change risks and policies in DSAs has been evolving with the increased recognition of their macro criticality. At present, DSAs reflect climate change risks and policies to the extent that they trigger mechanical risk signals in either the baseline or in stress scenarios,[33] or through the application of judgment[34] that could be informed by alternative scenarios, or both.

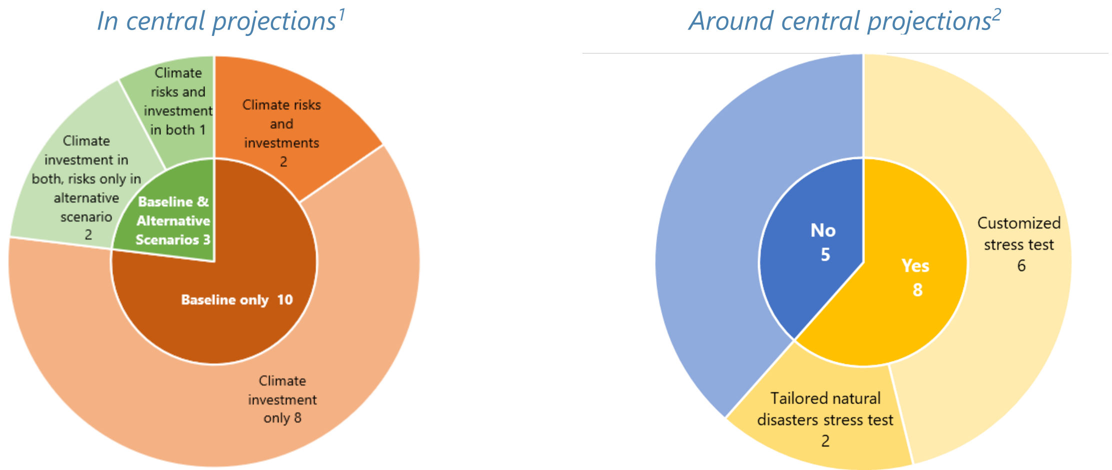
[33] The tailored natural disaster stress test is mandatory for the most vulnerable countries and can be used in other cases. A customized stress scenario can also be used to more flexibly capture the timing and transmission channels of climate change on external and overall public debt.

[34] This can be done directly or indirectly by extending the relevant time horizon of the risk assessment from 10 to 20 years, allowing more time for baseline threshold breaches to manifest.

35. The impact of climate change on central projections of debt outcomes has been mostly addressed in DSAs accompanying requests for a Fund Resilience and Sustainability Facility arrangement, but coverage has been uneven. Thirteen RSF arrangements have been approved for LICs since the arrangement's launch in 2022. In the DSAs accompanying those RSF requests, climate change issues have been covered using different tools in the framework (Figure 20). The authorities' plans for adaptation and mitigation investments in the baseline are covered in all cases. But the impact of climate change on central projections is reflected in the baseline in only three DSAs and its longer-term impact is considered in three cases with the help of alternative scenarios. The tailored natural disaster and customized stress tests were included in the DSA in two and six cases, respectively. The 2024 Supplementary Guidance Note formalized the requirement that analysis of the impact of climate change, investment, and policies on debt risks should be included in the DSAs of countries requesting an RSF arrangement and a World Bank Development Policy Operation with Catastrophic Deferred Drawdown Options (DPOs with CAT DDOs) and introduced a presumption for inclusion of such analysis in the DSAs of countries for which there are recently prepared in-depth analyses of climate change issues, such as the World Bank Country Climate and Development Reports (CCDRs) and IMF Climate Policy Diagnostics (CPD).

Figure 20. LIC-DSAs Accompanying RST Requests: Coverage of Climate Change Risks, 2022-24

(Number of DSAs)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

[1] The effects of slow-moving long-term shifts in climate, and changes to the frequency and intensity of extreme weather events on debt risks on central projections can be analyzed in the baseline or in alternative scenarios.

[2] The impact of extreme but plausible weather events on the distribution of outcomes around the central projection can be analyzed with the help of stress tests.

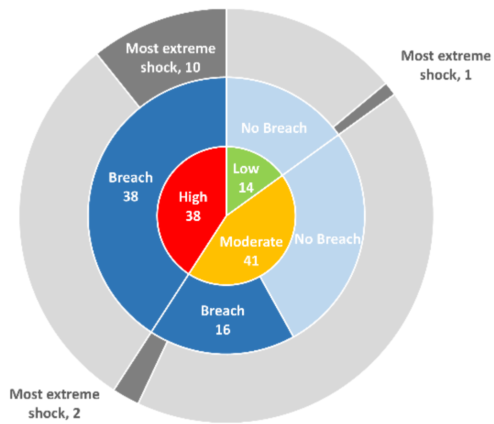
36. Across all DSAs, the tailored natural disaster stress test is the main vehicle for capturing climate-related debt risks, but it generates a mechanical risk signal only in the absence of baseline threshold breaches. It has been used in forty percent of the analyzed DSAs, triggering threshold breaches in 16 out of the 41 reviewed DSAs with mechanical signals of

Moderate risk of debt distress (Figure 21).[35] The tailored natural disaster shock was considered the most extreme for any of the four external debt burden indicators in less than one sixth of these cases. The natural disaster shock was pivotal to changing the mechanical rating from low to moderate risk in only two cases, though in both the final rating was downgraded to High risk with application of judgment (see below).

37. In a number of cases, the time horizon of the risk assessment was extended to better capture climate risks. Among the 21 judgment-based downgrades of risk ratings to reflect long-term considerations (Figure 13), nine were justified by extending the time horizon of the risk assessment to allow baseline threshold breaches to manifest and justify a High Risk overall rating.

Figure 21. LIC-DSAs: Frequency and Mechanical Signal from Tailored Natural Disaster Stress Test by Final Risk Rating, 2018-23

(Number of DSAs)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

Notes: The 3-level pie chart illustrates the frequency of use of the tailored natural disaster stress test by risk rating, and within each group whether it resulted in threshold breaches, and if so whether these breaches made the stress test the most extreme among those used in the analysis.

Ensuring Continuity in DSA for Frontier Markets Transitioning from LIC to MAC Status

The use of two related but separate frameworks for analyzing the risk of sovereign stress and debt sustainability for frontier markets creates the need for ensuring continuity in the assessment when countries graduate from one framework to the other. Ongoing work aims to identify the likelihood and magnitude of potential cliff effects in the risk and sustainability assessments, with the view of minimizing them as part of this review.

[35] By design, stress tests generate a mechanical risk signal only if there are no threshold breaches in the baseline, which is generally the case for countries rated at Low or Moderate risk of debt distress. When the mechanical signal is already at High risk such that there are already threshold breaches in the baseline, the stress tests do not provide any additional signal besides helping to contextualize sources of risks.

38. The use of two related but separate frameworks for analyzing the risk of sovereign stress and debt sustainability for frontier markets creates the need for ensuring continuity in the assessment when countries graduate from one framework to the other. Under Fund policies, the Sovereign Risk and Debt Sustainability Framework (MAC SRDSF) is the standardized approach for analyzing the risk of sovereign debt-related stress and public debt sustainability for market access countries (MACs), which refers to countries that are not eligible for the Fund's PRGT facilities.[36] All advanced economies and most emerging market economies are considered MACs.[37] For all other countries, which mainly rely on concessional financing, the joint IMF-World Bank Debt Sustainability Framework for Low Income Countries (LIC-DSF) is the applicable analytical tool. These countries may eventually graduate from concessional financing and access capital markets on a durable and stable basis, in which case they would migrate to the MAC SRDSF.

39. Ongoing work aims to identify the likelihood and magnitude of potential cliff effects in the risk and sustainability assessments, with the view of minimizing them as part of this review. Preliminary analysis of a non-representative sample of Frontier LICs points to the possibility that countries may experience a change in the medium-term risk assessments when graduating from the current LIC-DSF despite no fundamental change in the underlying risk, a result driven in part by the single risk assessment in the LIC-DSF for all time horizons. Ongoing work on the empirical significance of this and other transition issues will help inform the exploratory analysis of potentially missing country characteristics from the DCC mapping, with the view of minimizing any cliff effects in the risk and sustainability assessments. As necessary, other solutions, such as allowing for a more continuous risk assessment, will also be explored.

MAIN TAKEAWAYS AND FUTURE DIRECTIONS

A. Main Takeaways from Analysis of LIC-DSF Performance

40. Analysis of the predictive power and key design features of the LIC-DSF suggests that while it has withstood the test of time, the changing debt landscape has created pressure points that need to be addressed:

- **Performance and robustness of the LIC-DSF.** The out-of-sample performance of both underlying model and final risk ratings for external and overall public debt are in line with the model in-sample calibration, suggesting that the key design features of the framework, taken as a whole, including the calibration of debt-burden thresholds and countries grouping by DCC, continue to perform broadly as intended. All external debt burden indicators contribute to the

[36] A country is added to the list of PRGT-eligible members, if its GNI per capita is below a specific threshold (a multiple of the IDA operational cut-off) and it does not have the capacity to access international financial markets on a durable and substantial basis. Graduation from PRGT eligibility is based on a country meeting a certain income criterion or a market access criterion, and in all cases, provided that the country does not face serious short-term vulnerabilities, including as indicated by the latest available DSA. Addition to and graduation from the list of PRGT-eligible countries are assessed through periodical reviews (usually every two years).

[37] In certain cases, a MAC country can elect to use the LIC-DSF.

aggregate risk signal, though their marginal contribution differs by time horizon. Re-estimation of the external debt risk model with updated data suggests that while it continues to signal distress with a similar performance, the increased heterogeneity across LICs points to the opportunity to revise the framework to improve its predictive capacity and the need for adjustments to some of the key design features of the LIC-DSF. The main Threshold Approach continues to perform significantly better than the supplementary Probability Approach, which is rarely used in practice.

- **Domestic debt vulnerabilities.** The analysis also points to the need to strengthen the assessment of overall public debt distress, especially in light of the increased importance of domestic debt vulnerabilities in many LIC-DSF countries.

- **Incorporation of climate-change risks and policies.** The DSAs coverage of climate change risks and policies has been evolving with the increasing recognition of their macro-criticality, but both the scope and depth of coverage remains uneven, pointing to the need for developing additional tools to aid the assessment.

- **Reflecting country heterogeneity.** The grouping of countries by debt-carrying capacity (DCC) continues to play a useful role in keeping the framework simple and easy to use. "Cliff effects" from changes in countries DCC have not had a significant impact on final risk ratings and there is little evidence of instability in a country's DCC. Cliff-effects may nevertheless create some undesirable discontinuities in fiscal policy advice if not internalized in advance. At first cut, Frontier LICs do not appear to be disproportionately affected by potential gaps in the DCC definition, as their DCC generally aligns well with market signals of credit risk, and they are affected to a similar degree as other LICs by "false alarms" generated by the model. However, the increased heterogeneity among Frontier and other LICs, suggests a need for further analysis of existing and potentially missing variables from the DCC classification and other DCC modalities.

- **Judgment-based assessments.** Whereas the use of judgment has generally improved outcomes for countries that differ from the representative LIC, its use for reflecting long-term considerations in the risk assessment, against the backdrop of lack of time differentiation of risk ratings, has contributed to a higher rate of false alarms. Consideration of recent debt restructuring cases suggests that the sustainability assessments would benefit from being informed by model-based signals.

- **Forecasting challenges.** Projections in the short- and medium-term appear fairly accurate, but the drawing of definitive conclusions is complicated by the COVID-19 shock. Forecasts at longer horizons show some optimism bias in exports and revenue forecasts.

- **Debt coverage.** Comprehensive debt coverage remains key for properly capturing debt risks as borrowing outside the budgetary government has increased and state-owned enterprises (SOEs) remain a source of heightened fiscal risks in LICs. But cross-country differences in the completeness of debt coverage, including debt of SOEs, remain a challenge, potentially affecting the quality of some of the risk signals. Such concerns are in part mitigated by the more conservative calibration of the contingent liabilities stress test for some of the affected countries.

- **Continuity in assessment when countries graduate from the LIC-DSF.** Preliminary analysis of a non-representative sample of Frontier LICs points to the possibility that some countries may experience a change in the medium-term assessments when graduating to the MAC SRDSF.

B. Future Directions in LIC-DSF Review

41. The identified pressure points in the LIC-DSF design suggest there may be benefit in pursuing more granular approaches to the sustainability assessment and time horizon of risk ratings, as well as better accounting for domestic debt vulnerabilities and the increased heterogeneity across LICs. Without prejudicing the outcome of the next stage in the analysis, the following is a non-exhaustive list of options to explore in the quest to improve and future-proof the LIC-DSF:

1) **Revisiting and revising the underlying model specification.** Revising the model specification will seek to apply econometric advancements to improve the model estimation and performance, including in light of the updated data,[38] and better pinpoint the impact of macroeconomic fundamentals on the risk of debt distress, including:

    - **Exploring options to refine the grouping of a more heterogenous group of Frontier and other LICs by DCC in a consistent manner including:**

        - Further analysis of existing and potentially missing country characteristics from the CI-score, such as the size of the economy, domestic financial development, and interaction terms between global and country-specific factors;

        - Re-visiting the rules for reducing the heterogeneity of country fundamentals by mapping countries into DCC categories; and

        - Exploring elevating the role of the Continuous Model in well-defined instances, such as in cases where the CI-scores are within a narrow band around the DCC cut-off values (which may help moderate any undesirable impact of cliff-effects on policy advice) or for Frontier and other LICs where there is evidence that the DCC omits important characteristics.

    - **If the performance gap relative to a pure probability model proves to be large, expand the analysis to consider the merits of adopting such an approach subject to further guidance from the IMF and WB Boards.** This would arise if, in the process of revisiting and

[38] The global nature of the COVID-19 shock and the coordinated policy response to it diminish the information content of post-2019 data for the evolution of debt vulnerabilities in the future, which should be properly reflected in updating the models.

revising the underlying econometric model, refinements to the DCC-based Threshold Approach, possibly combined with an elevated role for the Continuous Model in well-defined instances, are insufficient to close the performance gap**.** The merits of adopting a pure probability approach would need to be weighed against the added complexity that would introduce**.** Staff will revert to the IMF and WB Boards for further guidance in this instance.

- **Exploring ways to better support fiscal space assessments,** including by extending the structured guidance in the 2024 LIC-DSF Supplementary Guidance Note (IMF and WB, 2024) on incorporation of the positive feedback effects from investment and policies in climate mitigation and adaption to other key developmental areas, such as achieving SDGs and preserving natural capital.

- **In the final model specification, consideration could be given to revisiting the weights assigned to false alarms and missed distress episodes in the model calibration.**

- **Re-evaluating the continued relevance of the Probability Approach, in light of its relatively poor predictive power and lack of use by teams.**

2) Exploring ways to improve key design features of the LIC-DSF:

- **Revisiting the contribution of the four external debt burden indicators in identifying debt distress episodes, with the view of possible simplification of the framework.**

- **Considering the feasibility of modelling the effects of domestic debt vulnerabilities, building upon the enhanced guidance on incorporating their impact in the overall public debt risk assessment in the 2024 LIC-DSF Supplementary Guidance Note, including by:**

    - Supporting the overall debt risk assessment with mechanical risk signals from fully-fledged econometric models of the existing group of overall debt indicators, or of potential alternative indicators, such as the overall public debt-to-revenues ratio and interest-only debt service indicators, as well as the integration of these models in the construction of CI-scores; and

    - Further developing the tools to analyze domestic debt vulnerabilities introduced by the Supplementary Guidance Note.

- **Exploring ways to bring in long-term considerations into the risk assessment via the use of judgment in a consistent manner across countries without triggering false alarms, including:**

    - Time-differentiation of the risk signal to improve interpretation and precision of risk ratings; and

- Development of a long-term climate change module that teams can use to meet the required coverage of climate issues.[39]

- **Developing models for mechanical signals for debt sustainability to better inform the application of judgment towards a final assessment.** This would also allow a clearer delineation between the assessment of the risk of sovereign stress and debt sustainability.

