INTERNATIONAL MONETARY FUND

IMF POLICY PAPER

REVIEW OF THE DEBT SUSTAINABILITY FRAMEWORK FOR MARKET ACCESS COUNTRIES

JANUARY 2021

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

- A **Press Release** summarizing the views of the Executive Board as expressed during its January 14, 2021 consideration of the staff report.
- The **Staff Report**, prepared by IMF staff and completed on November 25, 2020 for the Executive Board's consideration on January 14, 2021

[The documents listed below have been or will be separately released.]

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.

PRESS RELEASE

PR21/31

IMF Executive Board Reviews IMF Debt Sustainability Framework for Market Access Countries

FOR IMMEDIATE RELEASE

Washington, DC – February 3, 2021: The Executive Board of the International Monetary Fund (IMF) reviewed on January 14, 2021 the IMF Debt sustainability Framework for Market Access Countries (MAC DSA). The review revealed scope to improve the MAC DSA framework's ability to identify risk of sovereign stress and better align it with the IMF's lending framework, to be achieved by replacing the current approach with a new methodology.

The MAC DSA plays a key role in the Fund's core functions of surveillance and lending. In surveillance, the framework helps identify a member's vulnerability to sovereign stress to steer the member away from such stress. In Fund-supported programs, which often take place after the stress has already developed, the DSA helps determine if sovereign stress can be resolved via a combination of IMF financing and economic reforms, or if measures such as debt restructuring are needed to deliver medium-term debt sustainability. The framework is also used in developing IMF conditionality and informing the need for debt relief in debt restructuring operations undertaken in the context of Fund-supported programs.

Since its introduction in 2002, this framework has been reviewed in 2003, 2005, and 2011–13. The 2011–13 review introduced key features, including a risk-based approach through distinction between high and low scrutiny countries, standardization of writeup and publication requirements, realism tools to guard against optimistic economic projections, a heatmap summarizing debt vulnerabilities, and debt fancharts to give a sense of the uncertainty around the projected path of the debt/GDP ratio.

A careful review over the past two and a half years has revealed scope for further improvements, so as to predict sovereign stress with greater accuracy. The new framework includes a broader and more consistent debt coverage, a longer projection horizon, new tools at multiple horizons based on superior analytical methods that account for countries' structural characteristics, and enhanced transparency in the bottom-line assessments, including the exercise of judgment. Furthermore, the new tools support probabilistic debt sustainability assessments, as required by the Fund's lending framework.

The framework is expected to be operationalized in the final quarter of 2021/first quarter of 2022. This will be preceded by the completion of the accompanying Guidance Note and template, and extensive engagement with country authorities and other external stakeholders. The transition between the old and the new framework will be carefully managed to ensure consistency.

700 19th Street NW
Washington, DC 20431
USA

Executive Board Assessment[1]

Executive Directors welcomed the wide-ranging and comprehensive review of the Debt Sustainability Framework for Market Access Countries (MAC DSA), to be renamed "Sovereign Risk and Debt Sustainability Framework for Market Access Countries" (MAC SRDSF) to capture the full range of its analysis. Against the backdrop of rising vulnerabilities related to the pandemic, they broadly supported the proposed reforms aimed at improving the framework's capacity to predict sovereign stress, enhancing transparency and communication of its results, and aligning it with the three-zone sustainability assessment required under the exceptional access framework. Directors recognized that the framework would require some further technical fine-tuning in the run up to the preparation of the Staff Guidance Note and implementation.

Directors supported the continued application of the existing definition of debt sustainability, and most concurred that General Government (GG) debt, defined per GFSM 2014 classification, should be the default institutional coverage. A few Directors suggested that the expansion of debt coverage to GG be implemented in a phased manner, as two-fifths of EMs currently report data for the central government only. Directors welcomed the incorporation of public sector liquid financial assets as a mitigating factor, and most Directors supported the risk-based approach under which central bank liabilities and/or SOE contingent liabilities would need to be included in the debt perimeter. However, a few Directors advised the incorporation of a broader range of public sector assets and wider adoption of net public debt concepts in the framework. Directors stressed that capacity-development support would be needed to bring country data coverage to adequate levels. A few Directors preferred the continuation of the existing 5-year time horizon in certain cases in view of large uncertainties regarding public debt projections.

Directors welcomed the expanded realism toolkit for baseline projections and tools to assess sovereign risks at three horizons: short, medium, and long term. They supported the use of the proposed new tools, with slight adjustments, to produce the probabilistic debt sustainability assessments required in Fund-supported programs and evaluate the consistency of restructuring targets with restoring sustainability in debt restructuring cases. A number of Directors emphasized the need to adequately account for the impact of climate change on sovereign risk and debt sustainability. A few Directors questioned the expansion of the existing realism toolkit to cover exchange rate analysis, especially for pegged regimes. A number of Directors expressed concern about the use of perceptions-based third-party indicators to build the institutional quality variable used in the short- and medium-term models. In addition, these Directors asked to leave adequate room for judgment and, as a cross-check, compare results using alternative indicators of institutional quality that are not perceptions-based.

Directors agreed that a sovereign risk analysis should generally be prepared in both program and surveillance contexts. In a program context, staff reports should contain the full range of risk-of-sovereign-stress outputs for the medium and long term (but not for the near term), as well as an overall risk assessment. In surveillance and precautionary arrangement cases, most Directors endorsed full disclosure of sovereign risk analysis to the Board but limited

[1]At the conclusion of the discussion, the Managing Director, as Chairman of the Board, summarizes the views of Executive Directors, and this summary is transmitted to the country's authorities. An explanation of any qualifiers used in summings up can be found here: http://www.IMF.org/external/np/sec/misc/qualifiers.htm.

disclosure (omitting the near-term risk signal and assessment) to the public for a 12-month period, at which time full disclosure to the public would be reconsidered based on the experience gained with the new framework. A number of Directors expressed concern about the unintended consequences from potential market sensitivities of full disclosure of sovereign risk. A number of other Directors favored moving to full disclosure of sovereign risk analysis to the public immediately. Directors noted that implementing the limited disclosure options would require a targeted modification to the Transparency Policy, which would be proposed on a lapse-of-time basis.

Directors agreed that sustainability assessments should be required for arrangements involving GRA resources (including precautionary arrangements) as well as for the PCI. While most Directors agreed that sustainability assessments should be optional in surveillance cases, a few Directors favored preparing a sustainability assessment in surveillance cases with high risk of sovereign stress, with the results disclosed to the Board but not to the public, although a few other Directors would favor public disclosure even for such cases. With respect to program cases, a range of view were expressed. Some Directors preferred maintaining the current practice by which a three-zone assessment is included in staff reports in exceptional access cases but not in normal access cases. A few Directors suggested full disclosure (to the Board and the public) of three-zone assessments in both normal and exceptional access cases. In the end, Directors could go along with disclosure to the Board of three-zone assessments in both normal and exceptional access cases, and to the public only in exceptional access cases, with experience assessed at the end of a 12-month period.

In the context of precautionary arrangements, Directors agreed that sovereign risk assessments would be informed by the baseline scenario, while sustainability assessments would be informed both by the baseline and, when appropriate, by an adverse (full drawing) scenario. They agreed that the latter would be appropriate in exceptional access cases (excluding FCL cases), if shocks triggering a drawing are not adequately captured by the medium-term tools, or when review departments have doubts about the realism of the baseline that cannot be resolved through discussions with the country team, although a few Directors stressed that the appropriate use of the new realism tools should resolve any such doubts.

While most Directors supported the proposed timeline, with a carefully planned roll-out expected for Q4 2021 or Q1 2022, some Directors favored a more accelerated schedule, and a few others considered the proposed timeline could be ambitious. In this context, the transition between the old and the new framework should be carefully managed to ensure consistency. Directors looked forward to the preparation of a guidance note and new templates underpinning the new framework, accompanied by early engagement with a subset of country teams to test the new tools in parallel with the current framework. They encouraged the provision of appropriate capacity development support and maintaining close engagement with the Board as the framework is implemented, as well as ensuring an effective communication strategy with member-country authorities and external stakeholders during this process.

INTERNATIONAL MONETARY FUND

November 25, 2020

REVIEW OF THE DEBT SUSTAINABILITY FRAMEWORK FOR MARKET ACCESS COUNTRIES

EXECUTIVE SUMMARY

A careful review has revealed significant scope to modernize and better align the MAC DSA with its objectives and the IMF's lending framework. While the current framework has broadened the Fund's analysis of debt sustainability, its capacity to predict sovereign stress has been limited. In addition, because the framework relies on several separate indicators/outputs, its results are hard to summarize and communicate. For related reasons, it does not lend itself to supporting the three-zone sustainability assessment (sustainable with high probability; sustainable but not with high probability; unsustainable) required under the exceptional access framework.

This note proposes replacing the current framework with a new methodology based on risk assessments at three different horizons: *near-term*, based on a multivariate (logit) model predicting sovereign stress over 1–2 years; *medium term* (5 years) consisting of (i) a debt fanchart to assess prospects for debt stabilization, (ii) a module for more granular analysis of rollover risks, and (iii) triggered stress-tests to model specific risks (e.g. natural disasters, commodity price shocks, banking stress); and optional tools to analyze *long-term* risks (beyond 5 years). The new framework will require additional data and disclosure in some critical areas (debt coverage, liquid assets, holder and maturity profile of debt, and country-specific risks). This will entail additional resource costs, but is in line with institutional priorities, given rising global debt vulnerabilities. These costs will be contained with the help of automation.

Extensive testing has shown that the proposed framework has much better predictive accuracy than the current one. Furthermore, its results are easily communicated using mechanical risk signals (high/moderate/low) at both the near- and medium-term horizons. These could be reported in staff reports, as inputs into final, judgment-based risk assessments by staff at all three horizons. This would make the new framework more transparent than the current one, notwithstanding its greater sophistication.

In addition to predicting sovereign stress, the framework can be used to derive statements about debt stabilization under current policies and about debt sustainability. The former would be mandatory in surveillance cases, while three-zone debt sustainability assessments would be mandatory in program cases and optional in surveillance cases. Reflecting the framework's broad scope and purpose, the staff proposes renaming it "Sovereign Risk and Debt Sustainability Framework".

Approved By
Jeromin Zettelmeyer

Prepared by an IMF team led by S. Ali Abbas (MCD, formerly SPR) and Manrique Saenz, coordinated by Daniela Marchettini, and consisting of Tom Best, William Diao, Sandesh Dhungana, Amr Hosny, Maximilien Kaffo, Geoffrey Keim, Samuel LaRussa, Mike Li, Alexis Meyer Cirkel, Monica Petrescu, Modeste Some, and Tim Willems, all SPR. This paper has benefitted from extensive discussions with an interdepartmental working group, consultations with external experts and stakeholders, and invaluable guidance by Mark Flanagan (EUR), Hugh Bredenkamp (formerly SPR), and Craig Beaumont (SPR).

CONTENTS

BACKGROUND

1. The MAC DSA—the Fund's framework for assessing public debt sustainability in market access countries (MACs)—is critical for the exercise of the Fund's core functions of surveillance and lending.[1] [2] *In surveillance*, the framework helps identify a member's vulnerability to sovereign stress to steer the member away from such stress. In *Fund-supported programs*, which generally take place after stress has already developed, the DSA helps determine if the stress can be resolved via adjustment and Fund liquidity support, or if exceptional measures are needed to deliver "medium-term debt sustainability"—a pre-requisite for all Fund lending. The framework is also used to develop IMF debt conditionality and to inform debt relief and financing requirements in debt restructuring operations undertaken in the context of Fund supported programs, and as an input into fiscal space assessments.

2. The last review, undertaken in 2011-13, implemented significant reforms to the framework (Box 1).[3] The review set an expectation that debt coverage should encompass all material on- and off-balance sheet risks, including long-term risks, and introduced the option of projecting beyond the usual 5-year horizon. "Realism tools" to guard against baseline optimism were designed for growth, primary balance, and inflation projections. A heatmap summarizing breaches of vulnerability thresholds for debt, gross financing need (GFN), and debt profile indicators was introduced to inform and give more structure to the Fund's bottom-line assessments. Debt fancharts were rolled out to capture the distribution of risks around the baseline. Finally, the DSA template, write-up, and publication requirements were standardized to enhance transparency.

3. A careful review over the past two and a half years has revealed scope to further strengthen and modernize the framework, including in light of developments since the last review. While the 2013 framework was a major step forward from the previous one, introducing and operationalizing a broader definition of debt sustainability and strengthening the Fund's analysis of debt vulnerabilities, its capacity to predict sovereign stress has been mixed. Moreover, some perennial problems, such as uneven/inadequate debt coverage, baseline optimism, and unclear bottom-line assessments, have proven difficult to address. Judgment, which might be relied upon to offset these shortcomings, has faced challenges in suppressing the noise generated by the mechanical framework, and has not been applied in a transparent manner. In addition, the existing MAC DSA methodology does not provide a sufficient basis for probabilistic debt sustainability assessments. As such, it does not naturally lend itself to supporting the three-zone sustainability

[1]MACs refer to countries that principally receive financing through market-based instruments and on non-concessional terms. In contrast, countries that mainly rely on concessional financing usually use the debt sustainability framework for low-income countries (LIC DSF), which better accounts for the key role of the present value (PV) of debt in understanding debt-related vulnerabilities. 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 DSA [IMF (2017), section II.A].

[2]To assess sustainability of total (public and private) external debt, a separate framework is used (the "External DSA"). This framework will be reviewed separately in the future.