3) Other design considerations:

- **Options to improve forecast accuracy include:**
  - Strengthening the realism tools, including by leveraging on advances made in the context of the MAC SRDSF and expanding their coverage to all macroeconomic indicators that determine debt burden ratios; and
  - Exploring the merits of a more conservative approach comprising of the addition of an explicit flag for Degree of Confidence in the medium-term risk and sustainability assessments, based on results from realism tools, the breadth of debt coverage, results from the contingent liabilities stress test, and history of debt surprises;

- **Exploring ways to encourage comprehensive debt coverage, including debt of SOEs and other elements of the general government unless they pose limited fiscal risks, and broader debt transparency, while seeking ways to reduce the operational burden to country teams, including:**
  - Enhancing guidance on the targeted debt perimeter, including the conditions for excluding debt of SOEs and other elements of the general government that pose limited fiscal risks, leveraging where relevant on IMF and WB capacity development (CD);
  - Exploring the feasibility of accounting for recent improvements in debt transparency in the DCC classification, potentially using information from the annual World Bank's Debt Reporting Heatmap;[40]
  - Incentivizing the targeted debt coverage by exploring the feasibility of leveraging data from existing databases and simplifying the rules for incorporation of SOE debt; and
  - Further strengthening of the contingent liabilities stress test, for example by including a minimum calibration derived from existing databases.

[39] For instance, as outlined in the Supplementary Guidance Note.

[40] The Heat Map presents an assessment of debt transparency based on the availability, completeness, and timeliness of public debt statistics and debt management documents posted on national authorities' websites. The assessment is updated annually for all IDA-eligible countries. See www.worldbank.org/en/topic/debt/brief/debt-transparency-report/2023.

- **Ensuring continuity in assessment when countries graduate from the LIC-DSF.** An important aspect of the work on refining the classification of LICs by DCC, and the broader improvements in the LIC-DSF model (see above), including the Continuous Model, will be to mitigate the risk of potential cliff effects in risk ratings and sustainability assessments of Frontier LICs that may transition to use the MAC SRDSF. Exploring the feasibility of harmonizing the definition of overall public debt distress episodes in the LIC-DSF and the MAC SRDSF could aid these efforts. Depending on the likelihood of any residual discontinuity between the two frameworks for borderline country cases, consideration could also be given to a more tailored application of the tools for Frontier LICs, such as placing a higher weight on the cost of potential loss of market access as a result of false alarms in the model calibration. A related effort will take a closer look at the need for enhanced guidance on the application of existing SRDSF tools for borderline Frontier MACs.

- **Revisiting other technical aspects of the framework,** including but not limited to revisiting the setting of the discount rate (currently at 5 percent), consideration of how state-contingent debt instruments (SCDIs) will be evaluated in the new framework, the impact on the DSA risk assessment of the choice between currency- and residency-based approach to defining external public debt, and the implications for the setting of debt restructuring targets of the potential development of a mechanical signal to inform the final assessment of debt sustainability.

42. The work on updating and future-proofing the LIC-DSF outlined above will be guided by several key principles:

- **Keeping the framework transparent and easy to use.** The review will seek to ensure the tractability of the framework and templates, with the view of limiting transition costs and capacity development needs.

- **Reflecting analytical advances in the field.** In addressing the causes of emerging challenges, we will be drawing upon lessons learned from the recent MAC SRDSF review and early experience with its implementation.

- **Maintaining backward-looking continuity of the revised framework,** by ensuring that changes in debt risk and sustainability assessments relative to the existing LIC-DSF are well understood and economically sound. In general, this would imply minimizing the number of countries that would automatically be subject to a mechanical upgrade or downgrade relative to their latest assessment under the current LIC-DSF.

43. Staff propose to provide a further update to the IMF and WB Boards on the LIC-DSF review, planned for Summer 2025. In addition to presenting further work on different aspects of the LIC-DSF performance, it will also include the findings on feasibility of the various areas of future work outlined above and their potential to improve and future-proof the framework.

44. A robust program of engagement with key stakeholders will continue to be an integral part of the review. Given the importance of the framework for the IMF and the World Bank policies and country engagement, as well as for creditors and borrowers, a comprehensive engagement strategy with both internal and external stakeholders will remain an integral part of the review. In addition to the regular engagement with the IMF and WB Boards, staff plans outreach and listening sessions with CSOs, creditor and borrower forums, and experts in the field.

ISSUES FOR DISCUSSION

- How do Executive Directors view the performance of the LIC-DSF since the last review?
- Do the questions, listed in Section I, ¶6 and pursued in the analysis of the LIC-DSF performance in this paper, cover the main areas pertinent for the review?
- Do Executive Directors concur with the main takeaways from the analysis of the LIC-DSF performance presented in Section IV.A?
- How do Executive Directors view the areas of future work outlined in Section IV.B, the feasibility of which staff propose to explore in updating the framework?
- Are there areas that merit increased focus?

References

International Monetary Fund and World Bank, 2012, *Revisiting the Debt Sustainability Framework for Low-Income Countries*, available at: www.imf.org/-/media/Websites/IMF/imported-full-text-pdf/external/np/pp/eng/2012/_011212.ashx, (Washington: International Monetary Fund and World Bank).

———, 2017, "Review of the Debt Sustainability Framework in Low-Income Countries: Proposed Reforms," IMF Policy Paper and WB Board Report, available at: www.imf.org/-/media/Files/Publications/PP/2017/pp082217lic-dsf.ashx and documents1.worldbank.org/curated/en/823731506617907804/pdf/120124-BR-OUO-until-october-2-10-AM-IDA-R2017-0299.pdf, (Washington: International Monetary Fund and World Bank).

———, 2018, "Guidance Note on the Bank-Fund Debt Sustainability Framework for Low Income Countries," IMF Policy Paper and WB Board Report, available at: www.imf.org/-/media/Files/Publications/PP/2017/pp122617guidance-note-on-lic-dsf.ashx and documents1.worldbank.org/curated/en/513741518471205237/pdf/LIC-DSF-SGN-2017-Clean-Feb0718-02082018.pdf, (Washington: International Monetary Fund and World Bank).

———, 2024, "Supplement to 2018 Guidance Note on the Bank-Fund Debt Sustainability Framework for Low Income Countries," IMF Policy Paper and WB Board Report, available at: https://www.imf.org/-/media/Files/Publications/PP/2024/English/PPEA2024039.ashx, (Washington: International Monetary Fund and World Bank).

International Monetary Fund, 2024, "Mozambique: 2024 Article IV Consultation, Fourth Review under the Three-Year Arrangement under the Extended Credit Facility", Country Report No. 2024/219, July, available at: www.imf.org/-/media/Files/Publications/CR/2024/English/1mozea2024002-print-pdf.ashx (Washington: International Monetary Fund).

World Bank, 2023, "The World Bank's Role in and Use of the Low Income Country Debt Sustainability Framework," Independent Evaluation Group Report, available at: https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/LIC-DSF.pdf, (Washington: World Bank).

INTERNATIONAL MONETARY FUND

November 20, 2024

REVIEW OF THE BANK-FUND DEBT SUSTAINABILITY FRAMEWORK FOR LOW INCOME COUNTRIES—EXPLORING LIC-DSF PERFORMANCE AND SCOPE FOR IMPROVEMENTS—BACKGROUND PAPER

EXECUTIVE SUMMARY

This Background Paper provides technical information to accompany the main paper on Review of the Bank-Fund Debt Sustainability Framework for Low Income Countries—Exploring LIC-DSF Performance and Scope for Improvements. This material, presented in four chapters, covers the following issues:

- Analysis of technical aspects of the LIC-DSF econometric model and implementation;
- Analysis of the debt-carrying capacity as tool to parsimoniously capture country heterogeneity;
- Analysis of forecast accuracy;
- Analysis of debt perimeter issues.

Approved By
Guillaume Chabert, (IMF) and Manuela Francisco (WB)

Prepared by a team from the IMF's Strategy, Policy, and Review Department, led by Plamen Iossifov and coordinated by Karim Foda, consisting of Tamon Asonuma, Christopher Dielmann, Yukun Ding, João Capella-Ramos, Miguel Ricaurte, Moustapha Mbohou Mama, Jean-Claude Nachega, Sanan Mirzayev, Jasmin Sin, and Igor Zuccardi, under the overall guidance of Allison Holland (all SPR); and a team from the World Bank's Macroeconomics, Trade and Investment Global Practice, led by Luca Bandiera, consisting of Anna Corcuera, Mariela Caycho Arce, Wissam Harake, Juan Hernandez, Sebastian Horn, Muazu Ibrahim, David Mihalyi, Mellany Pintado Vasquez, Madi Sarsenbayev, Tessy Vasquez Baos, and Clemens Graf Von Luckner under the overall guidance of Frederico Gil Sander. Data analytics and visualization support was provided by Chen Chen, Zhuo Chen, Zhengyuan Xia, and Arthur Xie. Administrative assistance was provided by Eiman Afshar and Claudia Isern.

CONTENTS

INTRODUCTION

1. **This background paper delves deeper into the analysis of technical aspects of the LIC-DSF econometric model and implementation,** the use of debt-carrying capacity as tool to parsimoniously capture country heterogeneity, forecast accuracy, and debt perimeter issues.

TECHNICAL ASPECTS OF ECONOMETRIC MODEL AND LIC-DSF IMPLEMENTATION

A. 2017 Review Model

Model Overview

2. **The LIC-DSF is based on an econometric model that provides a one-year-ahead prediction of the risks of external and overall debt distress.** The model consists of four PROBIT regressions that include one debt burden indicator per specification and a common set of macroeconomic fundamentals. The probit regressions take the form of:

$$Prob(Debt\ Distress) = \Phi^{-1}(\alpha_j + \gamma_j d_j + \sum_{k=1}^{6} \beta_{j,k} X_k), \text{ where} \quad (1)$$

Debt Distress – dummy variables for the occurrence of external and overall public debt distress;

Φ^{-1} – inverse of the normal distribution;

$\boldsymbol{\alpha_j}$ – constant term;

$\boldsymbol{d_j}$ – four different external debt burden indicators (PV of external public debt to GDP and exports, and external public debt service to total revenues (excl. grants) and exports) are used to predict external public debt distress; one overall debt burden indicator (PV of overall debt to GDP) is used alongside the combined risk signal for external public debt distress to predict overall public debt distress;

X_k – other structural and macroeconomic fundamentals to capture countries' debt carrying capacity (CPIA, country real GDP growth, reserves-to-imports ratio in levels and squared, remittances as a share of GDP, and world growth);

$V_\beta < 0$ – The other structural and macroeconomic fundamentals are defined in such a way that higher values correspond to higher debt-carrying capacity, and hence lower distress probability. As a result, all of their coefficients are negative;

$j = 4$ – The four probits for external debt distress are estimated separately, while the one including the PV of overall debt to GDP is calibrated afterwards.

3. The dependent variables in the LIC DSF model capture the onsets of distress and non-distress episodes. They are based on a set of proxy dummy variables that combine information about the start of actual debt restructurings and timing assumptions (when the start date is not known) with other distress signals, such as outright defaults, accumulation of significant arrears, and large upfront disbursements under an IMF arrangement. The dependent variable for overall public debt distress is defined by combining external and domestic debt distress episodes. Only the start of the distress and non-distress episodes are coded as 1 and 0, respectively, with the rest of the observations in their respective spells coded as missing values. Distress episodes are determined by a distress signal from any of the proxy dummy variables irrespective of whether there are missing values for any of the other proxies, with additional rules about how to distinguish separate distress episodes, when distress signals occur in close succession. Non-distress episodes are defined as non-overlapping three-year spells with no distress signal from any of the proxy variables. The dependent variable is assigned a missing value if the values of the distress proxies are a mix of zeros and missing values.

4. In the 2017 review, the regression sample spanned the period 1970-2014 and covered 80 current and former LICs. The sample included 61 of the 68 current LICs, with the rest not included due to missing data on external debt burden indicators, and 19 former LICs that have since graduated to market-access status and no longer use the LIC-DSF. The use of a broad sample of LICs adds to the number of distress episodes on which the models are trained, allowing for their more precise estimation, and ensuring that the models continue to reflect the interlinkages between the dependent and explanatory variables, as countries develop.

5. The four continuous external debt distress PROBITs are estimated individually.[1] Distress signals are then generated by the untransformed Continuous Model and under the main "Threshold Approach and the supplementary "Probability Approach":

- **Continuous Model.** The regression coefficients are used to estimate empirical distress probabilities for each observation in the sample. Distress probability cut-offs (above which the model predicts a distress event) are then derived by minimizing a "loss function," in which the proportion of "missed crises" and "false alarms" by the four PROBITs enter with weights of 2/3 and 1/3, respectively. The obtained probability cut-offs are not used directly but serve as inputs for the Probability Approach.

- **The "Threshold Approach" is at the core of the LIC-DSF.** Under this approach, the absolute values of the coefficients on each of the explanatory variables from the continuous model, other than the debt burden indicators, are averaged across the 4 PROBITs and are used together with the 10-year averages of these variables to construct a Composite Indicator (CI) as of 2014. For

[1] Both the LHS and RHS variables enter the regressions as country-year observations.

each year in the sample, all 80 LICs are then assigned CI-scores equal to either the 25th, 75th or 50th percentile cut-offs for the CI values of the 61 current LICs included in the sample as of 2014,[2] depending on whether the 10-year moving historical averages of the Composite Indicator in that particular year falls in the 25th, 75th, or in the middle-two quartiles of the distribution of the 2014 CI-values of the 61 current LICs, respectively. As a result, in each year of the sample countries are assigned in three different groups by debt-carrying capacity – Weak, Strong, and Medium, respectively. For each PROBIT, the empirical probability of distress is then estimated as the sum of the constant, the product of the coefficient of the debt burden indicator and its untransformed values, and countries CI-scores that enter the sum with a coefficient of minus one. The same optimization procedure as for the Continuous Model is then used to derive distress probability cut-offs. In a final step, the "High-Risk" thresholds for the debt burden indicators are derived by DCC, by inverting the distress probability cut-offs for each PROBIT model at the 25th, 75th or 50th CI-score percentile cut-offs.

- **The supplementary "Probability Approach" is included in the LIC-DSF for auxiliary use when debt indicators are close to the thresholds.** Under it, the empirical distress probabilities are estimated using the coefficients from the Continuous Model and the values of the explanatory variables, fixed at their long-term, centered average as of the time of each DSA, except for debt burden indicators that enter the calculation with their untransformed forecasted values. The empirical distress probabilities are then compared to the probability cut-offs derived from the Continuous Model.

6. The thresholds for the PV of overall debt-to-GDP ratio are not estimated directly. Under the Threshold Approach, they are obtained from an ad-hoc application of the "signal-to-noise" method, using the same CI-scores and under the assumptions of no constant and coefficient on the overall debt-to-GDP ratio of one. The "signal-to-noise" method minimizes a "loss function" calibrated the same way as the one used in the external debt distress model. There is no Probability Approach for identifying overall public debt distress, as no additional PROBIT regressions have been estimated for overall debt burden indicators.

7. The simplicity of the Threshold Approach comes at the cost of an extra degree of separation from the estimated Continuous Model. Under the main Threshold Approach used in the LIC-DSF, the values of the debt burden indicators for each LIC are compared to their respective DCC-specific thresholds, making the results easy to interpret. The mapping of countries to DCC groups is a modelling choice that aims to parsimoniously capture the heterogeneity in macroeconomic fundamentals across LICs. While a specific mapping scheme may perform this task well at the time it was selected, there is no guarantee that it will continue to do so if countries macroeconomic characteristics diverge further. This can introduce biases in the risk signals that are separate from the performance of the underlying Continuous Model.

[2] The cut-offs were set at 2.69, 2.86 and 3.05 corresponding to the values of the 25th, 50th and 75th percentiles respectively.

Model Estimation

8. The predictive performance of the LIC-DSF model is evaluated in reference to the Receiver Operating Characteristics (ROC) curves of individual PROBITs and joint selection of distress probability cut-offs. The ROC curve is a plot of the True Positive rate of correctly predicting the onset of external debt distress versus the False Positive (false alarms) rate of misclassifying non-distress occurrences, using as thresholds different predicted probabilities derived from the four PROBIT models. Following the estimation of the four PROBITs, different cut-offs for the predicted probability of a debt distress will classify differently the sample observations into distress and non-distress ones. Each PROBIT can then be graphically represented as a curve on the ROC chart, in which the distress probability cut-offs are ordered from the highest to the lowest starting at the axes origin.[3] The bigger the vertical distance between a data point summarizing the model predictive power at a given probability cut-off and the diagonal line is, the better the performance of the model.[4]

9. All four PROBITs for external debt distress risk have strong predictive power, but the ones relying on export-based debt burden indicators do not add to the overall performance at the margin. The top panel of Figure 1 plots the ROC curves of the individual PROBITs obtained in the first round of estimating the model with the 1970-2014 dataset used in the 2017 Review. Also shown are the distress probability cut-off points selected by the optimization procedure to minimize the Loss function of the four PROBITs under the Threshold Approach. All PROBITs exhibit false alarm rates in the range of 25-31 percent and True Positive rates of over 75 percent, both of which are within the range of benchmarks for a successful crisis prediction model (top charts in Figure 3). The combination of the four PROBITs has in-sample True Positive rate of over 80 percent and False Positive rate of around 40 percent. However, the marginal contribution of the PROBITs for the PV of external debt-to-exports and external debt service-to-exports is negative, as conditional on the other PROBITs not raising a red flag, they do not help identify any missed debt distress episodes, while adding to the rate of false alarms (bottom panel of Figure 1) of the combination of the four PROBITs.

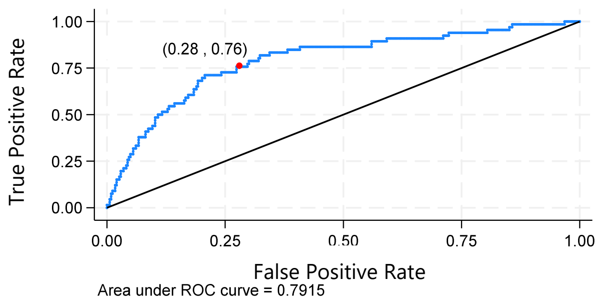
[3] At the axes origin (corresponding to zero False Alarms and zero True Positive Rate), the cut-off for the predicted probability is 1.0, at which point the model will correctly identify all non-distress occurrences but will mis-classify all distress events as non-distress ones. As the probability cut-off is lowered, the model is able to identify progressively more distress occurrences at the expense of misclassifying an increasing share of non-distress ones.

[4] The vertical distance is equal to the difference between the True Positive and False Positive rates generated by the distress probability cut-off. The 45-degree diagonal line corresponds to the ROC curve of an uninformative model corresponding to flipping a coin for each sample observation.

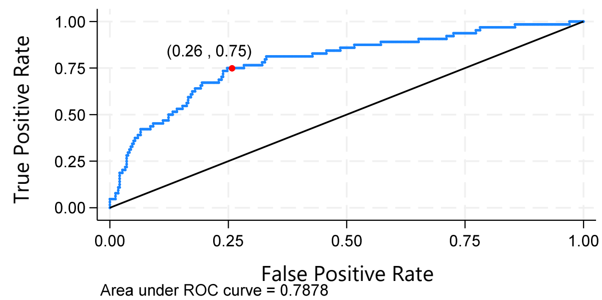
Figure 1. Individual PROBITs Performance under Threshold Approach and Contribution to Identifying Distress and Non-Distress Episodes in 2017 Review Dataset

ROC Curve Charts (Index)

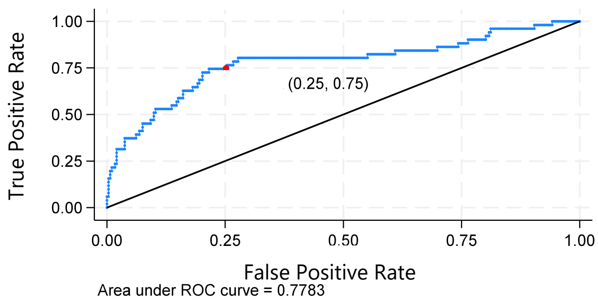
PV External Debt-to-GDP



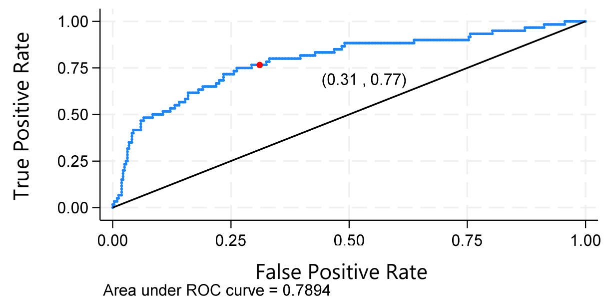
PV External Debt-to-Exports



External Debt Service-to-Revenue



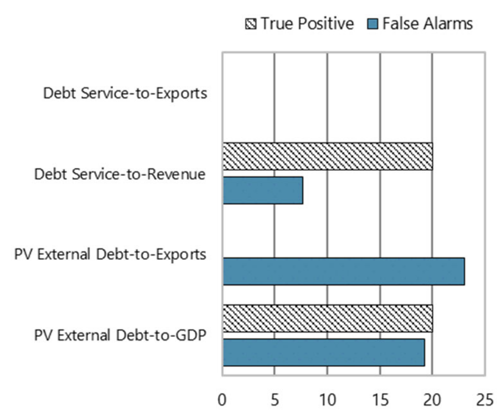
External Debt Service-to-Exports



Loss Function of Jointly Calibrated PROBITs (Index)

Statistics	Values
Missed Crises	0.18
False Alarms	0.38
Loss Function 1/	0.24

Marginal True and False Positive Rates of Individual PROBITs 2/ (Percent)



Source: Staff calculations based on the 2017 LIC DSF Econometric Model.

Notes: The four charts at the top show the Receiver Operating Characteristic (ROC) curves of the individual PROBITs, obtained in the first-round of the estimation, together with the distress probability cut-off point (red dot) selected by the optimization procedure to minimize the Loss function of the four PROBITs. The probability cut-offs are then inverted into DCC-specific thresholds. The estimates of the area under the ROC curve (AUC) can be used to compare the discriminatory power of different regression specifications (see Pepe, Margaret, Gary Longton, and Janes, Holly, 2009, "Estimation and Comparison of Receiver Operating Characteristic Curves," *The Stata Journal*, 9(1)). The AUC of the 45° diagonal line, which represents the ROC curve of an uninformative model, is 0.5. The maximum AUC is equal to 1.

1/ Weighted sum of missed crises (with weight of 0.67) and false alarms (with a weight of 0.33).

2/ True and False Positive Rates of individual PROBITs conditional on any of the other PROBITs not raising a red flag.

Issues Arising in Model Estimation

10. The optimization procedure used to jointly derive distress probability cut-offs from the four PROBITs is complex and its outcome is sensitive to how the dependent variable is defined, which can open a wedge between the in-sample calibration of the model and its performance in the LIC-DSF implementation. In the econometric model, only the start of the distress and non-distress episodes are coded as 1 and 0, respectively, with the rest of the observations in their respective spells coded as missing values. The model calibration based on this definition is optimized to correctly classify only the onsets of 3-year, non-overlapping, distress, and non-distress episodes, and not all country-year observations in the sample. While the focus on the onset of debt distress is standard in the crisis literature, the use of only the onsets of non-distress episodes is less common. It significantly reduces the sample of non-distress country-year observations and opens a wedge between the characteristics of the sample used for in-sample calibration of the model and the sample of its out-of-sample usage in DSAs. This is because, in practice, debt sustainability analyses are carried out in different years and the assessment results in a classification of countries as either being in distress or not experiencing one at a given point of time. As a result, the time series of final risk ratings more closely resemble country-year observations rather than onsets of spells of non-distress and distress episodes. The sensitivity of the optimization procedure to the definition of the dependent variable is illustrated in Figure 2, which presents the results of estimating the model with country-year non-distress occurrences and onsets of distress episodes in the 1970-2014 dataset used in the 2017 Review. Results show that using the same Loss Function as in Figure 1, the selected cut-off probabilities yield considerably lower false alarms rate (22 percent vs 38 percent) at the expense of higher rate of missed crises (38 percent, compared to 18 percent) (Figures 1 and 2).[5] In light of these findings, one can expect to see a higher rate of false alarms in the implementation of the LIC-DSF (that deals with country-year non-distress observations) compared to the in-sample calibration of the model (estimated with their onsets only). In practice, this effect may be offset by other features of the LIC-DSF implementation, such as the use of judgment to discount short-lived or marginal breaches, etc.

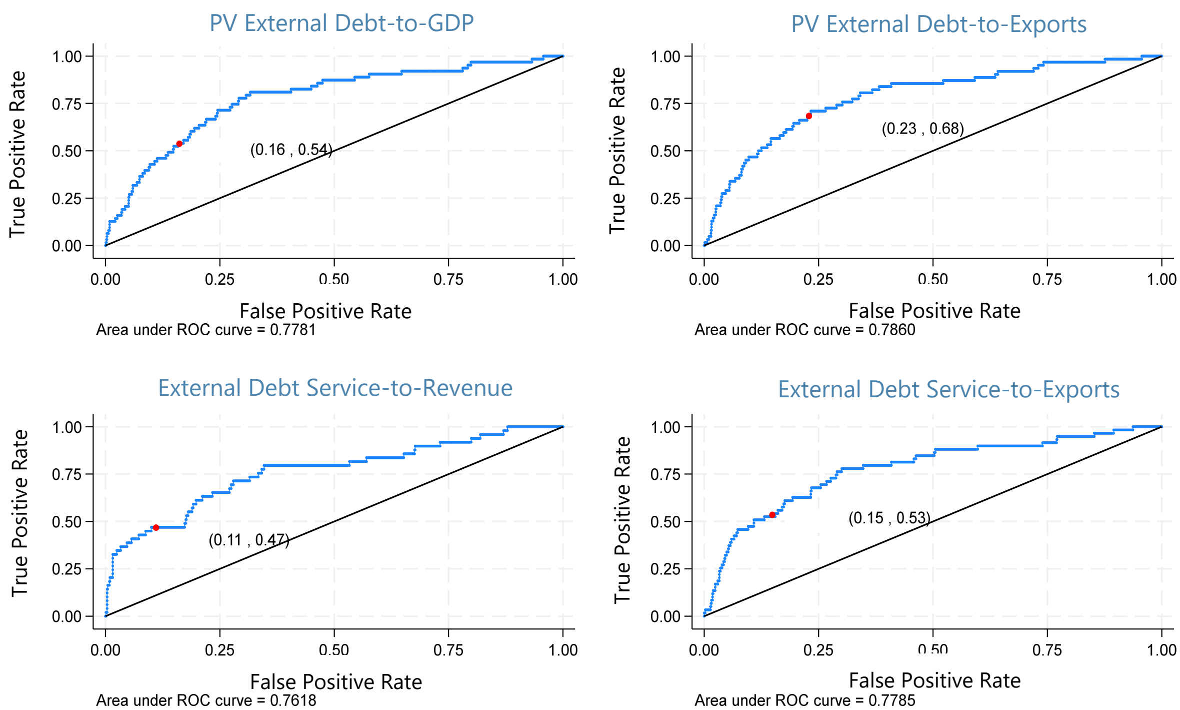
11. The model performance under the Threshold Approach is robust to alternative LIC samples used to derive the CI cut-off values for mapping countries into DCC groups. In the 2017 Review, these cut-off values were calculated with data for the 61 current LICs included in the sample. As a result, the 61 current LIC countries included in the data sample were classified into Weak, Medium and Strong DCC groups in 25:50:25 proportions. In the LIC-DSF implementation, the CI values for all 69 LIC-DSF users are being compared to the CI cut-off values, calibrated on 61 countries. As a result, the initial proportion of LIC-DSF users classified into Weak, Medium and Strong DCC groups was different from the 25:50:25 benchmark rule, with more countries ending up being classified in the Weak DCC category. However, re-estimation of the 2017 Review model under a

[5] This is due to the fact that under the revised definition of the dependent variable, the share of non-distress episodes in the sample, which can potentially be mis-classified as distress ones (and hence worsen the Loss Function) is larger. In choosing the distress probability cut-offs, for the same weights of missed distress and false alarms, the optimization procedure will try to reduce the penalty from the increased rate of false alarms by increasing the cut-off values, which results in more distress episodes being missed.

revised DCC classification scheme based on all 69 current LIC-DSF users shows little change in the predictive power of the model, suggesting that a better alignment between how the model is calibrated and used is feasible.

Figure 2. Sensitivity of Distress Probability Cut-offs under Threshold Approach to Definition of Non-Distress Episodes in 2017 Review Dataset

ROC Curve Charts (Index)



Loss Function of Jointly Calibrated PROBITs (Index)

Statistics	Values
Missed Crises	0.38
False Alarms	0.22
Loss Function 1/	0.33

Source: Staff calculations based on the 2017 LIC-DSF econometric model.

Notes: The four charts at the top show the Receiver Operating Characteristic (ROC) curves of the individual PROBITs, obtained in the first-round of the estimation, together with the distress probability cut-off point (red dot) selected by the optimization procedure to minimize the Loss function of the four PROBITs. The probability cut-offs are then inverted into DCC-specific thresholds. The estimates of the area under the ROC curve (AUC) can be used to compare the discriminatory power of different regression specifications (see Pepe, Margaret, Gary Longton, and Janes, Holly, 2009, "Estimation and Comparison of Receiver Operating Characteristic Curves," *The Stata Journal*, 9(1)). The AUC of the 45° diagonal line, which represents the ROC curve of an uninformative model, is 0.5. The maximum AUC is equal to 1.

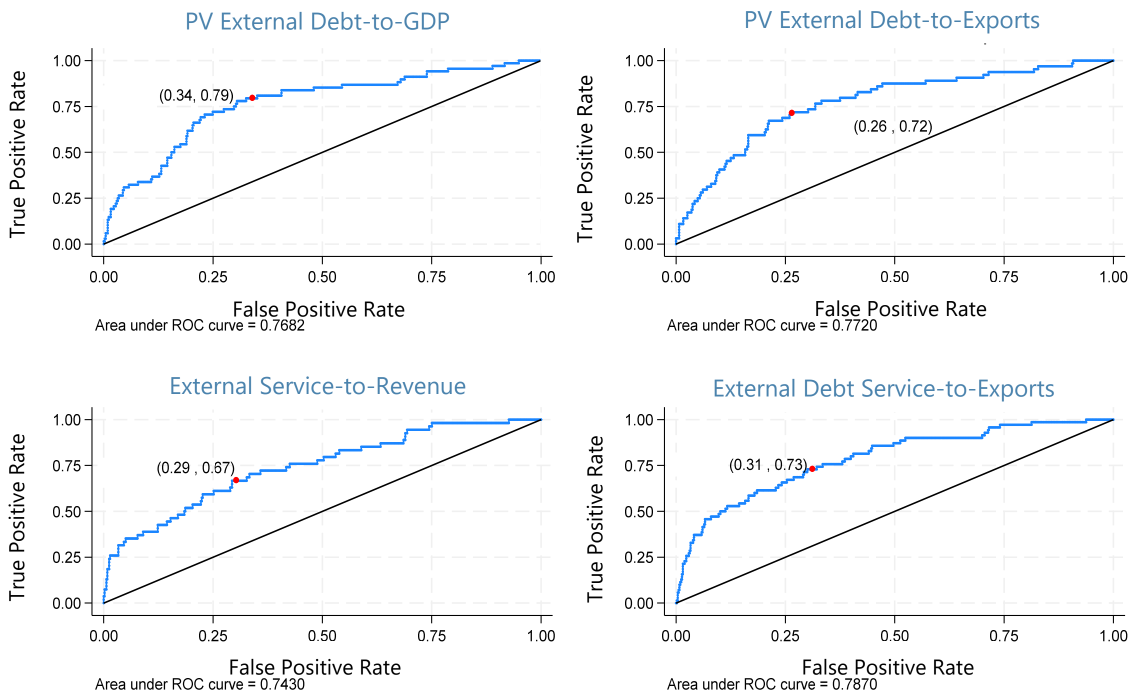
1/ Weighted sum of missed crises (with weight of 0.67) and false alarms (with a weight of 0.33).