[3]IMF (2011); followed by IMF (2013).

assessment (sustainable with high probability; sustainable but not with high probability; unsustainable) required under the 2016 exceptional access lending framework. Finally, while there is a trade-off between the need to capture diverse sovereign risk sources and the desire for simple, transparent models, advances in sovereign risk and debt sustainability analytics since 2011 open up the possibility of improving this trade-off.

Box 1. A Brief History of the MAC DSA

The MAC DSA was introduced in 2002 to improve the consistency and discipline of debt sustainability analyses.[1] The initial framework aimed to illustrate debt projections, the underlying assumptions regarding their drivers, and the sensitivity of the debt path to standard stress scenarios. The MAC DSA was reviewed in 2003 and 2005, to introduce some refinements.[2]

The last review of the MAC DSA, in 2011–13 introduced important reforms, in response to shortcomings revealed by the Global Financial Crisis (GFC) and euro area sovereign debt crises.[3] The review uncovered a number of shortcomings in the previous framework, including: inconsistent use or discussion of the public DSA; optimistic growth projections in several crisis countries, partly reflecting underestimated fiscal multipliers; the lack of a bottom-line sustainability assessment; limited scope for country heterogeneity; and ineffective tools to illustrate uncertainty (as shock scenarios were often dwarfed by ex-post debt increases, including from underappreciated sovereign-financial sector linkages).

The current MAC DSA framework, launched in 2013 in response to this review, introduced the following key reforms:

- A requirement for at least one DSA per year for program cases and one per Article IV cycle for non-program cases;
- A risk-based approach, i.e. a distinction between high- and low-scrutiny countries in analyzing debt vulnerabilities across the Fund's membership;
- New elements in the DSA template and output including: (i) an analysis of the realism of baseline projections, (ii) a heat map anchored by noise-to-signal (NTS) based thresholds for debt, GFNs, and debt profile indicators to summarize sustainability risks, and (iii) debt fancharts to capture the full distribution of risks around the baseline.

---

[1]IMF (2002) Assessing Sustainability.
[2]IMF (2005) Modifications to The Fund's Debt Sustainability Assessment Framework for Market-Access Countries.
[3]IMF (2011) Modernizing the Framework for Fiscal Policy and Public Debt Sustainability Analysis; IMF (2013) Staff Guidance Note for Public Debt Sustainability Analysis in Market Access Countries.

4. Against this backdrop, in this paper staff proposes a set of reforms to the framework. These aim to: (i) increase the robustness of sovereign risk analysis through broader debt coverage, a longer projection horizon and enhanced realism tools; (ii) improve the framework's capacity to predict sovereign stress through new analytical tools at three different time horizons that both account for countries' structural characteristics, and rely on continuous metrics rather than discrete, single-variable thresholds; and (iii) enhance transparency in exercising judgment and arriving at (horizon-based) bottom-line assessments. Furthermore, the new tools support probabilistic debt sustainability assessments, as required by the Fund's lending framework.

5. The remainder of this paper proceeds as follows: The next section presents a discussion of the related but distinct concepts of debt sustainability and risk of sovereign stress. This is followed by an assessment of the current framework, as envisaged in the 2011 review and implemented in the 2013 Guidance Note. Staff then sets out its reform proposals, both for surveillance cases and sustainability assessments in program cases. The final sections lay out reporting requirements and implementation issues.

DEFINITION OF DEBT SUSTAINABILITY

6. The Board-approved definition of public debt sustainability is as follows:

"In general terms, public debt can be regarded as sustainable when the primary balance needed to at least stabilize debt under both the baseline and realistic shock scenarios is economically and politically feasible, such that the level of debt is consistent with an acceptably low rollover risk and with preserving potential growth at a satisfactory level."[4]

7. The definition includes both solvency and liquidity requirements. It implies that a member is solvent (possibly conditional on a feasible set of policy adjustments), *and* that liquidity risks are contained. In this regard, the Fund's definition of debt sustainability is somewhat stronger than the definition sometimes found in academic literature, which focuses only on solvency. This can be justified in two ways. First, in practice, a clear-cut distinction between solvency and liquidity risks is impossible, since borrowing costs and market access depend on (actual and perceived) solvency. Hence, any attempt to model uncertainty around the baseline debt path (e.g., in the form of a fanchart) must account for liquidity risks. Second, the IMF's lending framework uses debt sustainability as an indicator of the capacity of the member to repay the Fund. The latter could be impaired not just by insolvency but also by lack of liquidity, particularly if this is persistent.

8. The definition is applied somewhat differently in surveillance and program contexts.

- In a *surveillance* context, the IMF's role is to alert the member to the likelihood of sovereign stress and help it steer away from it. Solvency and liquidity risks are equally relevant for sovereign stress, which can manifest itself in the form or high/rising borrowing spreads, high/rising inflation, loss of market access, default, among others. In signaling that such stress exists, the Fund takes no view on how it will be resolved (i.e., via adjustment, via some combination of adjustment and external financing, or through exceptional measures like debt restructuring). Sovereign stress is not the same as unsustainability of debt, which (according to the IMF's definition) implies that there is no feasible policy set to stabilize debt at a level consistent with manageable rollover risk and satisfactory potential growth.

- In a *program* context, assessments need to pay particular attention to (conditional) solvency, but cannot ignore market financing risks that remain *after* Fund financing has been incorporated in the baseline. Short-term liquidity risks can be reduced (and sometimes eliminated) by IMF financing.

[4]IMF (2013), Section I.

At the same time, liquidity risks matter even in the context of IMF-supported programs, for two reasons. First, market access after the successful completion of a program is one of the determinants of the member's capacity to repay. Second, for countries that must continue to borrow from the market during the program period,[5] liquidity risks can undermine the required access to finance and hence program success.

9. The foregoing illustrates the need to distinguish between three related, but distinct, concepts: sovereign stress, debt non-stabilization under the baseline and unsustainable debt (Box 2). According to the Board-approved definition, debt sustainability requires debt to stabilize with low financing risks under a *feasible* set of policies, but not necessarily under the policies assumed in the baseline. This distinction is not important in program contexts, as staff's baseline reflects a set of feasible policies that will stabilize the debt; but it is important for surveillance cases, when the baseline reflects the policies that are most likely to be implemented, whether or not they stabilize the debt. Even if debt does not stabilize under the baseline, it is possible that there is a feasible adjustment scenario under which debt would stabilize (while keeping financing risks low), consistent with sustainable debt under the Board-approved definition. Current reporting practices in surveillance cases sometimes appear to conflate debt non-stabilization under policies assumed in the baseline with debt sustainability (under feasible policies). Going forward, it would be good to use these terms more consistently.

10. To clarify the scope and purpose of the IMF's MAC DSA framework, staff proposes renaming it "Sovereign Risk and Debt Sustainability Framework" (SRDSF), while its output would be referred to as "Sovereign Risk and Debt Sustainability Analysis" (SRDSA). The Fund's DSA frameworks (both the LIC DSF and the MAC DSA) are first and foremost tools to analyze and warn about the risk of sovereign stress, both because this is of intrinsic interest to inform policies and because it provides information for deciding whether debt is sustainable. On this basis, they also allow staff to make a debt sustainability assessment (although this is a *requirement* only in a program context).[6] Acknowledging the multiple functions of the framework in its title will avoid confusion about its nature and purpose.

[5]Fund credit may have an inherent limit, since very large provision of Fund liquidity may create subordination risks for future private creditors, thus undermining market (re-)access prospects for the member.

[6]Paragraphs 72-73 outline how staff proposes to repurpose the stress framework illustrated in this paper to make sustainability assessments in program contexts.

Box 2. Risk of Sovereign Stress, Unsustainable Debt, Debt Non-Stabilization

It is helpful to distinguish between three concepts that are frequently referred to in Fund DSAs:

Risk of sovereign stress refers to the likelihood of a sovereign experiencing stress, regardless of whether and how that stress could be resolved (i.e., via adjustment, adjustment and financing, or exceptional measures like debt restructuring). Hence, high risk of sovereign stress is a broader concept than "unsustainable". Both the current and proposed tools are designed to capture risk of sovereign stress.

Unsustainable debt is the complement of the Fund's definition of "sustainable debt" noted in paragraph 6 above. It means that there is no set of politically and economically feasible policies that can stabilize the debt/GDP ratio with acceptably low rollover risk. When the Fund pronounces debt as unsustainable, this implies that no combination of adjustment and Fund financing can solve the problem, and debt restructuring (and/or exceptional bilateral support) is necessary.

Failure of debt to stabilize under policies assumed in baseline describes a situation in which a country's debt/GDP ratio does not stabilize under staff's best prediction of policies by the end of the projection horizon. In some surveillance cases, most notably the U.S. and Japan, this has been referred to as "debt is on an unsustainable path under current policies". While this was *not* intended to imply that the policies that would be required to stabilize debt are infeasible – implying unsustainable debt under the Fund's formal definition – there is a risk that the use of the word "unsustainable" conflates the two concepts, although they are quite different. Failure of the debt to stabilize under any set of feasible policies generally implies that a debt restructuring is necessary; while failure of the debt to stabilize under policies in baseline may only imply that macro/fiscal adjustment is necessary.

The relationship between the three concepts is summarized in the Venn diagram below. The risk of sovereign stress (red set) is a broader concept than unsustainability of debt (blue set). Failure of debt to stabilize under current policies (green set) may or may not imply sovereign stress and may or may not imply that debt is unsustainable in the IMF's definition, as there may be feasible policies to avert both. Finally, it is conceivable that debt is not sustainable in the IMF's definition even if it stabilizes under policies assumed in the baseline (a deterministic statement), namely, when the risk of not being able to meet debt obligations as a result of shocks or rollover difficulties is high and cannot be addressed with a feasible set of policies.

ASSESSMENT OF THE CURRENT FRAMEWORK

11. As part of the last review of the MAC DSA in 2011–13, the Board endorsed five broad objectives against which a sound public DSA framework should be assessed. These could be paraphrased as: (i) adequate coverage and disclosure of debt-related risks, (ii) discriminatory (predictive) capacity, (iii) realistic baselines, including a robust representation of uncertainty around the baseline, (iv) a risk-based approach, with countries at higher risk subject to greater scrutiny, and (v) a sharper output that allows an effective, transparent, and even-handed application of judgment.

12. Staff has carefully evaluated the existing framework against these objectives, while also taking into account new analytical methods and Fund policy requirements. This assessment has revealed several areas for improvement, detailed below and summarized in Table 1.

A. Adequate Coverage and Disclosure of Debt-Related Risks

13. Although the 2011–13 review introduced general government debt as the appropriate concept for DSAs, actual coverage has remained narrow in many cases. The 2011 Board paper envisaged anchoring Fund DSAs in the "general government debt concept, in line with the GFSM and Manual on Fiscal Transparency". The 2013 Guidance Note reaffirmed this ambition, even if the actual requirement was less stringent.[7] However, the current review has found that about two-fifths of EMs still restrict coverage to the central government, with little improvement in coverage over time (Annex I). In addition, the absence of a requirement to report the instrument and valuation basis for the debt reported in the DSA further obscures cross-country comparisons and undermines evenhandedness in the Fund's assessments of sovereign risks and debt sustainability.

14. There is insufficient information on the debtholder profile in the current framework, even though this can be a key amplifier or mitigant of sovereign risks. There are no disclosure requirements on who holds the debt (beyond a resident/non-resident split), even though assessing rollover risk and evaluating safeguards for Fund resources can depend critically on how much of the debt is held by specific categories of creditors.

15. Several additional coverage issues require attention. These include: how to (i) better incorporate liquid assets in the analytical tools (they currently only enter as part of judgment, or to help classify countries as low-scrutiny); (ii) report contingent liability risks arising from narrow institutional coverage, government guarantees, private-public partnerships (PPPs), and special purpose vehicles (SPVs); (iii) treat Fund credit intended for boosting reserves (there is no explicit guidance on this issue, although legally all Fund credit should be included in the public debt); (iv) account for central bank liabilities, such as FX swaps and liquidity paper (both increasingly important

[7]The 2013 Guidance Note emphasized "general government" in a box, but the actual requirement was cast as: "The coverage of public debt in the DSA should be as broad as possible, but consistent with the coverage of the fiscal accounts monitored for surveillance and program purposes".

in recent years);[8] and (v) deal with cases where the central bank holds a significant share of government debt (and consolidating these holdings could materially reduce measured debt and GFNs).

16. Finally, assessments have rarely considered long-term fiscal pressures. Longer horizons are essential for capacity-to-repay assessments in UFR cases and for setting/evaluating debt and GFN targets in restructurings undertaken in the context of Fund programs; they are routinely used in DSAs of other institutions, as well as the IMF-World Bank LIC DSF.[9] The 2013 guidance note allowed teams to project beyond 5 years where long-term fiscal pressures were relevant. However, long-term projections have been limited, potentially reflecting a lack of tools for analyzing specific risks.

B. Discriminatory (Predictive) Capacity

17. The 2011–13 review aimed at sharpening the analysis of sovereign risks beyond analyzing debt trajectories, by introducing thresholds for debt, GFN and other variables. The 2011 Board paper emphasized that "risks vary considerably across countries", including external, financial and public sector off-balance sheet risks, and suggested integrating the assessment of debt structure and liquidity into the DSA. To do this, staff estimated 7 thresholds (for debt, GFN, and five debt profile indicators), with breaches shown in a heatmap. The thresholds were calibrated to minimize errors in predicting crises[10] one year ahead. To accommodate some country heterogeneity, staff adopted different risk thresholds for AEs and EMs.