12. The model calibration is further affected by the revision of historical data and sample extension to 2022. The historical time series of the dependent and explanatory variables have been revised to reflect changes in existing data sources and the availability of new ones. An important source of revisions is updates of historical GDP series in the WEO database for a number of LICs, due to rebasing of nominal GDP and expansion of GDP series coverage of the gray economy. Newly available data sources have also changed the frequency and timing of distress signals. Comparison of the results from estimating the model under the Threshold Approach with the 2017 Review dataset and the revised data for 1970-2014 suggest that while the model performance remains broadly unchanged, as evidenced by the sustained high values of the PROBITs Area Under the ROC Curve (AUC), the derived distress probability cut-offs change, implying changes in the estimates of the debt indicator thresholds (Figure 3).

13. Estimation of the external debt risk model with the revised and extended data suggests that data revisions and the increased heterogeneity across LICs in the last decade have weakened the explanatory power of macro fundamentals. The explanatory power of individual PROBITs in the Continuous Model, as captured by the regressions area under ROC curve (AUC) values, are lower when estimated with the revised and extended data (Figure 3).

Figure 3. Sensitivity of Model Calibration under Threshold Approach to Revisions of 2017 Review Dataset

ROC Curve Charts (Index)



Loss function of Jointly Calibrated PROBITs (Index)

Statistics	Values
Missed Crises	0.33
False Alarms	0.37
Loss Function 1/	0.27

Source: Staff calculations based on the 2017 LIC-DSF econometric model.
Notes: The four charts at the top show the Receiver Operating Characteristic (ROC) curves of the individual PROBITs, obtained in the first-round of the estimation, together with the distress probability cut-off point (red dot) selected by the optimization procedure to minimize the Loss function of the four PROBITs. The probability cut-offs are then inverted into DCC-specific thresholds. The estimates of the area under the ROC curve (AUC) can be used to compare the discriminatory power of different regression specifications (see Pepe, Margaret, Gary Longton, and Janes, Holly, 2009, "Estimation and Comparison of Receiver Operating Characteristic Curves," *The Stata Journal*, 9(1)). The AUC of the 45° diagonal line, which represents the ROC curve of an uninformative model, is 0.5. The maximum AUC is equal to 1.

1/ Weighted sum of missed crises (with weight of 0.67) and false alarms (with a weight of 0.33).

B. LIC-DSF Implementation

Overview of Implementation Modalities

14. In the LIC-DSF implementation, the Threshold and Probability Approaches rely on forecasts and different transformations of explanatory variables, and the use of judgment, introducing a further degree of separation from the estimated Continuous Model. Mechanical risk signals are generated when a debt burden indicator breaches its threshold for more than one year at any time in the first 10-years of projections under the baseline (resulting in a High-Risk signal) or in standard, tailored, or customized stress-test scenarios (giving rise to a Moderate-Risk signal). The empirical distress probabilities over the forecast period are generated utilizing forecasts of the RHS variables.

- Under the Threshold Approach, the country CI-scores as of the time of each DSA are calculated using 10-year, centered, moving averages of macroeconomic fundamentals that include five historical outturns and five forecasted ones, instead of the 10-year, moving historical averages used in its calibration. The calculated CI-score is then kept fixed over the entire forecast period of each DSA.
- Under the Probability Approach, macroeconomic fundamentals other than the debt burden indicators are transformed into 16-year centered, moving averages that include five historical outturns and 11 forecasted ones. These values are then kept fixed over the entire forecast period of each DSA.

15. In determining the final risk ratings, the mechanical signals from the main Threshold Approach or the supplementary Probability Approach can be overlaid by judgment to capture country-specific considerations. Over 2018-23, judgment was applied in around one fifth of all DSAs for 30 of the 69 LIC-DSF users to reflect country-specific considerations not fully captured by the underlying model. The use of judgment has resulted in an improvement of the final risk rating in around 40 percent of the cases and in higher risk ratings in the rest. This can affect the rates of missed distress episodes and false alarms in the application of the framework separate from the model-based mechanical risk signals.

DEBT-CARRYING CAPACITY AS TOOL TO CAPTURE COUNTRY HETEROGENEITY

A. Overview

16. A key design feature of the LIC-DSF is the classification of countries by debt carrying capacity, which allows a differentiated treatment of distinct country groups while keeping the framework simple and easy to use. The 2017 LIC-DSF review expanded the measure of countries' debt carrying capacity (DCC) from the CPIA score (as a measure of policy and institutional strength) to a composite indicator (CI) score that also includes long-run averages of macroeconomic fundamentals that mainly capture external buffers. To arrive at debt burden thresholds for countries with Weak, Medium, and Strong DCC, the heterogeneity across LICs is reduced to the 25th, 50th, and 75th percentile cut-offs of CI scores, as calculated with data up to 2014 in the last LIC-DSF review, respectively (see Section I for details on the use of DCC under the Threshold Approach). In the 2017 Review, the predictive performance of the model under the Threshold Approach was closely aligned with that of the Continuous Model.

17. However, the effectiveness of the DCC mapping procedure can be affected by the increased heterogeneity among LICs in recent years. Since 2010, Frontier and other relatively wealthier LICs have decoupled from their peers to more closely resemble emerging markets (EMs), with an increased share of non-concessional debt (both external public debt owed to private creditors and domestic debt issued at market terms). As seen in Section I, this combined with the revision and extension of time series has affected the ability of the Threshold Approach, based on the existing DCC mapping procedure, to approximate the performance of the Continuous Model, necessitating revisiting of its design features. Furthermore, the increased country heterogeneity within existing DCC groups points to the need for analysis of potentially missing country characteristics that could lead to DCC misclassification.

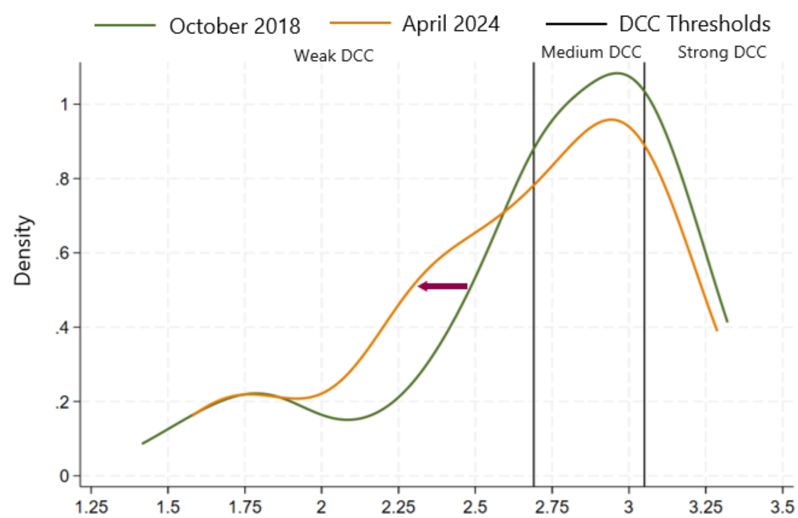
B. Developments in DCC Classification

18. The initial distribution of countries across DCC groups at the time of implementation of the current LIC-DSF in 2018 was more skewed than the model calibration. The use of the 25th, 50th, and 75th CI-score percentile cut-offs resulted in a symmetric distribution by DCC of the 61 LIC countries included in the regression sample, with half of them assigned Medium DCC and the rest split equally into the Weak and Strong DCC groups. Upon implementation of the framework in 2018, the CI scores of the additional 8 LIC-DSF users mostly fell into the Weak DCC group, resulting in nearly one-third of countries with Weak DCC. In addition, in the LIC-DSF implementation, the CI-scores are estimated using the 10-year, centered, moving averages of macroeconomic fundamentals that include five historical outturns and five forecasted ones, instead of the 10-year, historical averages used in its calibration. While an improvement over the purely backward-looking approach necessitated by the lack of forecasts in the model dataset, this resulted in the reshuffling of nine countries across DCC groups, though the number of countries in each DCC group remained broadly unchanged.

19. Since the initial round of LIC-DSF implementation, a weakening of macroeconomic fundamentals has led to a downward shift in CI scores that resulted in DCC downgrades for one quarter of LIC-DSF countries. Since 2018, the distribution of CI scores has shifted to the left. As a result, as at end-April 2024, nearly half of LIC-DSF countries (32 out of 69) are in the Weak DCC group (Figure 4). This is in line with the aims of the design of the DCC-classification scheme, i.e., to capture changes in the support from macroeconomic fundamentals to countries' ability to carry debt. Overall, 17 countries experienced DCC downgrades compared to 2018, while just 1 country experienced an upgrade.[6] Fragile and conflict-affected states (FCS) and/or small island developing states (SIDS) were disproportionately affected by downgrades into Weak DCC, accounting for 9 out 12 such cases. This points to potentially missing elements from the DCC classification that can control for specific features of FCS and SIDS, such as growth volatility. This is because average country growth may mask structural volatility in SIDS, which are more vulnerable to extreme weather events and sea-level rises. On the other hand, the DCC of larger and more diversified or resilient economies may be underestimated by the current DCC mapping, if their growth rates are more stable.

Figure 4. Distribution of CI scores across LIC-DSF countries, 2018 vs. 2024

(Density Functions)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Shown are CI score distributions for LIC-DSF countries as of Oct. 2018 and Apr. 2024. Distributions are shown as kernel probability density functions to illustrate a smoother representation of frequencies across countries.

20. Most DCC downgrades for borderline countries have been driven by weaker global conditions. The largest concentration of DCC downgrades occurred when CI scores were updated in April 2021 following the pandemic shock when 10 countries—most of which had CI scores near a cut-off to a lower DCC classification—experienced downgrades (another three downgrades had already occurred in October 2020).[7] With a weight in the CI score (Equation 2) over four times larger

[6] An additional 3 countries experienced temporary changes before reverting to their original DCCs, and one country (Guyana) experienced 3 instances of DCC changes. In one case, an upgrade was due to the inclusion of previously missing data that raised the CI score. In total, nearly one-third of countries, or 21 countries, experienced one or more DCC changes since 2018. This is similar in frequency as under the previous methodology based only the CPIA, when also 21 countries experienced DCC changes over 2011-17 (14 downgrades, 7 upgrades).

[7] CI-scores are updated twice-a-year reflecting forecast revisions in the April and October World Economic Outlooks, in addition to the once-a-year update of countries' CPIA scores. No CI-scores update took place at the time of the April 2020 WEO update, which took place at the height of the pandemic.

than that of country growth, the decline in the 10-year centered average of world growth was the leading contributor to the decline in CI scores (Figure 5).[8] Many of these cases affected borderline countries where relatively small changes in CI scores (around 3 percent on average) resulted in a DCC downgrade. Following these downgrades, further deterioration in country-specific macroeconomic fundamentals mainly kept countries in the lower DCC groups.

$$CI = 0.39*CPIA + 2.72*g + 2.02*\frac{Remittances}{GDP} + 4.05*\frac{Reserves}{Imports} - 3.99*\left(\frac{Reserves}{Imports}\right)^2 + 13.52*g_w \quad (2)$$

CPIA – World Bank's Country Policy and Institutional Assessment (CPIA) score;[9]
g – country's real GDP growth;
g_w – world growth;

21. At present, the DCC does not allow for differentiated impact of global growth on countries. The CI score formula places a significant weight on global growth, reflecting its empirical explanatory power in predicting debt distress episodes (equation 1). At the same time, countries are exposed to a varying degree to global conditions. Trade openness was explored as a potential variable in the 2017 Review but was less empirically significant at the time. Interacting trade openness, or a similar proxy, with global growth could help improve the ability of the DCC mapping to account for the differentiated impact of global conditions among LICs.

22. DCC downgrades on account on changes in country fundamentals have been less frequent and mostly driven by declining external buffers. Declines in reserve coverage have been behind some of the larger changes in CI scores, resulting in DCC downgrades. On the other hand, the only instances of DCC upgrades (three in total, two of which temporary), were led by improvements in reserve coverage.[10]

23. The higher share of countries in the lower DCC groups implies that more countries are assigned CI-scores that exceed their country-specific CI values, potentially overestimating their ability to carry debt. The DCC mapping assigns countries with Weak DCC the upper bound of the range of CI-scores in that category (the 25th percentile as estimated in the 2017 LIC-DSF review), resulting in overestimation of their ability to carry debt relative to the outcome from the Continuous model. Countries with Medium DCC are assigned the median CI-score (as estimated based on data from 2014 in the 2017 LIC-DSF review), with some individual country CI-scores being above and some below it, resulting in underestimation and overestimation of their ability to carry debt, respectively. Countries with Strong DCC are assigned the lower bound of the range of CI-scores in

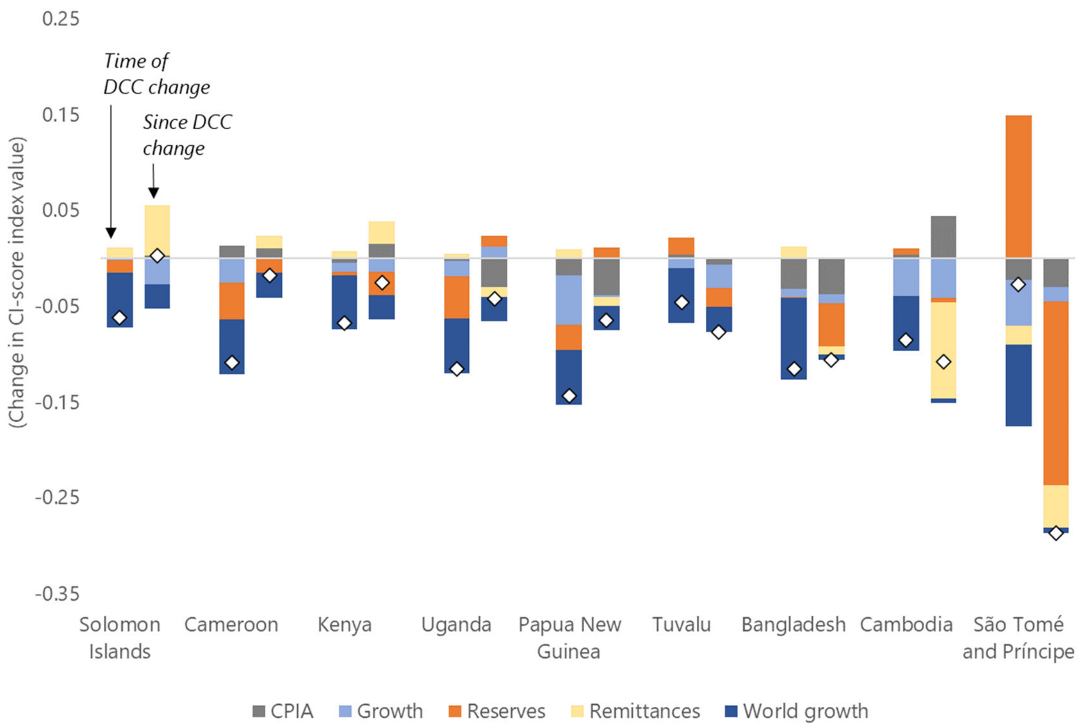
[8] The CI-score equation (Equation 2) consists of the non-debt variables ($\mathbf{X}_k$) included in the underlying PROBIT models (Equation 1). All variables are in percent, except the CPIA score. The weights are the average of the estimated coefficients across the four external PROBIT regressions.

[9] CPIA is an index compiled annually by the World Bank for all IDA-eligible countries, including blend countries. The index consists of 16 indicators grouped into four categories: (1) economic management; (2) structural policies; (3) policies for social inclusion and equity; and (4) public sector management and institutions. Countries are rated on their current status in each of these performance criteria, with scores from 1 (lowest) to 6 (highest).

[10] In other cases, DCC downgrades were led by CPIA (one case) and country growth (one case).

that group as of the 2017 review, resulting in an underestimation of their ability to carry debt. This implied overestimation of countries ability to carry debt has not resulted in a higher overall rate of missed crises in the LIC-DSF implementation.

Figure 5. LIC-DSF Users with DCC Downgrades Led by Changes in Global Growth: Contributions to CI-Score Changes

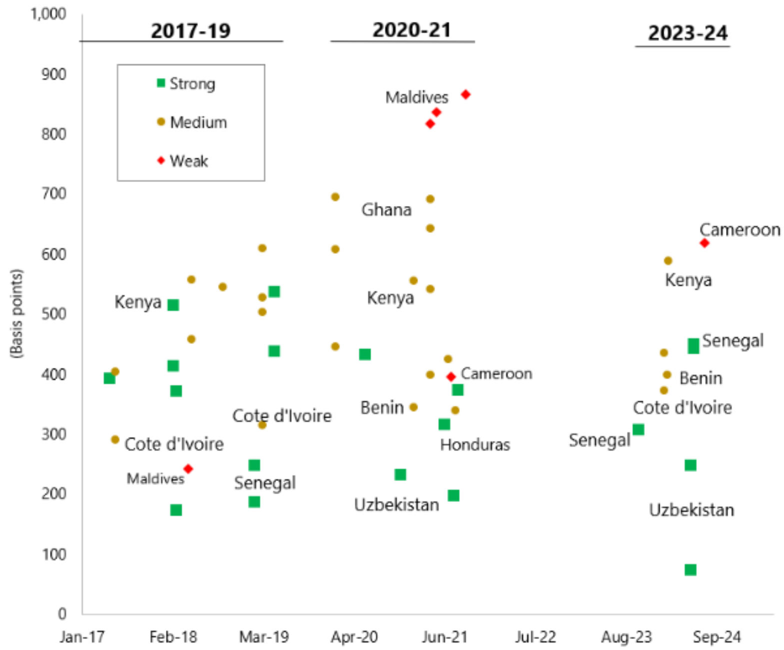


Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Countries shown are those that experienced DCC downgrades due to changes in CI scores in which a change in global growth was the leading contributor. Shown are contributions to CI-score changes that led to the DCC downgrade (first stacked column) and contributions to CI-score changes since the DCC downgrade until April 2024 (none of which resulted in another downgrade for Medium DCC countries).

C. DCC of Frontier LICs

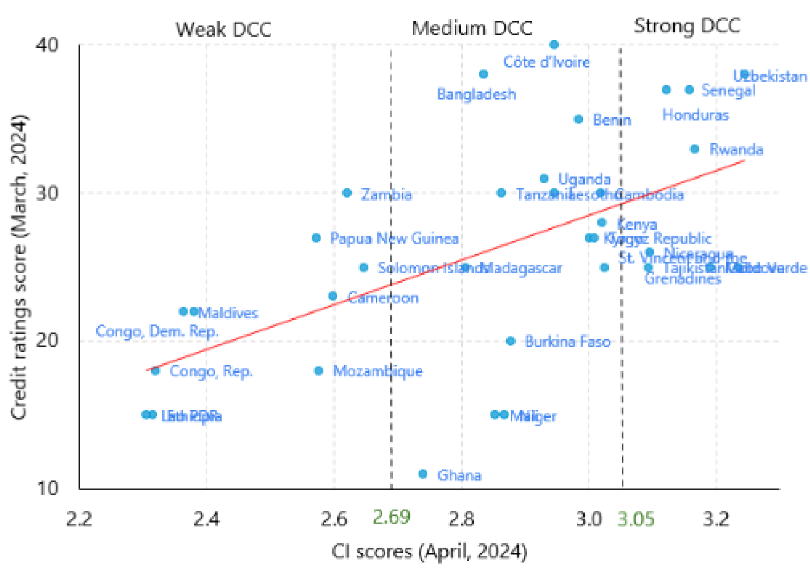
24. Market signals are generally aligned with Frontier LICs DCC classifications. Frontier LICs with stronger DCCs tend to issue at better spreads in international bond markets, including when controlling for maturity and debt levels. Since 2017, spreads for Strong DCC issuers averaged around 350 basis points, compared to around 490 basis points for Medium DCC issuers (Figure 6). Weak DCC issuers, who were fewer in number and generally issue at times of loose global financial conditions, faced higher spreads at around 630 basis points on average. Credit ratings provide similar signals, as countries with higher DCC classifications tend also to have higher credit ratings (Figure 7).

Figure 6. Frontier LICs: Sovereign Bond Spreads by DCC at Issuance, 2017-24



Source: Staff analysis of Bond Radar data and joint IMF-WB LIC-DSF country analyses.
Notes: Issuances as of August 2024. 2017 DCC is as of Oct. 2018. Mozambique, Congo, Rep. of and Togo are not shown due to missing data.

Figure 7. DCC's of LIC-DSF Countries by CI Score and Credit Ratings

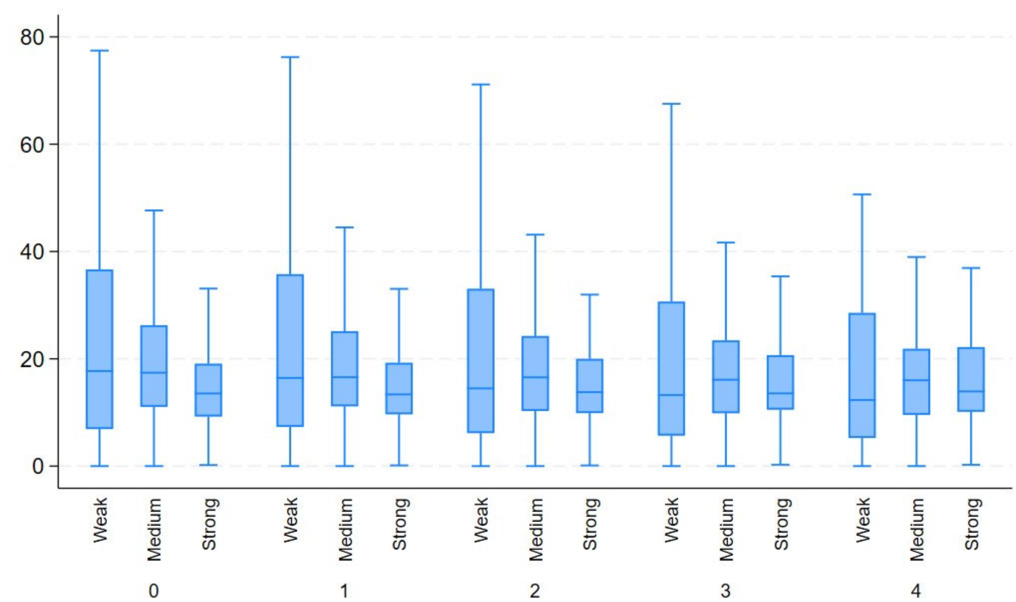


Source: IMF and WB staffs analysis based on LIC-DSF country database and data from Trading Economics.
Notes: AAA credit ratings assigned value 100, with a value of 5 subtracted for each subsequent credit rating.

25. At the same time, the fact that some countries with lower DCC groups issue at similar terms to those of countries with stronger DCC points to the need for further analysis of potentially missing country characteristics. To the extent that additional characteristics relevant for market access are missing from CI scores, capturing them in CI scores could potentially improve the mapping of countries into DCC groups and better reflect their debt repayment capacity, and as a result, improve precision in the risk and sustainability assessments.

26. Economic size and level of domestic financial development are potentially relevant characteristics that could be associated with greater capacity to carry debt. Larger countries tend to benefit from the absolute size of their economies, as intermediation costs and portfolio allocation considerations limit the appeal of smaller denomination bond issuances. Economic size can also serve as a proxy for economic diversification. In addition, higher degree of domestic financial development indicates greater ability for countries to finance themselves without crowding out investment by other sectors or exacerbating macro-financial risks. The absence of a measure of domestic financial development in the DCC also manifests in the fact that more countries with weak DCC have higher ratios of domestic debt to GDP than those of countries in higher DCC groups (Figure 8). For Frontier LICs, domestic market development contributes to the capacity of countries to smooth borrowing needs and finance themselves in adverse external conditions.

Figure 8. LIC DSAs: Domestic Debt-to-GDP Ratio by Debt-Carrying Capacity
(Percent)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Forecast horizon is equal to zero in the year when the DSA is produced (i.e., forecast of the yearly outturn during the same year).

FORECAST ACCURACY

27. Forecast accuracy can be examined by comparing data outturns with their projections made in DSAs prepared in earlier years. The term 'forecast errors' is used to denote the difference between the first actual data reading (e.g., 2022 data as reported in the 2023 DSA) and the forecasts made in earlier DSA vintages (e.g., 2022 forecasts in the 2018, 2019, …, 2022 DSAs) at different forecast horizons.[11] The forecast horizon represents the time gap between each earlier DSA vintage and the year when the actual data outturn was first reported (e.g., the 2022 forecast data made in the 2018 DSA has a forecast horizon of five years, as the 2022 outturn is first reported in the 2023 DSA).

28. The onset of the COVID-19 shock midway through the implementation of the existing LIC-DSF complicates drawing of definitive conclusions about the accuracy of forecasts. Forecasts are intended to capture the central tendency of economic indicators and, as such, are inherently limited in accounting for the impact of exogenous shocks, except through the mechanical extrapolation of their average effects on historical realizations. The impact of the COVID-19 pandemic—a "black-swan" event that occurred midway through the implementation of the current LIC-DSF—skews 'forecast errors', especially at forecast horizons longer than three years, at which differences between data outturns and their earlier forecasts entirely reflect projections made pre-COVID for outcomes that occurred after the shock. Results for shorter forecast horizons also encompass projections and outturns that both occurred before and after the onset of the pandemic. As a result, they are less skewed by the pandemic, though forecasts made in the immediate aftermath can also be noisy as they were made in a period of heightened uncertainty about the transmission channels of the shock to the real economy and the magnitude of any scarring effects.

29. Given the complex chain of events set in motion by the COVID-19 pandemic, the proper interpretation of the results from the analysis of forecast accuracy requires going beyond the headline numbers. For example, analysis of forecast errors of the external public debt-to-GDP ratio would, at first glance, suggest that country teams had been able to foresee the COVID-19 shock. The actual outturns of the PV of external public debt-to-GDP for both the median LIC and a wide range of countries around it are very close to the pre-pandemic forecasts made at different forecast horizons (Figure 9, top panel). In a number of cases, outturns have even fallen short of earlier projections for the nominal external public debt-to-GDP ratio. A more likely explanation of these findings is that they reflect the extended post-pandemic loss of market access and disruptions in the utilization of contracted but not yet disbursed concessional project loans experienced by many LICs. The bigger size of the positive forecast errors for the nominal debt ratio relative to the PV ratio further reflects the stepped up IFI support of the policy response to the pandemic that partially replaced non-concessional financing.

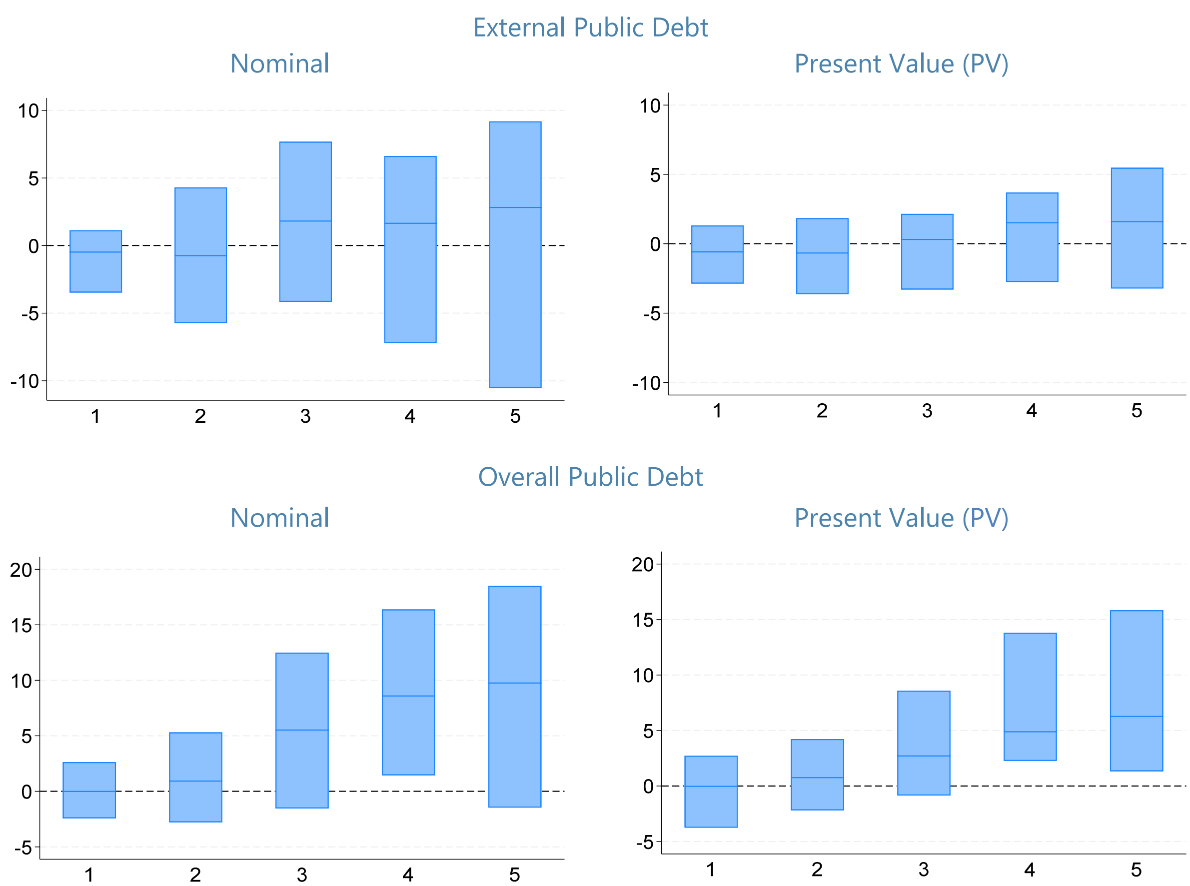
30. The forecast errors for the overall public debt-to-GDP ratio show more clearly the fiscal cost of the COVID-19 shock and subsequent loss of market access that forced LICs to ramp up

[11] As explained in this section, the presence of 'forecast errors' is not necessarily an indication of underlying issues with the quality of forecasts.

domestic borrowing. Results from the analysis show that actual outturns were higher for most LICs in both nominal and PV terms for medium-term forecast horizons (Figure 9, bottom panel). This reflects the fiscal cost of the policy response to the pandemic, a sizeable share of which was financed domestically. In this context, the significant forecast errors of overall debt-to-GDP ratios over the medium term are not a sign of optimism bias. Residuals in the debt accumulation equation, on average, contribute around 0.5 percentage points of GDP per year to the forecast error of the change in overall public debt (Figure 10, lower panel). These positive residuals—indicative of upward 'debt surprises'—reflect recognition of liabilities that extend beyond the DSA's debt perimeter. The relatively small average effect masks substantial contributions in specific country cases.

Figure 9. LIC-DSF Users: Medium-Term Forecast Errors for External and Overall Debt-to-GDP

(Difference between Outturns and Projections, Percentage points of GDP)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses performed under the 2018 LIC-DSF, over 2018-23.
Notes: Forecasts errors measure the difference between outturns and the forecasts of their values made in earlier DSA vintages. Positive values signal the presence of optimism bias, as projections underestimate the debt burden outturns. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported. The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at the given forecast horizon, while the horizontal line indicates the median.