18. The thresholds and heatmap supported staff policy messaging, but the predictive power of threshold-based signals has been weak. The introduction of thresholds for debt levels helped sensitize country authorities and Fund teams to the risks associated with high debt. The analysis of GFN levels and debt profile and market indicators mainstreamed liquidity and rollover risk analysis in Fund surveillance and lending. However, performance of the single-variable thresholds has been weak: in 57 percent of all stress episodes that occurred in AEs in 2007–13 and EMs in 2007–18, *both* the debt *and* the GFN indicators failed to flash (Annex I). Debt profile indicators, notably those capturing external vulnerabilities, actually showed greater discriminatory power relative to debt and GFN.

19. The limitations of the threshold approach were already known at the time of the 2011–13 review. The 2011 Board paper noted the "lack of empirical basis for generalized debt thresholds." When the thresholds were introduced in 2013, the final risk assessment was expected to

[8]For example, a bilateral swap line extended by the People's Bank of China was an important source of financing (roughly 20 percent of GDP) for the Mongolia 2017 EFF. Liquidity papers in the form of central bank-issued securities amounted to a non-trivial 5 percent of GDP in Argentina at end-2018 (and rolling them over at one-month durations proved challenging).

[9]The European Commission uses a 10-year horizon, the LIC-DSF a 20-year horizon.

[10]The terms "crises" and "stress episodes" are used interchangeably in the note and refer to any episode identified by one of six stress criteria: defaults, restructurings, large financing from the Fund or non-Fund official sources, high/rising spreads or loss of market access, high/rising inflation, financial repression (See Annex IV).

be based on the *overall* heatmap rather than on the breach of individual thresholds; however, the guidance note did not indicate *how* to aggregate the individual indicator breaches in the heatmap.

20. The present framework does not perform well in two additional respects:

- ***Similar heatmaps for countries with very different risk profiles:*** Similarities in the heatmaps of AEs and EMs facing different risks suggest two country buckets (AEs and EMs) may not be sufficient to capture the wide variation in debt carrying capacity across MAC DSA countries (Annex I);

- ***Lack of attention to the timing of and magnitude of specific risks.*** The heatmap treats threshold breaches the same regardless of whether they occur early or late in the projection horizon, although the urgency of action is much greater in case of the former; the heatmap also treats breaches in the previous year the same as those occurring during the projection horizon, even though breaches in the past do not require action unless they are expected to persist. Further, because the heatmap does not account for the magnitude of the breach, it may not change from year to year despite material changes in risks.

21. While re-estimation of current thresholds on a larger sample of countries or stress episodes might improve performance, it would not address the key limitations of the current empirical approach. *First*, several known predictors of sovereign stress – institutional quality, history of stress, cyclical imbalances, and global risk appetite – are omitted from the present heatmap.[11] *Second*, separate single-variable group-wide thresholds do not account for interactions among variables, which may be relevant to appropriately capture heterogeneity across countries.[12]

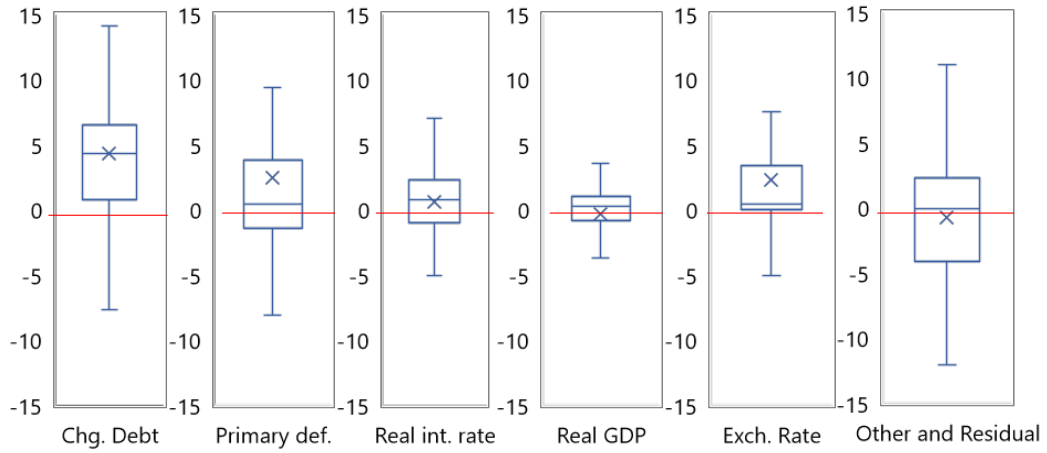
C. Baseline Realism and Robust Modeling of Uncertainty

22. The introduction of visual realism tools in the 2013 framework appears to have helped reduce optimism in baseline projections, but forecast debt trajectories remain optimistic. On average, projections errors for debt drivers covered by the realism tools—primary balance and real growth rate— were smaller than for those not covered (exchange rate and interest rate; see Figure 1). However, forecasts for the change in debt/GDP remain more optimistic than outturns, and medium-term debt stabilization is predicted more frequently than it occurs (Annex I).

[11]These variables feature prominently in private sector sovereign risk models, as well as other official DSA frameworks (EC framework, LIC DSF).

[12]Continuous models have been found to have better predictive performance than threshold-based approaches when applied to the same set of variables, because the latter only classify countries "in stress/not in stress", without giving weight to the severity of the breach, and do not account for variable interaction (see Berg and others, 2014).

Figure 1. Decomposition of 3-Year Forecast Errors in the Debt/GDP Ratio (2013–17)



Source: MAC DSA database.
Note: The figure indicates the difference between the outturns and the teams' forecasts. x is the average difference between outturns and forecasts, the horizontal line is the median difference, the box shows the interquartile range, and the whiskers are upper and lower limits within 1.5 times the 1st and 3rd quartile levels.

23. Separately, the 2013 framework's continued emphasis on deterministic stress-tests has likely constrained staff's ability to capture country-specific risks. The macro-fiscal stress-tests introduced in 2013 sought to account for the links between key variables. However, the stress-tests are deterministic, and the assumed shocks and their transmission mechanisms are set to be identical across all countries. Moreover, because the shocks are simulated individually, the stress-tests remain far from realistic stress scenarios, in which several things tend to go wrong at the same time.[13]

24. At the same time, stochastic tools like debt fancharts are underutilized and, as currently designed, suffer from methodological shortcomings. Relative to stress-tests, fancharts play a peripheral role in the current MAC DSA framework, even though the information contained in fancharts largely subsumes that in stress-tests.[14] The current fanchart methodology also suffers from shortcomings related to: (i) assumptions about the normality of the distribution from which debt driver shocks are derived; (ii) failure to provide an accurate picture of risks if the baseline is optimistic and/or risks are tilted to the downside; (iii) lack of standardization in deriving the asymmetric fanchart, precluding comparability across countries; and (iv) inability to account for uncertainty related to debt data revisions (base effects).

[13]While there is a combined macro-fiscal stress-test which consolidates the individual shocks, it does not inform the heatmap and was found to have weaker effects on debt and GFNs than observed in actual stress episodes.

[14]80 percent of the debt and GFN trajectories under the primary balance, interest rate, and exchange rate stress-tests fell within the 10-90th percentiles of the symmetric fanchart.

D. Classification of Countries into Low and High Scrutiny "Buckets"

25. While the principle that higher risk countries should receive more attention than lower risk countries is reasonable, its application in the form of low and high scrutiny buckets has revealed some problems.

- ***Countries classified as low scrutiny ended up in crisis.*** The weak near-term predictive power of the debt/GDP and GFN/GDP thresholds has resulted in several low scrutiny countries (e.g., Bosnia, Georgia) entering into stress since the introduction of the framework. The criteria for low scrutiny would have also precluded deeper analysis for several low debt and GFN AEs that experienced stress post-GFC before the introduction of the framework (e.g. Ireland, Iceland, Spain).

- ***Within the high-scrutiny group, the analysis is not adequately risk-based.*** The risk-based approach was introduced to help reconcile the need for deeper analyses in certain cases with constraints on staff resources. However, differences *within* the high-scrutiny bucket are not fully captured: for instance, there is virtually no use of triggered stress-tests (beyond the generalized contingent liability shock) that might shed light on key vulnerabilities.

E. Granularity, Aggregation, and Application of Judgment

26. The gaps in the 2013 framework have generated challenges for the effective, transparent, and evenhanded application of judgment. Flexibility in applying judgment is essential to ensuring country-specific nuances are adequately captured. However, the limited accuracy of the threshold approach, the lack of granularity in capturing risks, and the absence of a cohesive aggregation of outputs in the current framework limit the framework's ability to inform judgment. Specifically:

- *The predictive shortcomings of the mechanical tools place an undue burden on judgment*. The weak performance of the heatmap in the current framework requires teams to make up for shortcomings through judgment. However, the review finds that team judgment did not improve upon the performance of the heatmap in either stress or non-stress cases (Annex I).

- *Judgment rarely offsets the lack of tools to adequately capture country-specific aggravating and mitigating factors*. The current framework lacks tools for assessing key vulnerabilities facing groups of countries such as (i) natural disasters in small states; (ii) commodity price swings and resource depletion/discovery in commodity exporters; (iii) long-run fiscal costs; and (iv) large government assets.[15] Such factors could be incorporated in judgment, but in most DSAs under the current framework, the chapeau lacked a discussion of key risks and mitigating factors not already captured by the mechanical tools (Annex I).

[15]There is no guidance on when to invoke the option to perform a supplemental DSA using net debt.

- *Framework output is not cohesively aggregated to provide a starting point for the use of judgment.* The output of the current framework consists of many (at times conflicting) disaggregated results from the heatmap and other tools;[16] how to interpret and incorporate the results into a cohesive analysis is left to team's discretion. As a result, it is often difficult to understand whether—and, if so, why—staff disagrees with the framework's output (Annex I).

27. Finally, the current framework does not require *standardized reporting* of risks in surveillance cases, limiting the Fund's ability to communicate effectively. While Fund programs require a bottom-line statement on debt sustainability, there are no analogous requirements in surveillance cases. Teams use the DSA chapeau to summarize their assessments on risks, but these assessments can be crafted in many different ways, precluding meaningful cross-country comparisons, or comparisons for the same country over time. External stakeholders identified the lack of aggregate mechanical signals, unclear bottom-line assessments, and non-transparent application of judgment as the framework's most notable weaknesses, and highlighted that both the EC's framework and the LIC-DSF perform better in this respect.

[16]The tools outside the heatmap that do not produce any signals (e.g. the historical or constant primary balance scenarios, the boom-bust module, and the fanchart) pose an additional challenge to meaningful aggregation.

Table 1. Summary of Assessment of Current Framework and Scope for Improvement				
Objective	**2011 Board Paper**	**2013 GN implementation**	**Assessment of performance**	**Possible improvements**
Adequate coverage and disclosure of debt-related risks	- General Government (GG) debt to be the standard coverage - Emphasis on off-budget entities/contingent liability risks - Longer projection horizon if 5-year horizon inadequate - DSA to be based on gross debt, but net debt concept to be reported where available	- Coverage as broad as possible, but consistent with fiscal accounts used for surveillance or program monitoring - Optional complementary analysis based on longer horizon when debt burden indicators remain high at t+5 - DSA to use gross debt; net debt could be reported in write-up; liquid assets could allow for low scrutiny	- Coverage narrower than GG for 40 percent of EMs; reported debt concepts differ widely across countries; often deviate from GFSM - Longer than 5-year horizon used for only five MACs - Lack of guidance on treatment of certain claims/liabilities, e.g. swaps	- More firmly establish GG as minimum coverage for DSA, and encourage broader coverage where economic case exists - Account for risks from narrow coverage - Enhance disclosure, esp. on debt structure - Consider 10-year horizon for all countries; and introduce tools to analyze specific risks beyond 5 years (e.g., aging) - Integrate liquid assets in a more standardized way
Discriminatory capacity (by accounting for country heterogeneity)	- Emphasis on country-specific risks/vulnerabilities - Recognized weak empirical basis for group-wide debt thresholds - Need to leverage other tools (FSAP, balance sheet approach) - Integration of debt structure	- AE/EM thresholds for debt; GFN, debt profile, market indicators - Heatmap to visually summarize risks - Boom-bust module for cyclical risks - Symmetric and asymmetric debt fancharts (not integrated with heatmap)	- Weak early-warning capacity of single-variable thresholds and heatmap - Beyond AE/EM distinction, no adjustment for countries' differential debt carrying capacity - Conflation of near- and medium-term risks	- Use multivariate early warning model(s) with broader set of stress drivers, mitigants - Develop tools/metrics for analyzing medium-term debt and GFN vulnerabilities that take better account of differences in institutional capacity and creditor profile. - Analyze risks by horizon
Baseline realism, robust modelling of uncertainty	- Need to scrutinize the realism of debt driver projections - Need for stochastic simulations	- Visual realism tools for growth, primary balance and inflation - Deterministic stress-tests/scenarios, and symmetric/asymmetric fancharts	- Smaller forecast errors for variables covered by realism tools, but still-large errors for debt, and other drivers - Symmetric fanchart biased by baseline optimism; asymmetric fanchart not standardized	- Enhance realism toolkit to cover exchange rate, interest rate, stock-flow adjustments - Mainstream use of continuous, stochastic methods to better capture risk distribution - Standardize fanchart, robust to optimism
Risk-based approach	- Depth of analysis to be proportional to vulnerabilities	- Debt, GFN levels used to divide MACs into low/high scrutiny groups - Streamlined DSA for low-scrutiny countries	- Low scrutiny countries have entered stress; - Differences in types of vulnerabilities for high-scrutiny countries not captured	- Ensure minimum risk analytics for all countries - Introduce triggered stress-tests/modules for key vulnerabilities (e.g., commodity prices, natural disasters, aging population)
Standardization and transparency, with room for judgment	- Focus on debt, GFN projections - Judgment-based assessments, but expectation to sharpen/ standardize analysis of risks - Need for consistent/regular publication of DSAs	- Introduction of thresholds/heatmap to aid/sharpen messaging on risks - Team judgment to aggregate risks in detailed DSA write-up - Standardization of DSA template, write-up, publication requirements	- Absence of aggregate risk signals implying no clear mechanical bottom-line (unlike in EC framework and LIC DSF) - Lack of transparency on how final assessment arrived at	- Introduce horizon-based summary mechanical risk signals to provide clear starting points for application of judgment - Allow final horizon-based assessments to be judgment-based, but require explanation where judgment deviates from mechanical signals

STRUCTURE OF THE PROPOSED FRAMEWORK

28. The remainder of this paper lays out a root-and-branch reform aimed at fixing the main problems identified in the current framework, consistent with the high-level objectives set out in the 2011 paper and endorsed by the Board. Concretely, the reform aims to:

- increase the *robustness* of the analysis of sovereign risks through broader and more consistent debt coverage, a longer projection horizon, enhanced realism tools, and superior analytical methods;
- improve the framework's *discriminatory (predictive) capacity* by introducing a horizon-based approach that accounts for country-specific structural characteristics and uses continuous rather than discrete metrics;
- and enhance *transparency* in the bottom-line assessment (for each horizon) and in the exercise of judgment.

Furthermore, the new tools support probabilistic debt sustainability assessments, as required by the Fund's lending framework. Figure 2 illustrates the new framework and Table 2 summarizes the proposed reforms.

Figure 2. The Architecture of the Proposed MAC SRDSF

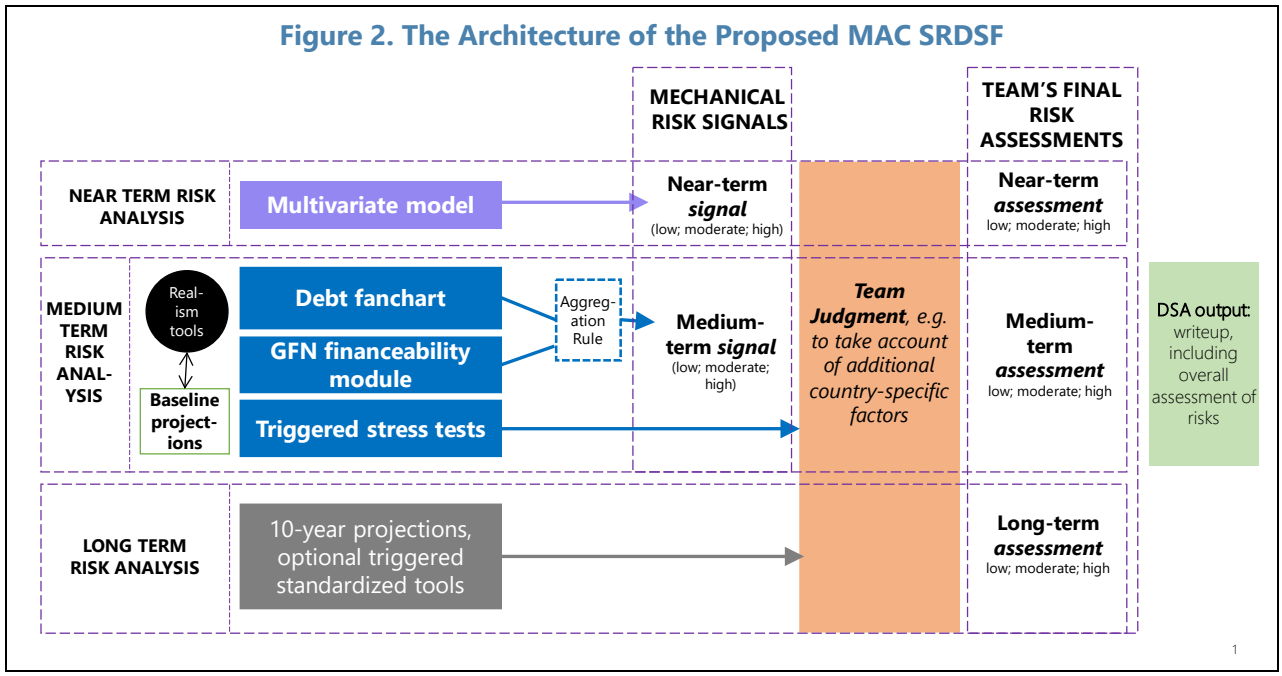


Table 2. Comparison of Key Features of Proposed and Existing Frameworks		
	Proposed framework	**Existing framework**
Coverage	• GG as default; justification required for narrower coverage; broader coverage (including central bank) in some cases • Disclosure requirements on coverage definitions, debtholder profile, and guidance on certain instruments (like swaps)	Narrower than GG in some cases; no disclosure requirement on coverage
Horizon	• 10-year debt and GFN projections for all cases • Risk assessments for near-, medium-, and long-term horizons	• 5-years projections • No distinction in horizons
Realism tools	• Cover additional drivers (exchange rate, financing terms on external debt, stock-flow adjustments), and public debt • In-depth tools for potential growth and fiscal multipliers.	Cover growth, inflation and primary balance.
Near-term risks		**Heatmap**
Stress indicators	• 10 indicators, in five categories: quality of institutions, stress history, cyclical, debt burden, and global[1]	• Debt and GFN levels, five debt profile and market indicators each producing a risk signal for heatmap
Composite index	• Multivariate logistic regression combines indicators in a continuous metric (fitted probability of stress)	
Signal derivation	• Stress probability split in low, moderate, and high-risk zones, (calibrated to 10% missed crisis and false alarm rates)	• No aggregation/overall signal
Medium-term risks		
	Debt fanchart	***Fanchart tool***
Stress indicators	Three indicators: i) probability debt does not stabilize in medium term, ii) fanchart width, and iii) debt level at t+5 controlling for debt-carrying capacity (fanchart accounts for deviation of baseline projections from historical trends via skewed shocks)	Visual tool based on symmetric shocks (asymmetric shocks used at team's discretion)
Composite index	Index based on 3 indicators weighted by predictive power	No signal/indicators; interpretation of fancharts at team's discretion.
Signal derivation	Index split in low, moderate, and high-risk zones, (calibrated to 10% missed crisis and false alarm rates)	
	GFN Tool	***Macro-fiscal shocks***
Stress indicators	Three indicators: (i) initial bank claims on government, (ii) maximum cumulative change in bank claims over projection period under a generalized stress scenario; (iii) average projected GFN/GDP in baseline.	• Effect of shocks to primary balance, real GDP growth, real interest rate, and exchange rate on debt and GFN levels reflected in heat map signals
Composite index	Index based on 3 indicators weighted by predictive power	
Signal derivation	Index split in low, moderate, and high-risk zones, (calibrated to 10% missed crisis and false alarm rates).	• No aggregate signal
Triggered stress-tests	• Simulate debt and GFN paths under: (i) contingent liabilities related to narrow coverage, (ii) banking crisis, (iii) natural disasters, (iv) commodity price shocks, and (v) REER shock. • Allows for customized stress-tests for idiosyncratic risks.	Allows for customized stress-tests for idiosyncratic risks.
Long-term risk analysis	Optional tools for risks from: population aging, natural resource discovery/depletion, debt amortizations; and climate change.	Option to extend debt and GFN projections
Judgment and communication	• Judgment-based risk *assessments* at each horizon, with deviation from mechanical signals explained. • Overall risk assessment based on team judgment.	No aggregate mechanical signals; lack of standardized bottom-line assessments; unclear application of judgment.

[1]Extensive in-sample and out-of-sample testing used to select regressors and check robustness of specification.

29. The proposed framework comprises tools to assess sovereign risks at three horizons:

- ***Near-term risks*** **(1–2 years ahead)** analyzed using a multivariate logit model, based on actual (i.e. not projected) data. This model aggregates information from a range of both stress drivers and mitigating factors and hence accounts for country-specific structural characteristics (thus allowing a more granular discrimination of countries than the current AE/EM bucketing).

- Tools to analyze ***medium-term risks*** **(up to 5-years ahead),** including: (i) a debt fanchart to probabilistically assess prospects for debt stabilization; (ii) a GFN module to analyze rollover risks, taking into account a country's creditor composition; and (iii) triggered/tailored stress-tests to assess country-specific risks not captured elsewhere.

- ***Long-term risks*** **(beyond 5 years)** captured by 10-year debt and GFN projections, and a suite of optional tools to analyze specific vulnerabilities.

30. Each element of the proposed framework has been rigorously tested and audited.[17] Overall predictive performance, as measured by the sum of missed crisis and false alarm rates, significantly improves relative to the existing framework's single-variable thresholds (see Table 3).

Table 3. Minimum Total Misspecification Error of the Proposed SRDSF Tools[1]

		NTS Ratio	Minimum total mis-specification error	*of which:* Missed crises	*of which:* False alarms	AUC
Near-term risk analysis	Logit model (1990-2015)[2]	30%	37%	10%	27%	0.88
Medium-term risk analysis	Medium-Term Index (2014-15)	14%	38%	27%	10%	0.85
	Debt fanchart (2010-15)	34%	43%	14%	29%	0.82
	GFN module (2014-15)	13%	42%	33%	9%	0.83
All indicators in heatmap (current framework)	OR condition (2006-16, AE and EM average)	76%	79%	12%	68%	
	AND condition (2006-16, AE and EM average)	NaN	100%	100%	0%	

Source: Fund staff calculations.

[1] The Total Misspecification Error (TME) is the sum of the probabilities of missed crisis and false alarm errors. The minimum TME provides information on the discriminatory capacity of the corresponding tools based on a single threshold that divides the space of possible results in two zones (high risk, predicting a crisis; and low risk, predicting no crisis). In the case of the current framework (last two lines) two options are explored for how this space could be divided: (1) a crisis is predicted if just one of the heatmap indicators flashes red ("OR condition"); (2) a crisis is predicted if all heatmap indicators flash red ("AND condition"). A crisis prediction based on the "OR condition" rarely misses a crisis (just 12%, about in line with the logit model and the debt fanchart tool), but at the cost of frequently sending false alarms (68%, much higher than any of the new tools). The TME is 79%, much worse than that of the new tools. A crisis prediction based on the "AND condition" never sends a false alarm (all crises are associated with at least one red signal) but misses all crises (no crisis is associated with the *entire* heatmap being red).

[2] Analysis based on 1,579 country-year observations for the logit model, 99 for the medium-term index, 403 for the debt fanchart, 125 for the GFN module, and 805 for the debt and GFN thresholds.

[17] Following the informal Board discussion held on May 29, 2020, all the analytical tools proposed in this paper have been subjected to an audit by an independent RES/ICD team. This paper reflects the results of this validation. Annexes IV and V respond to some points raised by the audit and discuss elements of the framework that were modified in reaction to the audit.

31. The design of the near- and medium-term tools is based on three steps:

i. **Identify relevant stress drivers and/or mitigating factors at each horizon.**

ii. **For each horizon, combine the relevant indicators into a continuous composite index.** Based on this composite index, a stress event (or crisis) is predicted if the value of the index exceeds a chosen threshold. For a given threshold and a sample period, one can calculate the percentage of missed crises (observed crisis events not predicted by the model, or Type I errors, divided by the number of crises observed) and the percentage of false alarms (observed non-crises events for which the model predicted a crisis, or Type II errors, divided by the number of non-crises observed). Hence, for each value on the composite index scale there is an associated percentage of missed crisis and false alarms that would be obtained if such index value were chosen as the threshold.

iii. **For easy and consistent communication, divide the index into three risk zones (low, moderate, high) based on two cutoffs** corresponding to probabilities of missed crises and false alarms. [18] Specifically:

- the risk of sovereign stress will be deemed "high" if the risk index exceeds the upper cut-off, corresponding to a false-alarm probability of 10 percent;
- the risk of sovereign stress will be deemed "low" if the risk index is below the lower cut-off, corresponding to a missed-crisis probability of 10 percent;
- the risk of sovereign stress will be deemed "moderate" for intermediate cases, with a risk index between the upper and lower cut-offs.

Steps (i) and (ii) are based on rigorous statistical procedures aiming to maximize predictive performance (see sections on tools for near- and medium-term risk analysis below, as well as Annexes V-VII). Step (iii) is underpinned by the good capacity of the tools to separate stress from non-stress episodes (see Table 3 and Box 3), which allows to calibrate the low- and high-risk thresholds to relatively low probabilities of missed crises and false alarms, respectively (10 percent), without implying a very wide moderate risk zone.