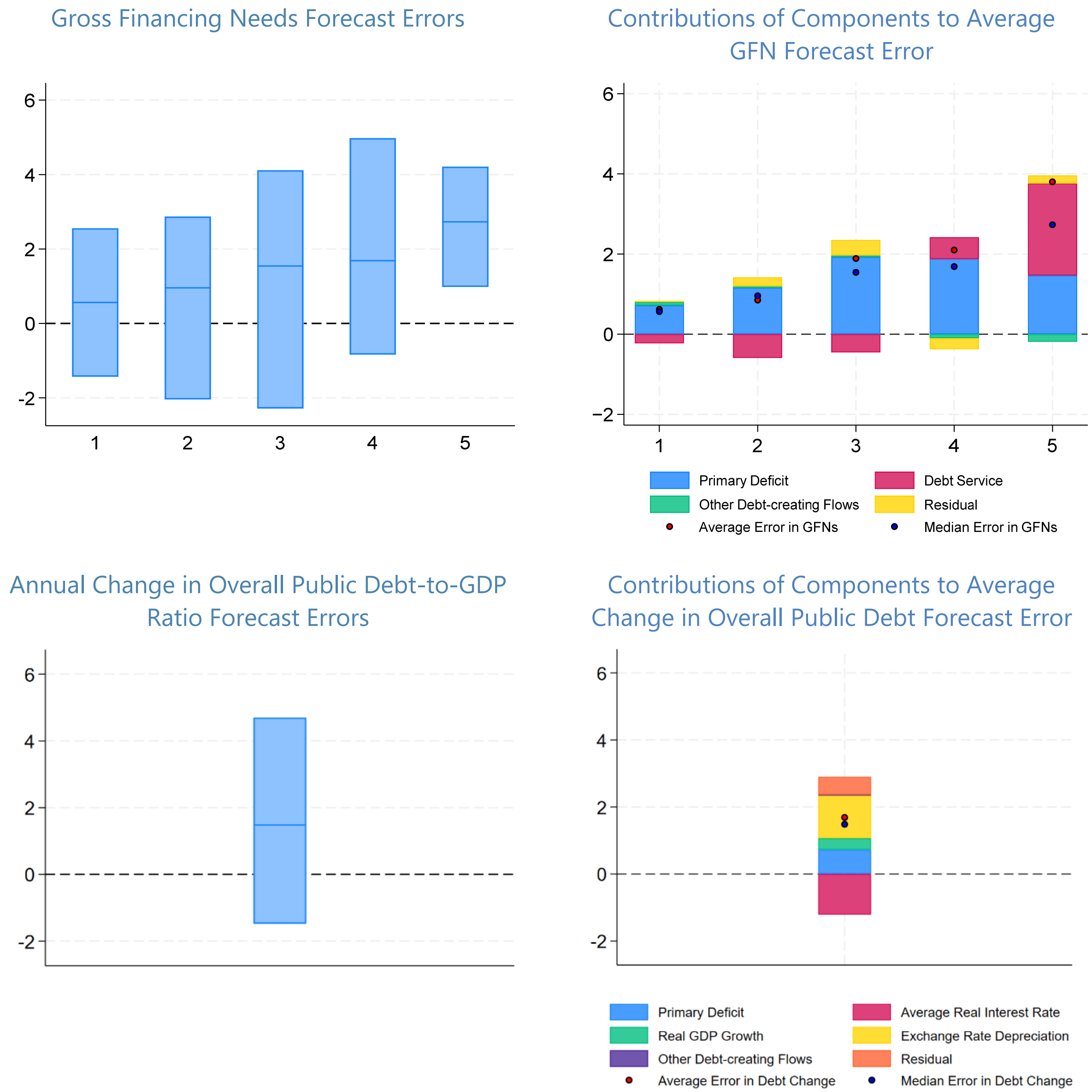
31. Primary fiscal deficits are the largest contributors to forecast errors for gross financing needs. Higher-than-expected primary deficits drove the larger-than-anticipated increases in gross financing needs and the overall public debt (Figure 10). At the same time, many LICs benefited from the Debt Service Suspension Initiative that deferred payments on the external official debt in the first two pandemic years, as seen by the negative contribution of debt service to the average GFN forecast error at shorter time horizons. However, the post-pandemic switch to domestic borrowing at market terms, the tightening of financial conditions, and the spread of deferred payouts under the DSSI over the following years in a PV neutral way resulted in elevated debt service costs that contributed to the positive average forecast errors over the medium term. Residuals play a relatively small role in GFN forecast errors.

32. Despite the noise in the data introduced by the COVID-19 shock, the short-term projections of external and public debt ratios and main macroeconomic indicators show little sign of optimism bias. For the median LIC-DSF country, forecast errors of the public debt-to-GDP ratios, as well as for the four external debt burden indicators remain near or below zero through the 2-year forecast horizon (Figures 9 and 12). This apparent absence of optimism bias also extends to projections of the main macroeconomic variables used in the calculation of external debt burden indicators (Figure 11).

33. There is some evidence that the strengthening of realism tools in the 2017 LIC-DSF review has improved forecast quality, though it is difficult to draw definitive conclusions given the noise in the data brought around by the 2008 GFC and the COVID-19 pandemic. Comparison of the magnitude of forecast errors in the 2018-23 and since the launch of the first LIC-DSF in 2005 shows smaller forecast errors since the 2017 LIC-DSF Review (Figures 11 and 13). This is consistent with the strengthening of realism tools in the latest LIC-DSF. Despite these advances, further improvements to data collection and reporting practices, as well as potential refinements to realism tools, remain critical to ensuring continued progress in strengthening forecast accuracy.

Figure 10. LIC-DSF Users: Analysis of Forecast Errors of Gross Financing Needs and Change in Overall Public Debt-to-GDP Ratio

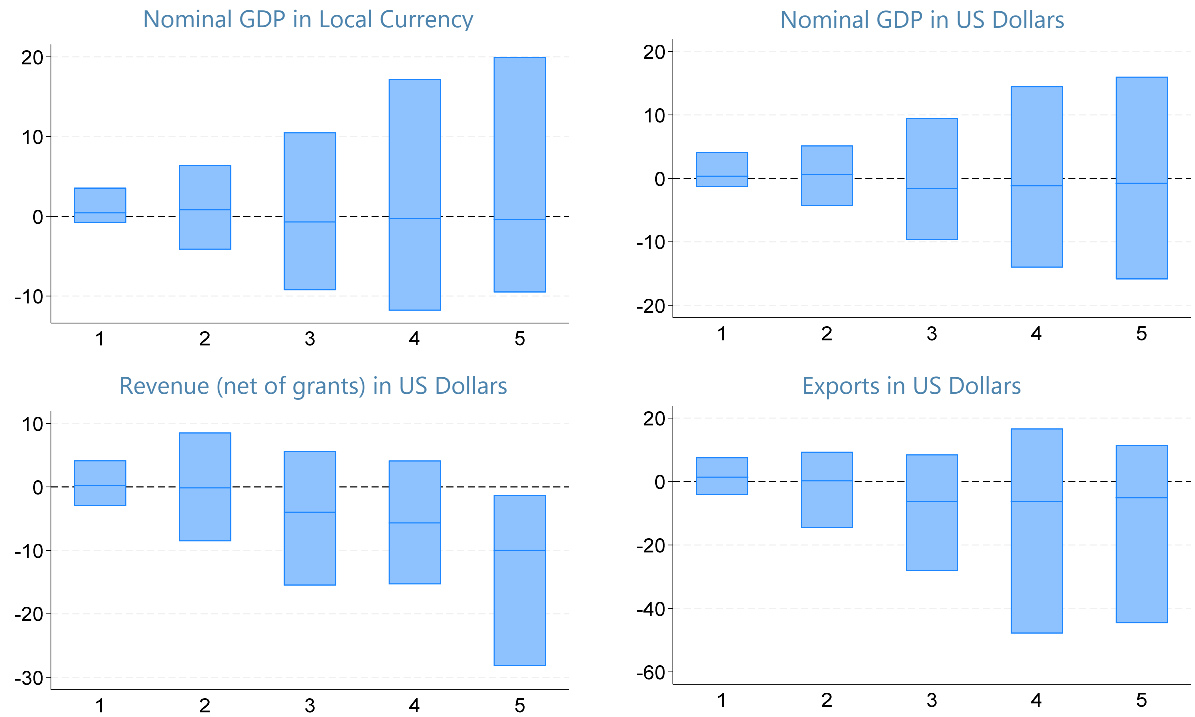
(Difference between Outturns and Projections, in Percentage Points)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses performed under the 2018 LIC-DSF, over 2018-23.
Notes: The top left panel shows the difference between the near-final (in-year) GFN projections, and their forecasts made in the four preceding years. The top right panel presents the median and average GFN forecast errors, along with the contributions of its components to the average GFN forecast error among LICs. The bottom left panel shows the difference between outturns and the forecasts made in earlier DSA vintages of the annual change in the overall public debt-to-GDP ratio over 2018-23 at all forecast horizons combined. The bottom right panel presents the median and average forecast errors for the annual change in overall public debt-to-GDP ratio, along with the contributions of its components to the average forecast error among LICs at all forecast horizons combined. Positive values signal the presence of Optimism Bias, as projections underestimate the outturns for GFNs and the annual change in overall public debt. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

Figure 11. LIC-DSF Users: Medium-Term Forecast Errors for Macroeconomic Indicators Entering Debt Burden Indicator Ratios

(Deviation of Outturn from Projections, Percent)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses performed under the 2018 LIC-DSF, over 2018-23.

Notes: Forecasts errors in this Figure measure the percentage deviation of outturns from the forecasts of their values made in earlier DSA vintages, expressed as a percentage of outturns. Negative values signal the presence of Optimism Bias, as the projections overestimate the outturn of variables used as denominators of debt burden indicators. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported. The lower and upper hinges of each box correspond to the 25th and 75th percentile of the distribution of the indicator at the given forecast horizon, while the horizontal line indicates the median.

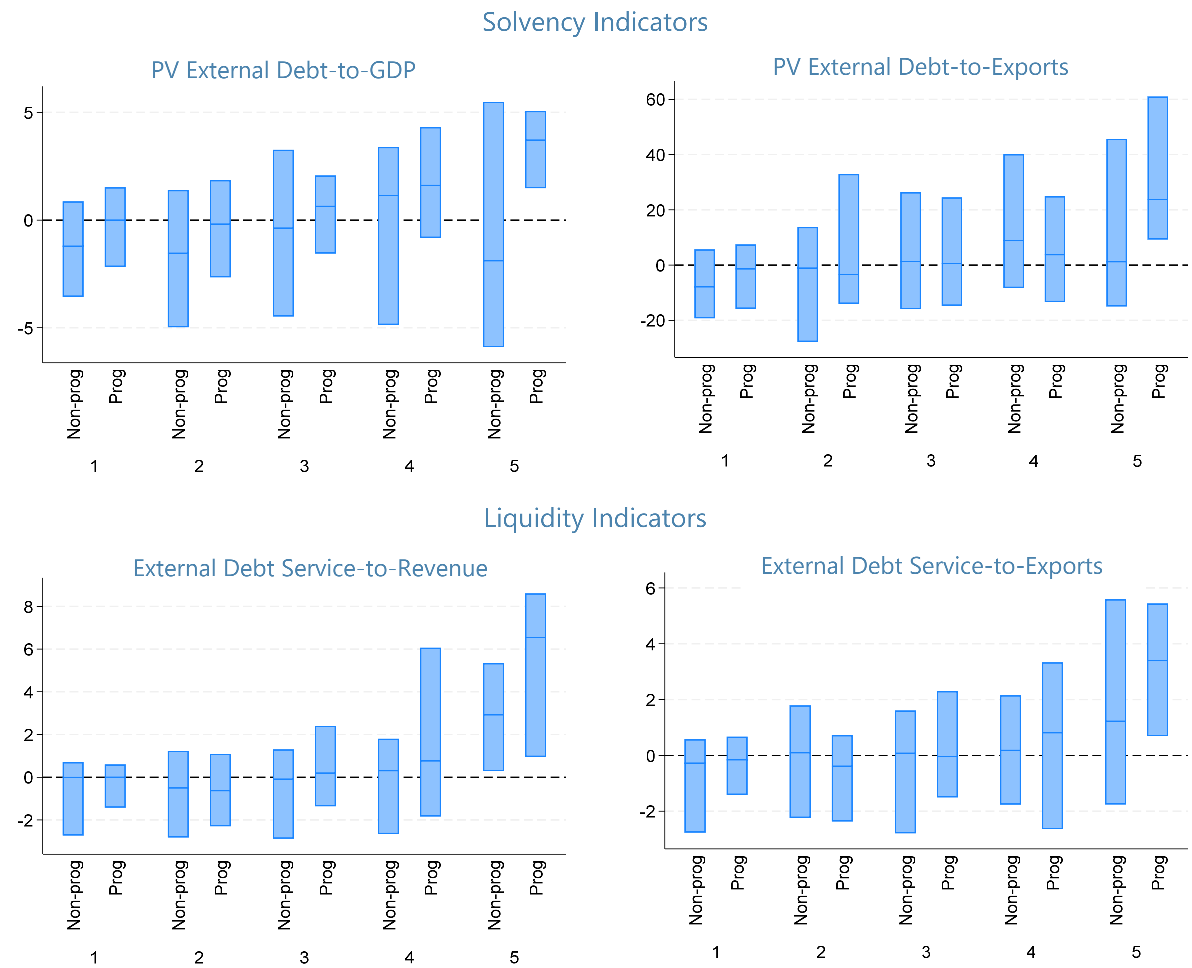
34. There is some evidence that forecast errors for external debt burden indicators over the medium term and revenues and exports over the long run are higher in countries with Fund program arrangements. The forecast errors of the external debt burden indicators for the median LIC-DSF country are higher for program than non-program countries in the medium term (Figure 12). Disentangling the drivers of this finding is complicated by the fact that countries that request a Fund program are typically subject to greater unanticipated shocks than their peers and their baseline forecasts assume successful implementation of difficult reforms, which has a higher probability of not materializing than the most likely policies reflected in the baseline for non-program countries.[12] Forecasts of other macroeconomic indicators appear to play a role especially over forecast horizons beyond the 3-4 years mark, as illustrated by the bigger negative differences

[12] Dreher (2006), for instance, argues that IMF programs often assume successful policy implementation, resulting in forecasts that may not fully account for the complexities and challenges of execution. Furthermore, Ho and Mauro (2014) find greater optimism bias in long-term forecasts for IMF program countries.

between outturns and earlier forecasts for fiscal revenues (excluding grants) and exports in US dollars (Figure 13). The differences become even more pronounced in the long run. At the same time, the patterns of forecast errors are similar across program and non-program countries for GDP in US dollars, with evidence of overestimation in projections. Nominal GDP in local currency is, however, consistently under-projected over forecast horizons greater than five years, reflecting in part frequent revisions of historical data.

Figure 12. LIC-DSF Users: Medium-Term Forecast Errors for External Debt Burden Indicators by Program Status

(Difference between Outturns and Projections, Percentage points)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses performed under the 2018 LIC-DSF, over 2018-23.