[18]There are several approaches to defining the three zones, including (1) equisized by number of countries (as in the Vulnerability Exercise), (2) using thresholds calibrated on missed crisis and false alarm rates, or (3) using thresholds calibrated on stress probabilities. (2) and (3) have the advantage that the stress probability associated with a particular risk classification does not change over time. Because thresholds calibrated based on missed crisis and false alarm rates imply a stress probability and vice versa, (2) and (3) would give almost the same result. The only difference is that in approach (2), thresholds are associated with the same probabilities of missed crises and false alarms for all tools while threshold stress probabilities vary slightly across tools (see Box 3), whereas in approach (3), the thresholds would be associated with the same crisis probabilities across tools, but vary slightly in terms of probabilities of missed crises and false alarms. Because the Fund has an institutional interest to avoid both missed crises and false alarms, staff is of the view that defining a maximum tolerance for missed crises and false alarms consistently across all tools of the framework is the preferred approach.

Box 3. Interpretation of "High", "Medium" and "Low" Risk Signals

Under the proposed framework, the risk of sovereign stress is classified as "high", "medium" or "low" based on the level of risk identified by the models. The thresholds separating the three buckets are set to target misclassification rates based on historical data. Specifically, the threshold separating the "low" and "medium" signal is chosen so that only about 10 percent of actual crises are misclassified as "low" risk (missed crisis rate of 10 percent). Similarly, the threshold separating "medium" and "high" signal is chosen so that only 10 percent of non-crisis events are misclassified as "high" risk (false alarm rate of 10 percent). This choice of thresholds reflects a maximum tolerance level for potential misclassifications, set at 10 percent.

Each risk category is associated with an ex-post crisis probability. While thresholds are designed to target misclassification rates (probability of signal, conditional on the outcome), ex-post probabilities (probability of outcome, conditional on the signal) provide guidance to country teams and country authorities regarding the risk level associated with each category. The difference in ex-post probabilities across ratings captures the extent to which the model can discriminate between crisis and non-crisis outcomes.

The table below summarizes the accuracy of the new tools in terms of both misclassification rates and ex-post probabilities (see Annex V-VII for more details and information on the underlying samples):

- By construction, reported "missed crisis (L)" and "false alarm (H)" rates are close to 0.1.
- For the *near-term* tool, the (average) probability of stress following a high-risk signal, $P(S|H)$, is 0.40, while the (average) probability of stress following a low-risk signal, $P(S|L)$, is 0.02. For the *medium-term* tool, $P(S|H) = 0.43$ and $P(S|L) = 0.04$.

Importantly, while the table reports the average crisis probability of each risk bucket, there is some variation in risk scores within each bucket. The *lowest* stress probability associated with a high risk signal (corresponding to an index realization right at the cut-off between medium and high risk) is about 0.2 for both tools. The *highest* stress probability associated with a low-risk signal (corresponding to a realization at the cut-off between low and medium risk) is 0.09 for the near-term and <0.1 for the medium-term tool.

The table can also be used to calculate the sensitivity (probability of correctly identifying stress outcomes, $P(H|S)$) and specificity (probability of correctly identifying no-stress outcomes, $P(L|N)$) of the new tools. The near-term tool has sensitivity of 0.55 (80/146), and specificity of 0.73 (1034/1411). For the *medium-term* tool, specificity and sensitivity are 0.55 and 0.58 (6/11 and 51/88, respectively).[1]

Table. In-Sample Risk Signals and Outcomes for the Near- and Medium-Term Risk Tools

		Signal risk level				Probability of signal	
		low (L)	medium (M)	high (H)	total	conditional on outcome	Interpretation
			Near-term risk tool				
Outcome	Stress (S)	16	50	80	146	$P(L\|S) = \frac{16}{146} = 0.11$	Missed crisis (L)
	No stress (N)	1034	257	120	1411	$P(M\|S) = \frac{50}{146} = 0.34$	Missed crisis (M)
	total	1050	307	200	1557	$P(H\|N) = \frac{120}{1411} = 0.09$	False alarm (H)
Prob. of stress conditional on signal		$P(S\|L) = \frac{16}{1050} = 0.02$	$P(S\|M) = \frac{50}{307} = 0.16$	$P(S\|H) = \frac{80}{200} = 0.40$			
Prob. of no stress cond. on signal		$P(N\|L) = \frac{1034}{1050} = 0.98$	$P(N\|M) = \frac{257}{307} = 0.84$	$P(N\|H) = \frac{120}{200} = 0.60$			
			Medium-term risk tool				
Outcome	Stress (S)	2	3	6	11	$P(L\|S) = \frac{2}{11} = 0.18$	Missed crisis (L)
	No stress (N)	51	29	8	88	$P(M\|S) = \frac{3}{11} = 0.27$	Missed crisis (M)
	Outcomes total	53	32	14	99	$P(H\|N) = \frac{8}{88} = 0.09$	False alarm (H)
Prob. of stress conditional on signal		$P(S\|L) = \frac{2}{53} = 0.04$	$P(S\|M) = \frac{3}{32} = 0.09$	$P(S\|H) = \frac{6}{14} = 0.43$			
Prob. of no stress cond. on signal		$P(N\|L) = \frac{51}{53} = 0.96$	$P(N\|M) = \frac{29}{32} = 0.91$	$P(N\|H) = \frac{8}{14} = 0.57$			

[1]Ex-post probabilities and sensitivity/specificity are related through Bayes' Law. For example, $P(S|H) = P(H|S) * \frac{P(S)}{P(H)}$, where $P(S)$ denotes the prior (unconditional) probability of stress and $P(H)$ the probability of a high risk signal.

32. The three risk signals defined can also be interpreted in terms of ex-post crisis probabilities, that is, the probability that a stress event will occur conditional on a signal (Box 3). The *highest* stress probability associated with a *low-risk* signal (corresponding to a realization at the cut-off between low and medium risk) is below 10 percent for both the near-term and medium-term tool. The *lowest* stress probability associated with a *high-risk* signal (corresponding to an index realization right at the cut-off between medium and high risk) is about 20 percent for both tools. Hence, a high-risk signal does not necessarily mean that a stress event is likely to happen; but it means that the probability for it to happen is high enough (above 20 percent) for the risk to be taken very seriously.[19]

33. These tool-based "mechanical" risk signals would form the basis for an *overall* assessment of risks (low, moderate, high). Recognizing that, despite good predictive performance, standardized analytical tools do not allow to account for all factors that may be relevant in specific countries/circumstances, the two signals and the analytics at the longer-term horizons will only serve as starting points for teams, who can then add their judgment to arrive at final risk assessments for the near-, medium- and long-term horizons (also low, moderate, high). In addition, surveillance teams will be required to report if the debt/GDP ratio stabilizes under the baseline. A probabilistic debt sustainability assessment, based on a particular application of the tools (see paragraphs 72-73), will be required in Fund programs, but will remain optional in surveillance cases. Hence, the new framework delivers the necessary information to analyze all three concepts discussed in Box 2: risk of sovereign stress; failure of debt to stabilize under current policies (both indicating a need for policy action); and lack of debt sustainability (indicating a need for restructuring and/or exceptional bilateral support).

34. Judgment will inform the SRDSA at several stages of the analysis, constrained by the review process, and made transparent in SRDSA write-ups (see Appendix):

- At the stage where the mechanical signals are generated, country-specific characteristics can in some cases be captured through the consideration of additional variables (e.g. accounting for liquid assets in the GFN financeability module or counting them as part of international reserves in the logit (¶49)); triggered stress tests (¶63); and incorporation of long-term risks (¶70).
- At the stage where staff risk assessment is generated at each horizon, after taking the mechanical signal (for near- and medium term) as a starting point (¶33). Furthermore, in the case of medium-term assessments, stress tests (which can also be customized to introduce country-specific considerations) may also lead to staff assessments that differ from the mechanical signal. The general presumption is that deviations from the mechanical signal would not exceed one notch (¶68). In long-term risk assessments, optional modules and other considerations can inform judgment (¶70).

[19]Raising the threshold levels separating the medium- from the high-risk zones would imply that stress risks above 20 percent would still be considered "moderate". They would also imply a higher probability of missed stress events conditional on a low or moderate risk signal. At the current threshold levels, the average risk of a stress outcome conditional on a moderate risk signal is about 0.16 for the near term tool and 0.09 for the medium-term tool. These missed crisis probabilities would rise if the upper thresholds were to be raised.

- When generating an overall risk assessment based on the horizon-specific assessments, the presumption is that the overall risk assessments will remain within the range of staff risk assessments generated for each horizon.

The application of judgment would be constrained by the review process and would need to be justified appropriately in the SRDSA write-up, generally to account for country-specific considerations not appropriately captured by the statistical framework.[20]

35. While Covid-19 may raise some issues for the SRDSF, staff is of the view that these can be handled with some adjustments (Box 4). While a fuller evaluation will need to wait at least until the end of the 1-2 year prediction window of the near-term framework (that is, until mid-2022), a preliminary analysis indicates that the risk assessments derived from the proposed tools perform significantly better than the current framework in terms of predicting stress during in the first four months of 2020. Specifically, staff found a high correlation between the evolution of the proposed mechanical risk indices and observed sovereign rating downgrades. Nevertheless, staff is alert to possible challenges that might arise from the application of the new framework in the post-Covid context. Box 4 and Annex VI describe how these might be addressed, with additional detail to be fleshed out in the staff guidance note. The next review of the framework will offer the opportunity to recalibrate the tools, based on information that encompasses both the current crisis and its aftermath.

DEBT COVERAGE

36. Gross debt will remain the core concept in the SRDSF, but with an enhanced role for liquid assets. Government assets, particularly when liquid and foreign currency denominated, can have important implications for both solvency and liquidity, but differences in the coverage, quality and availability of such statistics argue against a net debt definition. However, the new analytical tools introduce specific customizations for liquid assets where data quality and materiality permit, alongside enhanced guidance where such customization is not possible (Annex II).

37. General Government (GG) debt, defined per GFSM 2014, will be the default institutional coverage. Since GG encapsulates all non-market government-controlled entities, it provides the most intuitive statistical concept of "government" and is already the most commonly used institutional coverage in the present MAC DSA. The *Fiscal Monitor* also uses this coverage level for reporting data on AE and EM debt stocks and fiscal flows. There is, thus, an obvious case to set GG as the explicit minimum benchmark coverage for MACs going forward, and to create incentives for reporting on this basis.

[20]As part of the Multipronged Approach to Debt Sustainability, a suite of tools has been developed, in addition to debt sustainability frameworks, to support debt analysis that could also help inform staff's judgment. These include dynamic general equilibrium models to analyze debt-investment-growth linkages and a framework to assess fiscal space.

Box 4. Implications of COVID-19 for the New Framework

Covid-19 has had a marked impact on the global macro-economic outlook, and therefore on prospects for debt sustainability. In the near-term, the crisis has led to sharp output contractions and generated substantial spending needs, causing a surge in fiscal deficits and public debt-to-GDP ratios. While output is expected to recover when the public health crisis abates, there is significant uncertainty around both the timing of this recovery and the extent to which output will be persistently affected, for example due to labor market hysteresis and lower capital accumulation.

The nature of the crisis also poses some challenges for the SRDSF. While the proposed new tools have provided useful and intuitive results during the initial phase of the crisis,[1] this box considers whether any adjustments are needed to ensure that the models continue to perform as intended in the post-shock period.

The SRDSF's realism tools would allow for some modification of inputs to ensure they remain relevant during the post-shock period, with residual issues to be discussed in SRDSA write-ups. Updates to baseline forecasts will continue to be based on country teams' analysis. However, given the importance of baseline forecasts for the proposed medium-term tools, some adjustments to the SRDSF's realism tools may also be warranted, to ensure that the expected reversion from extraordinarily high deficits and depressed output levels is not mistakenly classified as over-optimism. For example, when examining prospective fiscal adjustment, it would generally be appropriate to measure underlying primary balances excluding temporary crisis-related support.[2] Similarly, when assessing output projections, it may be useful to ignore temporary contractions and rebounds associated with the short-term supply effect of shutdowns. Where numerical estimates are unavailable, SRDSAs should discuss the plausible scale of these effects, and whether they are sufficient to account for any 'flags' raised by the realism tools. Finally, during the recovery phase an adjustment to the fanchart methodology (outlined in the next sections) is warranted to limit the number of instances where the optimism correction mechanism is incorrectly triggered (Annex VI).

The rise in medium-term uncertainty also deserves consideration, but on balance staff's view is that adjustments to the mechanical tools are not needed. The debt fanchart and GFN tools rely on historical data to calibrate forecast uncertainty and the likelihood and potential magnitude of shocks to the baseline. As such, they are not designed to reflect temporary rises in uncertainty. While this issue is partly mitigated by the tools' medium-term horizon, a concern remains that risks may be underestimated by this approach at present. At this stage, staff's view is that introducing a mechanical adjustment to capture increased uncertainty is not feasible, given the challenges associated with objective measurement. However, the new SRDSF template would allow teams to explore the impact of scaling up shocks, which could be used as a basis for applying judgment to the final risk ratings where appropriate.

Unless there is evidence that re-estimation would substantially improve the framework's out-of-sample performance, staff does not envisage re-estimating the various models until the next review. The coronavirus shock is likely to create a 'structural break' in many macro-economic time series, with movements of a magnitude that far exceed experience outside of events such as wars and natural disasters. While this break could affect the predictive power of the models, it also means that re-estimating the models on data that includes the shock would not necessarily result in a better model for the post-shock period. Regular updates would also be challenging from a communications perspective, as they may result in changes to risk signals despite broadly unchanged country-level fundamentals. Hence, it is envisaged that parameters would be 'frozen' until the next review of the framework. This said, a one-off re-estimation could be considered once the dynamics of the crisis and ultimate recovery are clearer, if this is likely to generate a clear improvement in the framework's out-of-sample performance.