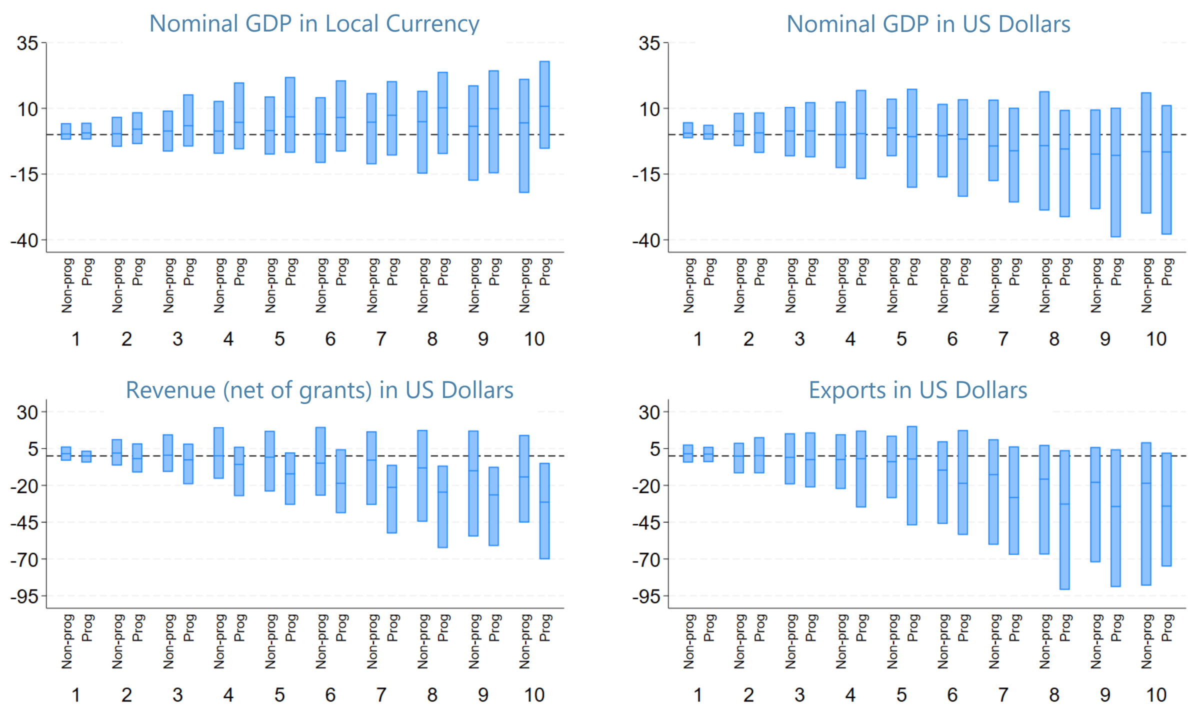
Notes: Forecasts errors measure the difference between outturns and the forecasts of their values made in earlier DSA vintages. Positive values signal the presence of Optimism Bias, as projections underestimate the debt burden outturns. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

35. There is some evidence that forecast errors for revenues and exports are more pronounced in commodity exporters in the long run. The forecast errors of these variables for the median LIC-DSF country are higher for commodity exporters than other LICs beyond a forecast horizon of five years (Figure 14).[13]

36. The relatively good track record in forecasting the nominal GDP in local currency and US dollars for the median LIC masks an underlying optimism bias in forecasts of real GDP growth (Figure 15). In a number of cases, this is more than offset by under-projection of the GDP deflator, which forecast errors have the opposite sign and are of larger magnitude than those of real GDP growth. In other cases, frequent revisions of historical data help offset the impact on the level of GDP.

Figure 13. LIC-DSF Users: Long-Term Forecast Errors for Macroeconomic Indicators Entering Debt Burden Ratios by Program Status

(Deviation of Outturn from Projections, Percent)



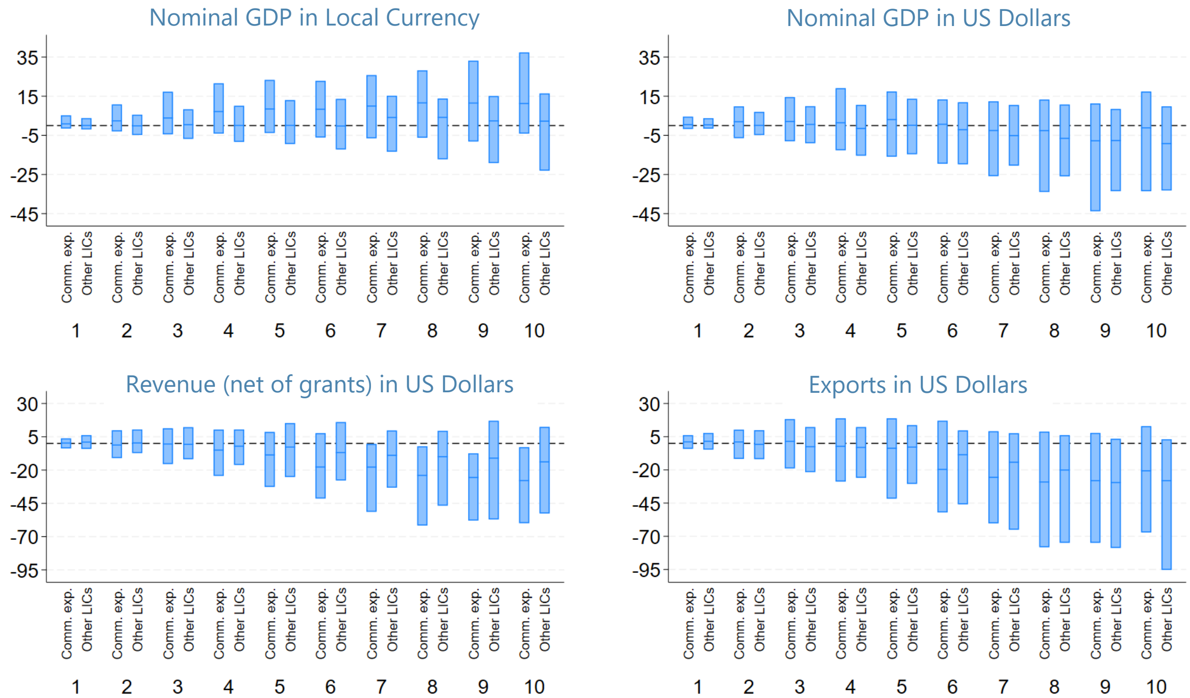
Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

Notes: Forecasts errors in this chart measure the percentage deviation of outturns from the forecasts of their values made in earlier DSA vintages. Negative values signal the presence of Optimism Bias, as the projections overestimate the outturn of variables that enter as denominators of the debt indicators. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

[13] The analysis of forecast errors for the four external debt burden indicators does not reveal any discernible differences in patterns between commodity exporters to other LICs in the short to medium term.

Figure 14. LIC-DSF Users: Long-Term Forecast Errors for Macroeconomic Indicators Entering Debt Burden Ratios for Commodity Exporters and Other LICs

(Deviation of Outturn from Projections, Percent)

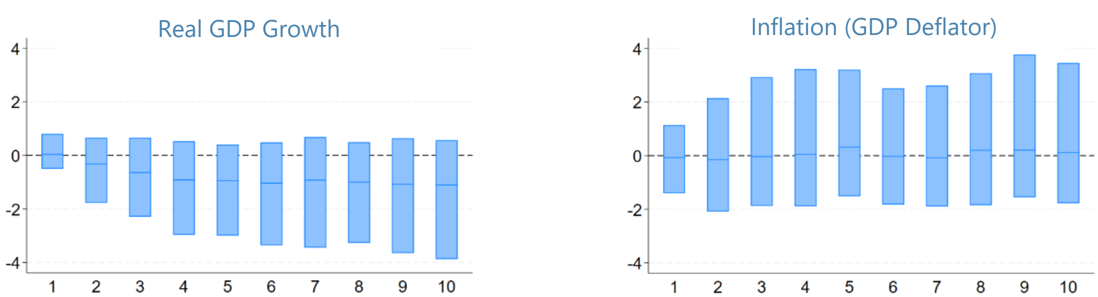


Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.

Notes: Forecasts errors in this chart measure the percentage deviation of outturns from the forecasts of their values made in earlier DSA vintages. Negative values signal the presence of Optimism Bias, as the projections overestimate the outturn of variables that enter as denominators of the debt indicators. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

Figure 15. LIC-DSF Users: Long-Term Forecast Errors for Real GDP Growth and Inflation

(Difference between Outturns and Projections, Percentage points)



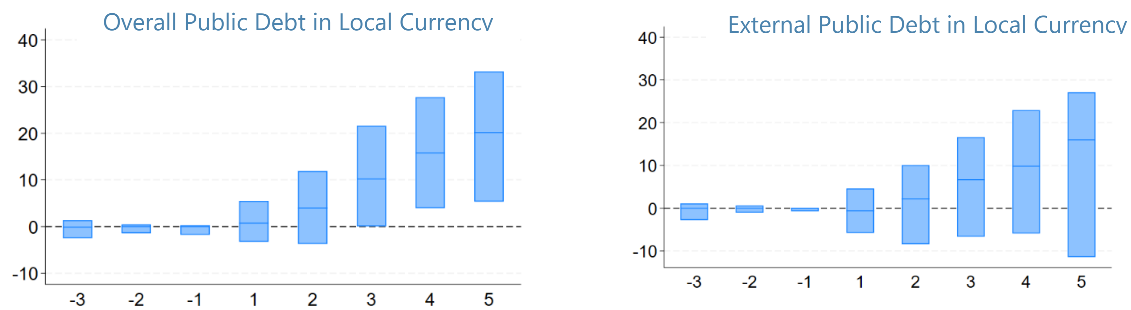
Source: Staff analysis of joint IMF-WB LIC-DSF country analyses, over 2006-23.
Notes: Forecasts errors in this Figure measure the difference between outturns and the forecasts of their values made in earlier DSA vintages. For real GDP growth, negative values signal the presence of Optimism Bias, as the projections overestimate the outturn. For inflation, positive values signal Optimism Bias, as the projections underestimate the outturn. The X-axes show the time gap between each earlier DSA vintage and the year when the actual data was first reported.

37. Differences between data outturns and their earlier forecasts can also be due to revisions of historical data series made after a forecast has been made. Analysis of the direction of updates to projections across the whole forecast horizon between DSA vintages separated from one another by different time spans over 2006-23 (i.e., forecast revisions) shows that nominal GDP levels, along with overall and external public debt (over 2018-23), have consistently been revised upwards (see Figure 16). In the case of nominal GDP in levels, this is in part driven by statistical methodology improvements, such as GDP rebasing and expansion of coverage that result in higher nominal levels of GDP that affect the level of the entire time series. This is seen from the negative sign of historical revisions, which indicate that the first reported value of GDP has been subsequently revised up (Figure 16).[14]

Figure 16. LIC-DSF Users: Historical Revisions and Forecast Errors

(Percent)

Medium Term, 2018-23

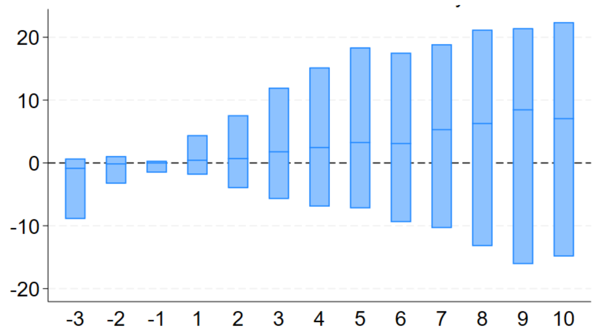


[14] Historical revisions are defined as the difference between the earliest actual data reading (e.g., 2022 data as reported in the 2023 DSA) and any revisions to the 2022 outturn as reported in subsequent DSAs (2024, 2025, etc.). The historical revision horizon represents the time gap between each subsequent DSA vintage and the year when the actual data outturn was first reported (e.g., revisions to the 2022 data outturn made in the 2025 DSA are made at revision horizon two, as the 2022 outturn is first reported in the 2023 DSA).

Figure 16. LIC-DSF Users: Historical Revisions and Forecast Errors (concluded)

Long Term, 2006-23

Nominal GDP in Local Currency



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses, over 2018-23 (medium term) 2006-23 (long-term).
Notes: This Figure shows the difference between the first actual data reading for a given year and: (i) for historical revisions, the actual data readings for that year as reported in subsequent DSAs; and (ii) for forecast errors, the forecasts of their values made in earlier DSA vintages; both differences are expressed as a percentage of the first actual data reading. Negative values signal that the first actual data reading for a given year was lower than either its subsequent revisions or earlier forecasts. The X-axes show the time gap between the year when the actual data outturn was first reported and each DSA vintage; this difference is negative for historical revisions (shown at -1, -2, -3) as they occur in subsequent DSA vintages, and positive for forecast errors (shown at 1, 2, ..., 10) as they occur in earlier DSA vintages.

DEBT PERIMETER ISSUES

38. **The LIC-DSF targets near-complete coverage of the public sector to capture relevant risks and promote comparability of the assessments across countries.** The existing framework targets a near-complete coverage of both external and domestic, public, and publicly guaranteed debt. The institutional coverage includes both the general government (or its individual components, such as central, state, and local governments, social security funds and extra budgetary funds, if consolidated data is not available), the central bank,[15] and all state-owned non-financial enterprises (or SOEs). Under clearly defined circumstances, specific SOEs may be excluded, if the enterprise poses limited fiscal risks. A narrower definition of public debt can contribute to unexpected increases from materialization of fiscal risks from sources outside the defined perimeter. The quality of the LIC-DSF model calibration and DSA risk assessments is also affected by cross-country variations in debt coverage, including due to data reporting.

A. Debt Coverage in Model Dataset

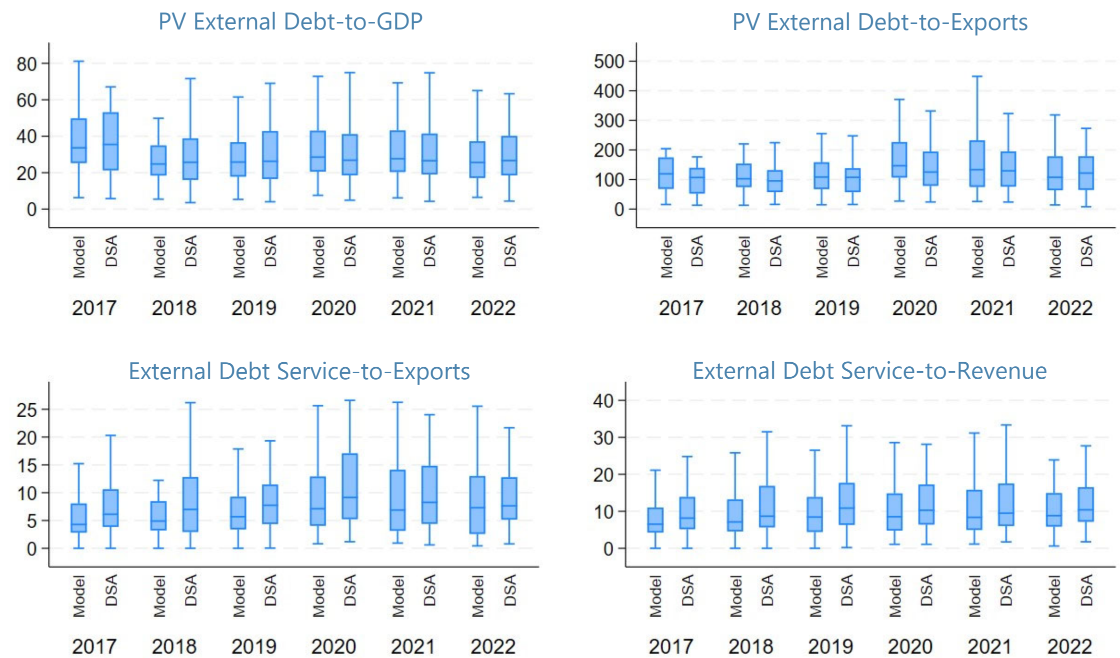
39. **Differences in debt coverage between the data used in the model estimation and the DSAs themselves do not appear systematic.** A comparison of the distributions of the values of the external debt burden indicators in the model database updated through 2022 and the data used in DSAs shows that interquartile ranges are broadly aligned for solvency risk metrics, but generally

[15] For the central bank, only debt contracted on behalf of the government or taken for purposes other than to support the implementation of monetary policies or reserve management is counted as public debt.

somewhat higher for liquidity metrics though gaps have narrowed in latest years (Figure 17). Differences in definitions and methodological approaches can explain the discrepancies between the external debt data used in the model, which is based on the World Bank's International Debt Statistics, and the one used in practice in the DSAs. This includes different criteria for inclusion of SOE data; different assumptions in projecting cross-currency exchange rates and variable interest rates; use of residency versus currency criteria to define external debt; the reporting of debt service based on contracted amounts in the WB IDS versus actual payments in the DSAs; etc.