---

[1]Initial results suggest tools have done a good job of identifying countries most likely to experience stress, particularly when updated to use post-coronavirus macro-economic projections.

[2]While the cyclically adjusted primary balance excludes the impact of the 'automatic stabilizers', it is still affected by discretionary measures, such as additional public health spending or temporary increases in transfers to affected households/workers.

38. Countries currently reporting on a central government (CG) or an incomplete GG basis will need to justify why the narrow coverage is appropriate.[21] Furthermore, countries with narrower coverage than GG will be subject to a mandatory contingent liability stress-test to capture omitted risk exposures. The stress-test will be calibrated based on countries' historical nonfinancial contingent liabilities per FAD database.[22]

39. In some circumstances, there may be a need to expand the coverage beyond GG to fully capture both sovereign risks and potential mitigants. A full or partial non-financial public sector (NFPS) coverage could be appropriate if it captures material fiscal risks from SOEs, aligns with national legislative requirements, or anchors policy discussions (including the production of official statistics at this level). A similar rationale could warrant inclusion of public banks involved in quasi-fiscal activities, so as to yield a full or partial consolidated public sector (CPS) concept.

40. The new framework will tackle the complex issues regarding central bank consolidation (whether done in the context of a proper CPS reporting or not). The framework will propose consolidations only in cases of central banks with large negative capital positions and/or where the country team considers the central bank to be involved in significant direct monetary financing of the budget and/or quasi-fiscal activities.[23] In other cases, the framework will incorporate mitigating factors from central bank holdings of government debt into its analysis without the use of consolidation (Annex II). In addition, some central bank liabilities could represent material risk and warrant a risk-based inclusion even when the central bank is not consolidated with the government. This includes central bank bilateral FX swap liabilities that do not represent normal central bank monetary or liquidity operations, or are not extinguishable by the central bank without actions detrimental to government debt levels (Annex II).

41. Fund credit disbursements will continue to be included as public debt for SRDSA purposes. Fund credit is legally an obligation of the member country and thus must always be included in a public debt sustainability analysis. There are also clear economic reasons for doing so: Fund credit can be disbursed either to finance the budget, or to raise FX reserves to a safe level. In the former case, Fund money merely substitutes for other types of budget financing; in the latter, it substitutes for other types of sovereign borrowing required to shore up reserves. Thus, in both cases, borrowing from the Fund by a country is conceptually a mirror image of the counterfactual change in the public debt ratio that would have happened in any case. This said, borrowing from the Fund would also reduce debt-related risks as assessed by the proposed framework, both

[21]A situation where narrower coverage would be appropriate would be when government functions are concentrated at the CG (for example, in small states). In cases where narrower coverage is due to data shortcomings, teams will be expected to outline in staff reports the authorities' plans to address these shortcomings, including any TA needs.

[22]The contingent liability stress-test would also be applied to GG or broader coverage countries with omitted fiscal risk exposures outside their chosen coverage.

[23]Such consolidation would imply that (i) central bank claims on the government are netted out *and* (ii) central bank debt liabilities (excluding currency and deposits held by residents) are added.

mechanically through higher reserves and lower rollover risk and due to the impact of IMF financing and conditionality on the macroeconomic baseline.[24]

42. The inclusion of contingent liabilities in the debt projections should reflect the likelihood of their materialization. Contingent liabilities are not generally expected to be included in GG debt. However, if teams are able to anticipate and estimate their materializations (such as uncalled government guarantees, legal settlements, bank recapitalization needs), these should be included in the debt projections (with a corresponding adjustment in the contingent liabilities stress-test). In particular—and consistent with GFSM principles—government guarantees should be included fully as public debt if there is a high likelihood of their materialization. The liabilities of government-controlled non-market SOEs or SPVs, which are recognized as part of the GG as per GFSM principles, should also be part of GG debt; if they are not, they should be added manually to the debt definition used in the SRDSF.[25]

43. The new framework will also require strengthened debt disclosures. Although the proposed approach creates incentives for a more consistent coverage of fiscal risks going forward, expecting a harmonization of coverage in the near-term is not a realistic goal. Accordingly, enhanced disclosure will be essential to support evenhandedness, and to help achieve greater harmonization over time. In particular, the SRDSA would be expected to include metadata on institutional and instrument coverage and the valuation method (nominal, face, or market value). Where available, this will be accompanied by a consolidation table showing the gross debt outstanding by level of government; the cross-holdings that are consolidated away; and the final consolidated debt position (Appendix I).[26]

44. Finally, enhanced reporting of debt profile vulnerabilities is proposed. Present reporting includes no information on governing law or the share of marketable debt; all FX debt is reported in just one bucket; and maturity and holder profiles are reported in just two buckets (short term versus the rest; resident versus non-resident).The new framework will include a breakdown of

[24]Under the proposed framework, IMF financing could mechanically affect the assessment of debt risks and sustainability through four channels: (1) the debt terms in the proposed logit model; (2) the FX reserves term in the logit model; (3) the debt level and index in the fanchart tool; (4) creditor composition, which enters the GFN tool. How these four effects impact the mechanical assessment depends on the counterfactual. If borrowing from the IMF allows the country to temporarily run larger deficits relative to the counterfactual, the net effect is ambiguous (as risks would mechanically rise via (1) and (3) but fall via (2) and (4)). If the counterfactual is similar or higher borrowing from the market as the country borrows from the IMF, the net effect is to lower debt risks. On top of these mechanical effects associated with IMF financing, the SRDSA will be affected through the assumed economic impact of Fund programs (resulting from both financing and conditionality), reflected in the macroeconomic baseline assumed by the framework. These should unambiguously improve sustainability, via external stabilization (contributing to lower public debt ratios in countries with FX-denominated debt), lower borrowing costs and possibly higher growth.

[25]An illustration of this principle is the use of an "augmented" debt measure for China which includes the debt of Local Government Financing Vehicles and other government funds that, although legally separate from the government, perform government functions. Similar criteria have been used in other countries (Belgium, Brazil, Russia, United Kingdom).

[26]Cross-holdings between government agencies have proven relevant in recent crisis cases (e.g., Argentina, Barbados) as a mitigating factor for debt burden and refinancing risks.

domestic/foreign law debt and marketable/non-marketable debt, the currency composition of FX debt, and additional information about the holder profile (foreign official, foreign private, domestic central bank, domestic commercial banks, and domestic nonbank) drawing on the Arslanalp-Tsuda database (in turn based on several cross-country databases).[27] As will be shown below, the inclusion of holder information is essential to allow an empirically grounded analysis of rollover risks.[28]

ENHANCED VISUAL REALISM TOOLS

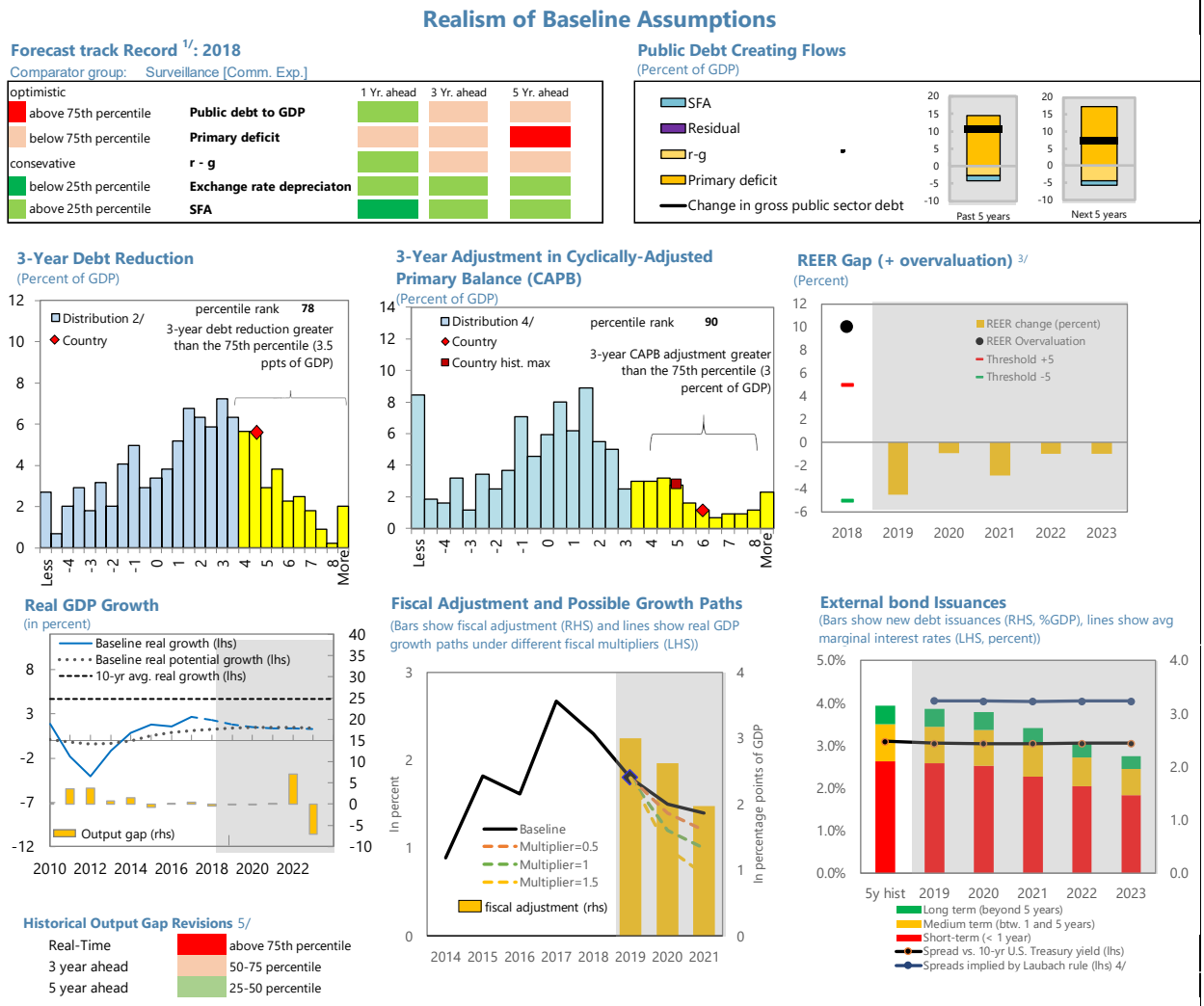
45. Enhanced realism tools are proposed to support more realistic baseline projections, a key input into medium-term risk analytics. Staff proposes to expand the existing realism tools (covering growth, inflation and primary balance) to encompass *all* debt drivers—including exchange rates, financing terms on external borrowing, stock-flow adjustments—and public debt itself. In addition, in light of evidence indicating systematic bias in output gap estimates (Annex I), the framework will introduce new tools to assess the realism of assumed fiscal multipliers and potential growth rates (see Annex III). Finally, to account for differences in forecast error distributions across commodity and non-commodity exporters, as well as surveillance vs. program cases, comparator buckets will be defined accordingly. This should allow more nuanced conclusions on the realism of baseline projections. Finally, adjustments are proposed to ensure that the tools maintain their relevance following the COVID-19 shock (Box 4).

46. As in the existing framework, the realism tools are intended to illuminate key assumptions underlying the projections. The tools will continue to put the baseline in context, in order to focus discussion on features of the macroeconomic framework that may differ from past and cross-country experience and suggest potential optimism or pessimism. Teams would be encouraged to utilize these tools early on as part of the iterative process of producing the baseline macro framework. In cases where tools flag differences from cross-country or a country's historical experience, these may well be explained by country-specific factors Such justifications should be clearly discussed in the SRDSA (see Appendix I for an illustrative example). In other cases, a re-examination, and possibly revision, of the macro projections may be warranted as part of the iterative process of producing a SRDSA, which includes the internal review process. The new suite of tools is summarized in the Figure 3 below and detailed in Annex III.

[27]See Arslanalp and Tsuda (2012) and Arslanalp and Tsuda (2014).

[28]An additional important piece of information, particularly at the time of a program request, is the holder profile of debt amortization. While this data is currently not compiled, it can be approximated with cooperation from country authorities (Annex VII). Provision of this information will also bring reporting requirements in the MAC DSA closer to those in the LIC DSF where teams already break down their financing assumptions into different creditor classes.

Figure 3. Proposed Realism Tools



Source: Fund staff.

Note: The tools in the top row (from left) analyze forecast record for debt drivers *vis-à-vis* a relevant comparator group (red cells indicating forecast optimism) and compare past and projected drivers of debt dynamics to check for large shifts. The two left charts in the middle row compare the projected three-year debt reduction and increase in the cyclically adjusted primary balance with the past distribution of such changes (changes corresponding to the yellow shaded portions of the distribution are unusual and may signal overoptimism). The REER gap chart indicates whether an initial overvaluation is expected to be unwound. Finally, charts in the bottom row check for closure of output gap by the end of projection period, output gap optimism based on the track record on past output ga revisions, check consistency between fiscal adjustment and growth assumptions using plausible multipliers, and assess the realism of new external issuance assumptions based on the history of issuance in the last five years and by comparing assumed spreads with those implied by the Laubach (2009) rule.

1/ Projections made in the spring WEO vintage.

2/ Data cover annual observations from 1990 to 2018 for MAC advanced and emerging economies. Percent of sample on vertical axis.

3/ Starting point reflects the team assessment of the initial overvaluation from EBA (or EBA-Lite).

4/ The Laubach (2009) rule is a linear rule assuming bond spreads increase by about 4 bps in response to a 1 ppt increase in the projected debt-to-GDP ratio.

5/ Calculated as the percentile rank of the country's output gap revisions (defined as the difference between real time/period ahead estimates and final estimates in the latest October WEO) in the total distribution of revisions across the data sample.

TOOL FOR NEAR-TERM RISK ANALYSIS

47. Staff proposes a multivariate logistic regression as the workhorse tool for near-term risk analysis. The proposed logit model combines current heatmap indicators with other relevant stress-drivers/mitigants to yield a single continuous probability of stress 1–2 years ahead. The specification includes a parsimonious set of 10 regressors, is intuitive and was the subject of extensive testing:

- Regressors are organized in five buckets (Table 4): *institutional quality*;[29] *stress history*;[30] *cyclical* (current account balance/GDP, 3-year real effective exchange rate appreciation, lagged credit/GDP gap); *debt burden and buffers* (change in public debt/GDP, public debt/revenue, FX public debt/GDP, and international reserves/GDP), and *global* (change in VIX). These buckets and the individual regressors were selected after consulting the relevant empirical literature, extensive statistical testing, and internal and external peer reviews (Annex V).[31]

Table 4. Specification of Multivariate Logit Model

Bucket	Regressor	Coeff.	Std. Coeff.
Institutional Quality		-1.073 ***	-0.377
Stress History		0.514 ***	0.1006
Cyclical	Current account balance/GDP	-0.024 **	-0.095
	REER (3-year change)	0.013 **	0.070
	Credit/GDP gap (t-1) (if + ve)	0.086 ***	0.258
Debt burden and buffers	Δ(Public debt/GDP)	0.052 ***	0.1182
	Public debt/revenue	0.002 ***	0.1213
	FX public debt/GDP	0.024 ***	0.1601
	International reserves/GDP	-0.034 ***	-0.2348
Global	ΔVIX	0.015 ***	0.1373
Number of Observations			1,579
LR chi2			246.70
Pseudo R2			0.25

Source: Fund staff calculations.
Note: Stars indicate statistical significance at the 1 percent (***) and 5 percent (**) levels. Standardized coefficients are scaled by variable standard deviations, thus providing a measure of relative importance (Long, 1997). For instance, the standardized coefficient for the FX public debt to GDP is about 1.4 times the magnitude of the coefficient for the change in public debt-to-GDP. This implies that ceteris paribus, a 1 standard deviation higher FX public debt-to-GDP ratio (about 16.8 percent of GDP, see Table AV.5) would have roughly the same effect on the stress probability as a 1.4 standard deviation increase in change in public debt-to-GDP (approximately 7.5 percent of GDP, see Table AV.5).

- The specification can be given a clear economic interpretation:
  - **Quality of institutions and stress history** capture countries' heterogeneity in a continuous way. Reflecting slow moving structural characteristics, these variables relate to a country's

[29]The "quality of institutions" index is derived as the simple average of the indicators *Government Effectiveness* and *Regulatory Quality* from the Worldwide Governance Indicators (WGI) database (Kaufmann *et al.*, 2010). The WGI are a summary measure of several hundred individual variables measuring perceptions of governance, drawn from 31 separate data sources constructed by 25 different organizations, ranging from think-tanks to governments, multilateral organizations and commercial firms.

[30]The stress history variable obtains as: a unit impulse is generated when stress occurs and then the impulse decays geometrically with AR-coefficient 0.9 (implying that the past is being "forgotten" at a rate of 10% per year).

[31]Consistent with Board advice at the midpoint stage, staff consulted with several external experts on the model, including: C. Reinhart and K. Rogoff (Harvard), E. Duggar (Moody's), L. Giorgianni (Tudor) and S. Pamies (EC). Comments received by an independent RES/ICD audit team, and reactions to these comments, are described in Annex V.

"debt carrying capacity" (Fournier and Fall, 2015; Fournier and Bétin, 2018),[32] similarly to the LIC DSF. Staff tested several candidates to proxy this concept and the two indicators selected by staff, particularly the quality of institutions index, significantly outperform other variables in terms of statistical significance, coefficient magnitude and robustness to different samples (Annex V).

- **Cyclical indicators** provide information on country-specific buildup of vulnerabilities and reflect the fact that the source of sovereign stress can be the external position, the financial sector, and/or a weakening of the fiscal position.[33]
- **Debt burden and buffer indicators** provide information on the vulnerabilities associated with debt level/dynamics/structure and on the risk-mitigating role of reserves (and liquid assets).[34]
- Finally, **global variables** are meant to proxy investor risk appetite (or push factors for international capital flows). They capture the fact that crises in countries exposed to international markets are often triggered at times when global financial conditions deteriorate. The change in the VIX proxies the change in global tolerance for risk-taking, which can trigger financing problems in weaker sovereigns.[35,36]

48. The estimated Logit Stress Probability (LSP) will be used to assign countries to low-, moderate- and high-risk of sovereign stress (the "near-term mechanical risk signal"), where the low- and high-risk cutoffs are calibrated to keep the rate of missed crises and false alarms at 10 percent, respectively. The corresponding stress probability cutoffs are 9 percent (at the threshold between the low and moderate risk signal) and 20.5 percent, respectively. Hence, a low risk signal means that the model estimates the probability of near-term sovereign stress at less than 9 percent, while a high risk signal means that it is estimated to be higher than 20.5 percent. This said, the model output includes the estimated logit stress probability in each individual case (see Appendix).

[32]In addition to empirical literature, the quality of institutions is a key indicator for the assessment of debt carrying capacity in the LIC DSF, it is considered a key credit factor by rating agencies (See Moody's Sovereign Ratings Methodology, S&P Sovereign Rating Methodology, and Fitch Sovereign Rating Criteria), and used by other institutions, such as the Bank of Canada, to assess eligibility and inform investment decisions in the management of foreign exchange reserves (See Muller and Bourque, 2017).

[33]The change in debt is classified as a debt-burden variable because it contributes to debt accumulation but can also be interpreted as a cyclical variable, flagging a deterioration in the fiscal position.

[34]In line with recent literature (Bassanetti, Cottarelli, and Presbitero, 2019), the dynamics of debt/GDP appear to be more relevant than its level in predicting near-term stress. Debt/revenue and FX-denominated public debt/GDP capture risks associated with high debt relative to revenues and currency mismatches, respectively.

[35]Rey (2015), and Obstfeld, Ostry and Qureshi (2019).

[36]Staff also aimed to incorporate systemic risk stemming from regional spillovers (not captured by the change in the VIX) and tested several variables, which however were insignificant with the exception of the share of currency union MACs in stress (Annex V). Acknowledging that the ongoing transformations in the governance of the eurozone may address these risks, the default setting of the logit model mutes this variable, which however can be "switched on" if country teams consider spillover risks within the CU a material risk.

49. Guidance will be provided to address some special cases. To reflect some countries' specificities that are relevant for the risk assessment but not adequately captured by the standard model (for instance, the use of liquid assets (Annex V) and the treatment of "safe havens"), specific direction will be provided in the guidance note.

50. The output of the near-term risk tool is designed to support policy messaging. Given that the tool issues a signal of sovereign stress 1–2 years ahead, which allows some time to adjust policies, the output is designed to empower this discussion. Specifically, the template will show the contributions of the five regressor buckets to the year-on-year increase in the LSP, which will support discussions on the policy actions to minimize the corresponding risk sources. Finally, to allow for cross-country comparisons, the estimated probability of stress will also be reported against contemporaneous values for peers (with teams having discretion over the choice of peer buckets). See Appendix I for a mock-up of the proposed output.

TOOLS FOR MEDIUM-TERM RISK ANALYSIS

51. Medium-term risk analysis is based on three modules, which analyze both solvency and medium-term rollover risks. First, baseline projections are used to produce an improved debt fanchart which is designed to be robust to baseline optimism and to assess prospects for debt/GDP stabilization in a probabilistic way. Second, a GFN module analyzes financing risks, taking into account the country's debtholder profile. Finally, a set of triggered/tailored stress-tests helps capture certain specific risks facing countries that are not fully covered by the fanchart and GFN tools.

A. Debt Fanchart

52. Staff propose a two-step procedure to derive an improved debt fanchart that will replace both the current fancharts and the standardized macro-fiscal stress-tests (see Annex VI for details on the methodology and metrics).

- **In the first step, the team's baseline is compared with a "historical fanchart",** independent from the team's projections, as a diagnostic tool for baseline realism.
- **The second step produces the final fanchart, based on the results of the first step:**

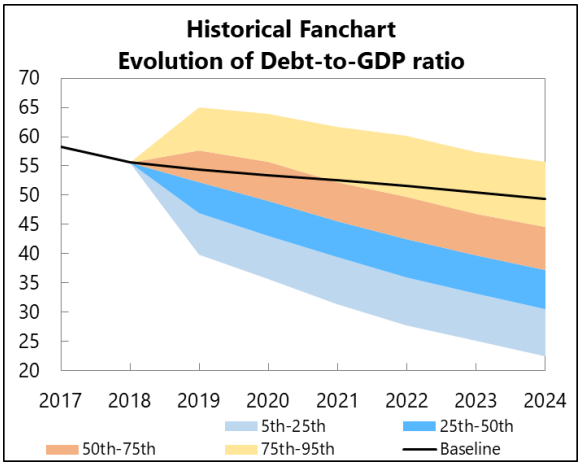
  i. If the first step *does not* reveal an optimism issue, the second step generates a "standard" fanchart (Figure 4, country 1) showing a symmetric balance of risks.

  ii. If the first step *reveals* an optimism issue, the second step generates a "realism adjusted" fanchart showing an asymmetric balance of risks (Figure 4, country 2).

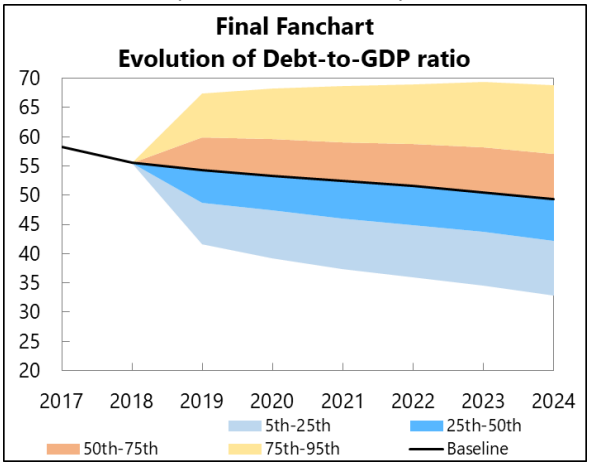
During the Covid-19 recovery phase (2021-22), staff proposes a modified version of the historical fanchart, partly dependent on the team's baseline, in order to limit the number of instances where the realism correction is incorrectly triggered (Annex VI).

Figure 4. Application of the New Fanchart Methodology

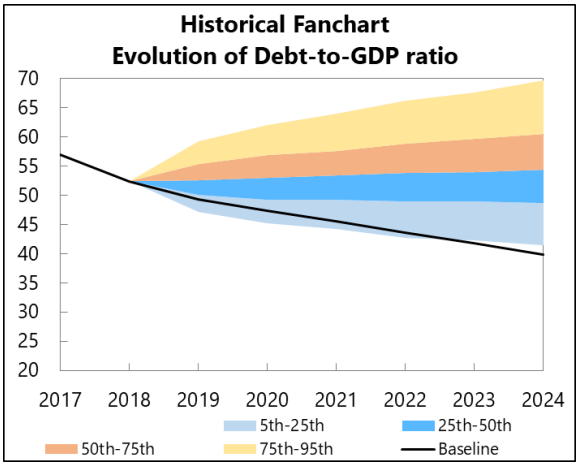
Country 1 (2019): Baseline above 20th percentile of historical fanchart



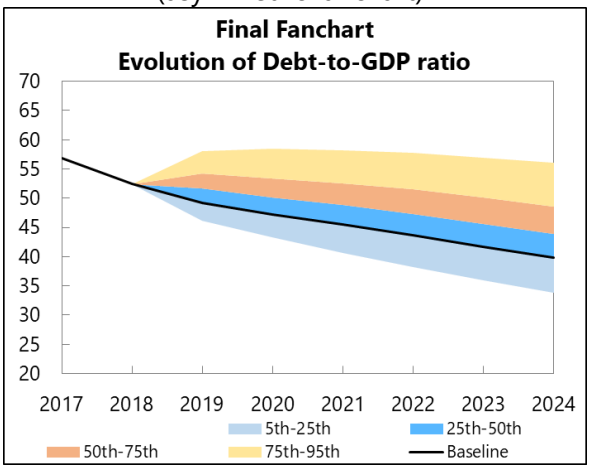
Country 1 (2019): Final fanchart obtained by applying de-meaned shocks to baseline (standard fanchart)



Country 2 (2019): Baseline below 20th percentile of historical fanchart



Country 2 (2019): Final fanchart obtained by applying skewed shocks to baseline (asymmetric fanchart)



Source: Fund staff estimates.

53. Like the logit analysis for the near-term, the new fanchart can be used to predict stress in the medium-term. To do so, staff examined the predictive power of several potential fanchart-based metrics over the period 2010–17. Three metrics stood out:

- **Probability that the debt does not stabilize in the medium-term,** defined as the probability that the projected baseline primary balance at t+5 will be lower than the balance required to stabilize t+5 debt. This metric expresses the probability that debt will not be on a stable path by the end of the projection horizon.