Figure 17. Comparison of Distribution of External Debt Burden Indicators in Model and DSA Datasets, 2017-22

(Percent)



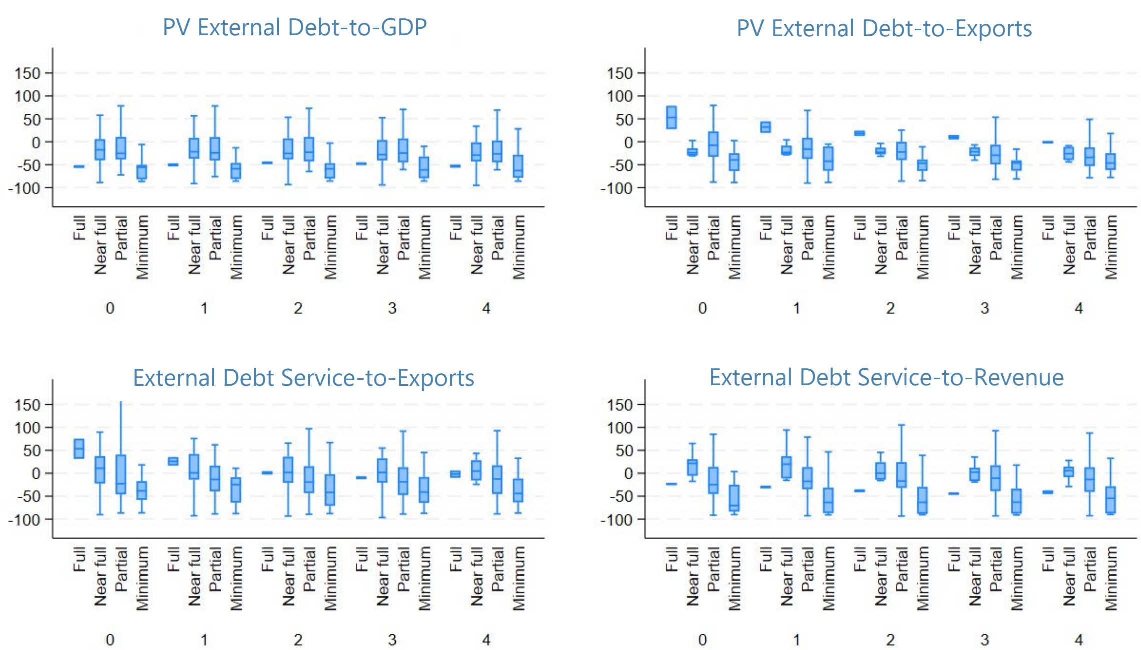
Sources: Staff calculations drawing on different debt databases and DSAs.
Notes: Model database refers to the revised and extended data series used in the current review. DSA database is constructed from data used in DSAs.

B. Debt Coverage in DSAs

40. Incomplete data coverage can lead to biases in the risk assessments. This issue arises from the fact that countries are judged against uniform thresholds, irrespective of the breadth of actual debt coverage. Analysis of the magnitude of the average deviations of debt indicators from their thresholds over the medium-term suggests that countries with less comprehensive debt coverage tend to exhibit smaller threshold breaches in High Risk cases and are farther from the thresholds from below in Moderate Risk ones with corresponding effect on fiscal space assessments, potentially affecting the quality of some risk signals (Figures 18 and 19).[16] This implies a relatively weaker risk signal for countries with less comprehensive coverage relative to a similar country with better coverage, and may miss an important element of fiscal risk.

41. Including estimated contingent liabilities in the debt perimeter could impact both the overall and external risk ratings of countries with incomplete data coverage. If the estimated contingent liabilities (CLs) are brought into the DSA baseline for Low and Moderate external risk cases with no breaches in the baseline, this will generate a mechanical High risk signal in about one-quarter of them, indicating at least one breach in the baseline. Among those with CLs above the median value across LIC-DSF users, about half would face such mechanical risk signals.[17] The CL stress test offers a valuable channel to incorporate estimates of fiscal exposures, yet incorporation of liabilities intended for the baseline (due to data constraints, etc.) may materially result in more optimistic risk assessments.

Figure 18. High Risk LIC DSAs: Deviations of External Debt Indicators from Thresholds by Debt Coverage
(Percent)



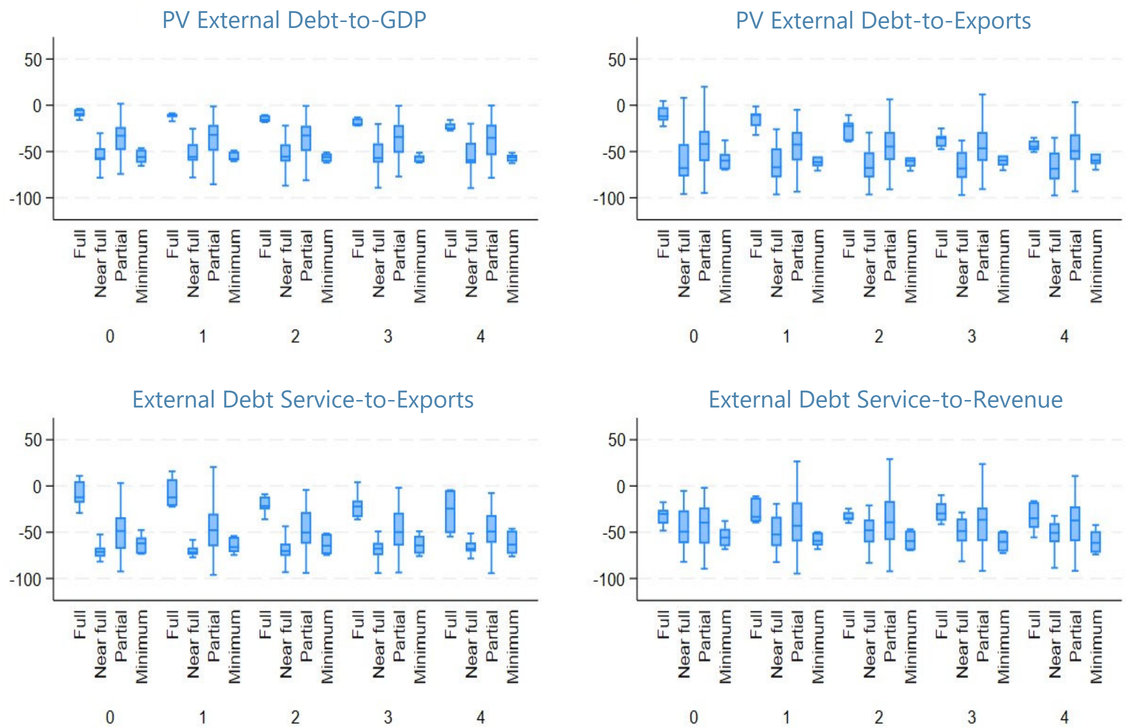
Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Positive deviations signal breaches of thresholds. Forecast horizon is equal to zero in the year when the DSA is produced (i.e., forecast of the yearly outturn during the same year). Debt coverage groups: Full (general government (defined as covering central, state and local governments and at least one of Social Security Fund and Extrabudgetary Funds) and guarantees and non-guaranteed SOE debt); Near full (narrower budgetary government, guarantees and non-guaranteed SOE debt); Partial (either general or narrower budgetary government and guarantees); Minimum (either general or narrower budgetary government).

[16] Analysis is carried out only for debt indicators, for which there is a sufficient number of countries in different debt coverage groups in the High and Moderate Risk groups.

[17] Results are similar when looking at overall risk rating.

Figure 19. Moderate Risk LIC DSAs: Deviations of External Debt Indicators from Thresholds by Debt Coverage

(Percent)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Notes: Positive deviations signal breaches of thresholds. Forecast horizon is equal to zero in the year when the DSA is produced (i.e., forecast of the yearly outturn during the same year). Debt coverage groups: Full (general government (defined as covering central, state and local governments and at least one of Social Security Fund and Extrabudgetary Funds) and guarantees and non-guaranteed SOE debt); Near full (narrower budgetary government, guarantees and non-guaranteed SOE debt); Partial (either general or narrower budgetary government and guarantees); Minimum (either general or narrower budgetary government).

42. **Incomplete coverage often reflects data constraints in capturing SOE debt fully and consistently.** There is a mixed practice of including all or a subset of larger SOEs in DSA baselines. Differences can often be attributed to i) whether liabilities of different institutional sectors are consolidated; ii) whether SOEs are part of the definition of government used in the DSA; and iii) data availability between external and domestic debt and/or among debt stocks, borrowing plans, and revenue projections. SOEs often have large crossholdings of public sector liabilities (SOEs hold government securities or accumulate claims against each other), which should be consolidated to avoid double counting public sector liabilities. Projecting SOE revenues, deficit, and debt service requires insight into sectoral trends and may also not be consistently available across SOEs. While

data on external borrowing of SOEs is typically available (through World Bank International Debt Statistics (IDS)), data on domestic SOE debt is not readily available.

43. **SOEs remain a source of heightened fiscal risks in LICs.** While most SOEs do not borrow externally, there are a number of cases in which large external debts on SOE books have been transferred to the government (Box 1).

Box 1. Examples of Materialization of Fiscal Risks from SOEs

Kenya. Kenya Airways (KQ) faced financial difficulties starting in 2014. By 2022, KQ had a negative equity position and the highest cost base among all airlines in the Sub-Saharan Africa, even after a 75 percent cost reduction. The airlines' total debt reached US$835 million and was insolvent by end-2022. Of this amount, US$638 million was the remaining balance (out of original US$750 million) of debt guaranteed by the Government of Kenya (GoK). Even after repeated extraordinary budget support and moratorium on debt service for the amounts owed to the GoK, KQ was unable to service its external debt. Following a call on the guarantees, GoK begun servicing KQ's guaranteed senior external debt of US$95 million between October 2022 and April 2023. In addition to external guarantees, the GoK had extended domestic guarantees of FX-denominated loans of US$225 million to a consortium of local banks with a 10-year maturity to KQ in 2017. Although the debt is owed to domestic banks, it is included in the external debt stock as DSA is prepared on currency basis. In 2024, local banks called on the government guarantee for an amount of US$150 million on KQ's FX-denominated obligations, creating additional fiscal financing needs.

Mozambique. In 2013-14, three SOEs, Proindicus, MAM, and Ematum borrowed about US$2 billion (over 13 percent of GDP) through government guaranteed loans for the purchase of vessels and equipment to develop the tuna fishing industry and improve maritime security. The government guarantee was issued without proper legal procedure and breached legal limits. Two external commercial creditors, Credit Swiss and VTB Capital, provided the loans on the international market, channeled through a UAE-based shipbuilding company, Privinvest. An independent audit conducted by a third party subsequently revealed that US$ 500 million was diverted from the investment projects. Government officials and bankers involved in arranging the deals were accused of misrepresenting the use of proceeds, the amount and maturity structure of Mozambique's public debt, and the country's capacity to repay to the creditors. As a result, Mozambique stock of public and publicly guarantee debt, including domestic debt, jumped from 86 percent of GDP in 2015 to 125 percent of GDP in 2016, including 8 percent of GDP from the "hidden debt" and significant exchange rate depreciation which exacerbated the external debt burden. The stock of arrears on public and publicly guaranteed external debt reached about US$1.2 billion.

Vanuatu. Over recent years, Air Vanuatu, the national airline, has been experiencing serious operational and financial difficulties. In March 2021, the government took temporary control of Air Vanuatu in order to restructure the enterprise. Following the placement of Air Vanuatu under voluntary liquidation in May 2024, a significant portion of the debt owed by Air Vanuatu to creditors has been included in the DSA debt perimeter, given the high likelihood that the government will take over Air Vanuatu's liabilities.[1]

Sources:
IMF, 2023, "Kenya: 2023 Article IV Consultation-Sixth Reviews under the Extended Fund Facility and Extended Credit Facility Arrangements, Country Report No. 2024/013, January, (Washington: IMF).
IMF, 2024a, Mozambique: 2024 Article IV Consultation, Fourth Review under the Three-Year Arrangement under the Extended Credit Facility", Country Report No. 2024/219, July, (Washington: IMF).
IMF, 2024b, Vanuatu: 2024 Article IV Consultation," Country Report No. 2024/278, September, (Washington: IMF).

[1] In September 2024, the restructuring of Air Vanuatu was concluded with reaching a creditor compromise which resulted in a reduction of headcount and debt liability for the airline.

44. Differences among currency and residency criteria are mostly material for countries in regional currency unions. When data is available, DSAs should be recorded on a creditor residency basis, as it aligns better with balance of payments flows, provides consistency between DSA and the macroframework, captures the risks of sudden capital flight, provides a more accurate picture of rollover risks related to the sovereign-bank nexus. In practice, the difference between residency and currency-based debt definitions of external debt is generally assessed by country teams to be immaterial. However, in 18 percent of DSAs with currency-based definition, the difference is assessed as being material.[18] In the case of countries in regional currency unions (e.g., WAEMU), where the majority of the domestic currency debt is held by regional (non-domestic) investors, switching debt definition would have measurable impact and could result in changes to risk ratings. This points to the need for greater focus on overall public debt vulnerabilities in countries members of a currency union.

45. A narrow debt perimeter and gaps in debt definition may result in large debt flow residuals and revisions to debt stocks that can contribute to perceived optimism bias in projections. Unidentified, or residual, debt-creating flows signal that debt is being accumulated outside the primary balance and the automatic debt dynamics captured in the model.[19] The average residuals in historical external and overall public debt accumulation are positive across LIC-DSF users and amount to 0.6 percent and 1.8 percent of GDP per year, respectively (Figure 20). Upward revisions to historical debt stock figures across DSA vintages capture the recognition of additional liabilities retrospectively, either due to data omissions or expansion of sectoral coverage. Upward revisions to external and overall public debt stocks amounted to 0.6 percent of GDP and 1.3 percent of GDP per year on average, respectively (Figure 20). The upward revisions were most notable for countries that have recently completed or are undergoing a restructuring process (Zambia, Malawi), likely reflecting additional debt reconciliation efforts.[20] At the same time, the forecast errors for the residual in the debt accumulation equation for overall public debt are smaller (Figure 10), reflecting the fact that while in the majority of cases this residual is set to zero in forecasts, some teams do factor in historical trends and project it, which would mitigate the impact on optimism bias.

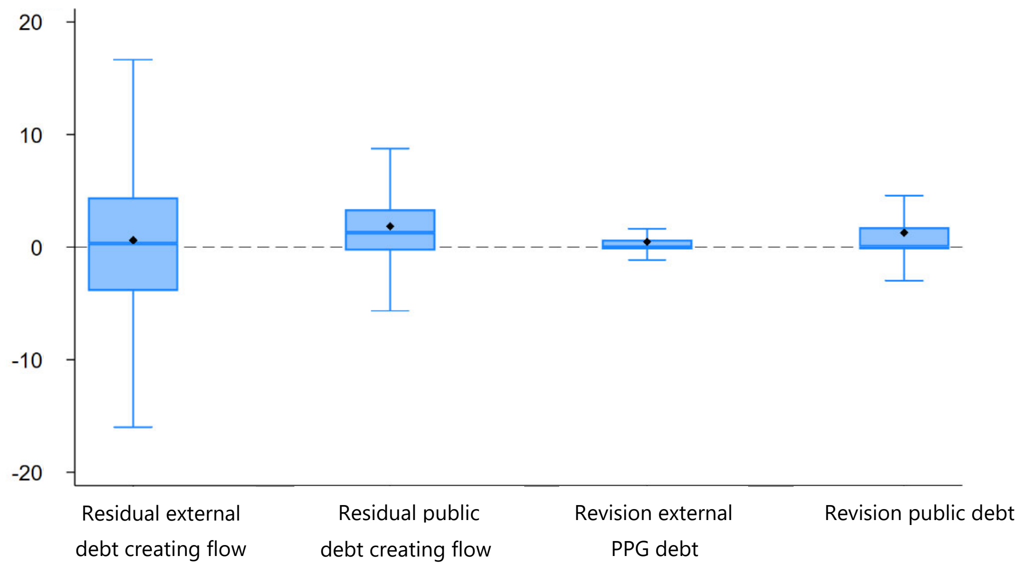
[18] There have been some improvements with 4 additional countries switching to residency-based definition over time and one country reversing back to currency-based definition.

[19] Unidentified debt-creating flows are also influenced by underlying data quality.

[20] These results align with long-term review of debt revisions in Horn et al., 2024, "*Hidden Debt Revelations," Policy Research Working Paper, 10907, (Washington: World Bank).*

Figure 20. Contributions to Annual Debt Stock Changes of Residuals in Debt Accumulating Equation and Data Revisions, 2013-23

(Percent of GDP)



Source: Staff analysis of joint IMF-WB LIC-DSF country analyses.
Note: Averages shown as diamond markers.