- **Debt level at t+5, controlling for debt-carrying capacity,** derived as the product of the debt level at t+5 multiplied by the institutional quality index[37] used in the near-term risk tool. Intuitively, this metric captures the risks posed by high levels of sovereign debt, while seeking to control for cross-country variation in the levels of debt at which such risks arise, (partly) dependent on debt-carrying capacity.

- **Fanchart width,** derived as the distance between the 5th and 95th percentiles of the debt distribution in the final projection year (t+5).[38] Intuitively, this metric captures the volatility of the country's debt drivers, and the potential for highly adverse debt realizations in the future, even if starting from a low level.

54. The three metrics are subsequently aggregated into a composite Debt Fanchart Index (or "DFI") that weights each metric by its predictive power and is used to classify countries into risk groups. Following a similar approach as for the logit model, three risk zones (low-, medium-, and high-) are derived by calibrating a low- and a high-risk threshold for the DFI associated with a 10 percent missed crisis and false alarm rates, respectively (Annex VI).

55. Guidance will specify some adjustments to the standard methodology in special cases, including: countries where the public sector holds large financial assets (Annex II), countries which have experienced obvious structural breaks, and countries that are close to or have recently reached a debt restructuring agreement. Additional details on these cases (and on countries that underwent restructuring in the past) can be found in Annex VI.

56. The output of debt fancharts will naturally support staff's policy messaging on debt stabilization, the level of debt, and the associated risks (Appendix I). The DFI value will be reported on a continuous scale of past DFI values associated with stress and non-stress episodes; comparisons with peers will also be included. Moreover, the output will be designed to bring out the relative contributions of the three metrics to the DFI. This will facilitate sharper discussions on the necessary policy adjustments.

B. Gross Financing Needs Module

57. The ability to consistently meet given levels of financing needs can differ materially across MACs. Countries with large domestic investor bases (like Japan and the United States) tend to be able to easily handle seemingly-large GFNs while economies with small financial sectors (like Argentina) can face serious challenges meeting even moderately-sized GFNs. Dependence on relatively stable versus relatively volatile investor classes also helps to explain different

[37]For this purpose, the institutional quality index is rescaled, so that zero corresponds to the highest institutional quality and one to the lowest.

[38]Metrics based on the width of the fanchart for stochastic DSA simulations are also used in determining the risk signal in the DSA methodology of the European Commission (2017), and debt sustainability approaches published by ECB staff (Bouabdallah *et al.*, 2017).

susceptibilities to liquidity pressures across countries. And finally, the presence of large government asset buffers can be an important mitigating factor to suppress risk.

58. Thus, the proposed GFN module analyzes how large the demand for additional financing might be in case of shocks and whether residual financing sources would have space to increase their government exposures. The analysis in this module is based on three steps (see Annex VII for further details on the methodology and metrics):

- *Projecting GFNs and financing*: As in the current MAC DSA, the starting point is GFN-to-GDP projections. However, teams will now be asked also to allocate domestic issuance among the central bank, commercial banks, and other private creditors; similarly, foreign issuance will be divided among foreign official and foreign private creditors.[39]

- *Modeling risk*: The centerpiece of the tool is a generalized stress scenario that includes adverse shocks in three key areas: (i) *macro-fiscal variables*, similar to the combined macro-fiscal shock scenario in existing MAC DSA; (ii) *maturities*, to illustrate implications of shortening of maturity of issuances in stressed conditions; and (iii) *access to external debt markets*, to examine the impact of a loss in foreign private investor appetite for a sovereign's debt.[40]

- *Analyzing residual financing*: The stress scenario described in the previous bullet creates a need for financing at a time when the pool of potential creditors has shrunk. It is assumed that the residual financing need is absorbed by domestic banks. However, customizations will be available to take account of relevant mitigating/aggravating factors like the availability of government liquid asset holdings and the domestic non-banking sector's role as risk dampener/amplifier.

59. Banks' ability to absorb additional financing needs is informed by three metrics.

- **Initial bank claims on the government expressed as a percentage of banking system assets.** Intuitively, this indicator gauges the degree to which the banking system is already exposed to the government and may be less able to *further* increase holdings.

- **The maximum cumulative change over the projection period.** This indicator gives a sense of the size and "financeability" of the financing demand that could be placed on the banking system if stress were to materialize.

[39]A default procedure for the disaggregation, which preserves the debtholder stock shares (*à la* Arslanalp-Tsuda database), would be available to teams. This procedure can also be used to approximate amortization by creditors.

[40]Specifically, the stress test includes a "holder shock" in which rollover rates of foreign investors drop to 67 percent, and these investors are unwilling to finance any additional new borrowing for a period of two years. Importantly, the risk signal derived from the test is not sensitive to how exactly the holder shock (or the remaining shocks) are defined, as they do not depend on the absolute level of the residual financing need according to the test but rather on the probability of false alarms and missed crises associated with this level, based on past stress events. For example, a more severe shock would lead to a higher level of residual financing but also to a higher probability of false alarms. Hence, the thresholds that determine whether the risk of debt distress is high, medium or low would be set lower than if the shock were defined to be less severe (see Annex VII).

- **The average level of GFN-to-GDP ratios over the projection horizons**. As in the current framework, the GFN level will continue to inform the analysis of risks, as there is evidence that this indicator has power to predict medium-term stress; however, the use of an average rather than year-by-year projections helps to separate a high one-year GFN level (less concerning) from consistently high GFNs (more concerning).

60. The three indicators are combined into an aggregate GFN Financeability Index (GFI), weighted by their explanatory power. As with the fanchart and logit model, the GFI is divided in three-risk zones (high, moderate, low), using thresholds associated with a 10 percent missed crisis and false alarm rates.

61. The tool incorporates some standardized customizations and guidance will be provided to better account for important factors, including: (i) the use of government financial assets to offset funding pressures (Annex II), (ii) the role of the domestic non-bank financial sector as a residual creditor, (iii) when non-bank financial intermediaries are a source of government funding risk, (iv) the timing of the onset of stress, and (v) the use of more granular information available to teams (see Annex VII for details).

62. The translation of the signal into a policy recommendation would focus on GFN levels and the financing mix for a given (baseline) GFN. On GFN levels, relevant policy recommendations in case of a moderate- or high-risk signal could be focused on some combination of fiscal adjustment or debt management (e.g. maturity extension) so that GFN levels and bank claims remain contained. The discussion could also include the holder structure and ways to finance the same baseline level of GFN but in less risky ways, depending on each country's circumstances.

C. Triggered Stress-Tests for Country-Specific Vulnerabilities

63. The medium-term toolkit is completed by triggered stress-tests that simulate debt and GFN paths under major specific shocks faced by sovereigns. These tools aim to enhance the discriminatory capacity of the framework by capturing an additional aspect of country heterogeneity and align it with the recommendations of the 2018 Interim Surveillance Review and 2018 Review of Conditionality, which called for better tailored scenario analysis and policy advice. Country teams will be encouraged to customize the tests' default parameters, based on historical and empirical evidence, to reflect country-specific considerations where appropriate.

64. The insights of the stress-tests complement those of the other two medium-term tools. The country's history on which the fanchart is based may not include some risks to which the country is susceptible in the future (e.g. climate change); and even where the fanchart captures past shocks, the country may have undertaken measures that would attenuate or aggravate the impact of such events (e.g. state-contingent debt instruments). Similarly, triggered stress-tests can complement the GFN generalized stress scenario as they correspond to a specific situation relevant for a country; and focus on the evolution of GFNs if such tail risks materialize.

65. Five specific vulnerabilities, common to groups of sovereigns, are proposed to be covered by the stress-tests (Table 5)**:** (i) contingent liabilities related to narrow coverage, (ii) banking crisis, (iii) natural disasters, (iv) commodity price shocks, and (v) REER shock. For some countries, more than one triggered test may apply; for others, none. Teams will also have the opportunity to create customized scenarios to model idiosyncratic risks, as in the current framework.

Table 5. Triggered Stress-Tests

Stress-test	Rationale	Trigger	Default Shock Design	User Customization
Contingent liabilities related to narrow coverage	Highlight risks from narrow debt coverage (and incentivize broader coverage)	Countries with shortfalls between their actual and appropriate debt coverage	Default set to zero if coverage is deemed appropriate, otherwise the average fiscal cost of observed nonfinancial contingent liabilities per FAD database	Tailor default parameters based on country-specific information
Banking crisis	Capture vulnerabilities of countries with oversized and vulnerable banking systems	Existing mechanical criteria and vulnerable countries per MCM FSI heatmaps	Fiscal cost implied from loss of 10 percent of financial assets (as in the existing framework), after assuming tier 1 capital	Tailor fiscal cost based on results of stress-tests implemented in financial sector assessment programs (FSAPs)
Natural disasters	Incorporate events outside country's history and/or where future impact could be different from the past	Countries at risk of natural disasters per past damages, exposure and vulnerability	One-off shock to public debt-to-GDP ratio (+7 percent of GDP) and real GDP growth (-5 ppts) representing the worst quintile impact of staff event analysis of past natural disasters	Tailor default parameters to account for recent structural or financial resilience initiatives (including any state-contingent instruments) and if baseline already assumes some event or not
Commodity price shock	Assess the impact of a negative ToT shock, for both commodity exporters and importers	Commodity exporters, and vulnerable commodity importers	Shock fiscal revenues (expenditures) to GDP ratios by -0.5 (+0.9) ppts for each 10 percent decrease (increase) in export (import) prices for commodity exporters (importers).[1] Additionally, shock baseline interest rate and growth	Tailor default parameters based on the likely impact of risks from countries' GRAMs and/or country-specific information (including by accounting for revenue collateralization and state-contingent instruments where relevant)
REER shock	Capture risks associated with large sudden currency movements, tied to the realism tool on REER.	Countries with high initial over-valuation, and REER change over t-t+5 that is insufficient to reduce overvaluation <5%.	A nominal one-off depreciation sufficient to close the country's overvaluation gap over the projection horizon.	Tailor size or duration of shock depending on country-specific information.

[1]To shock commodity prices, staff will use a price gap defined as the difference between the baseline commodity price and a+1 (-1) SD for importers (exporters) using RES commodity (fuel and non-fuel) price fancharts. The fiscal impacts reported come from Spatafora and Samake (2012). One may also consider shocking expenditures in exporters with large subsidies.

D. Medium-Term Overall Signal

66. The results from the debt fanchart and GFN module are aggregated into one overall medium-term risk signal that captures both solvency and liquidity risks. This makes sense for two reasons: first, to produce a more informative, and hence more accurate, medium-term signal; second, because neither of the underlying risk measures reflects pure solvency or liquidity risks. For example, while the fanchart-based tool can be thought of as capturing mostly solvency risk, the width of the fanchart reflects the past behavior of interest rates (among other debt drivers) and hence liquidity risk. Conversely, while the GFN signal can be interpreted as picking up mainly liquidity risk, it is also influenced by variables that are typically associated with solvency, such as the debt level.

67. The single Medium-Term Index (or "MTI") is created as the average of the DFI and GFI. The use of the simple average reflects the empirical finding that the DFI and GFI have roughly similar predictive capacities of stress (as measured by the respective AUCs) and thus should be equally weighted inside the composite index. This index is then divided into three risk zones, using the same approach as for the GFI and DFI, generating a single medium-term mechanical signal (high, moderate, or low). As described in ¶32 and Box 3, a low-risk signal can be interpreted as indicating a single-digit probability of sovereign stress and a high-risk signal as indicating a stress probability in excess of 20 percent.

68. In a further step, the result(s) of (a) triggered stress-test(s) can be used by the team to modify the MTI signal in the final medium-term risk assessment. In particular, when (i) the corresponding scenario has a high probability in the RAM; (ii) the scenario is judged to not be fully captured by the fanchart; and (iii) the scenario generates a debt trajectory above the 75th percentile of the fanchart, there is a presumption that the final MT assessment would be one notch worse than the mechanical signal. This said, the extent of the correction will remain judgment based.

LONG-TERM RISK TOOLS

69. To provide confidence around the debt and GFN levels/trajectories after 5-years, a 10-year horizon is proposed for debt and GFN projections. This horizon covers the Fund repayment period, and is essential to analyze debt restructuring cases, where longer repayment horizons are the norm. The longer horizon also allows teams to bring into the SRDSA important risks (and mitigating factors) in the 5 to 10-year window. This doesn't require a full financial program over 10 years, but merely a careful extrapolation of the main debt drivers. The extension is limited to the baseline; uncertainty around baseline projections (i.e., fancharts and stress-tests) is not modeled after 5 years. Guidance will elaborate how teams could extend their projections out to 10 years in a manner that is internally consistent and ensures evenhandedness.

70. A set of optional modules will be available to country teams to analyze key longer-term risks and inform staff's final long-term assessment. This will close an important gap in the current framework, which offers no analytical tools beyond the 5-year horizon. The use of the modules will be optional and will provide qualitative inputs to staff judgment. As such, there will be

no mechanical signals associated with the tools, but teams will be required to analyze key risks where these risks are identified (Table 6) and report their judgment if risks are low, moderate or high (reporting numerical estimates will be optional). The modules cover the following risk categories:[41]

- Population aging, including long-term sustainability risks stemming from pensions and social security benefits, and from health care costs;
- The scaling up/down of natural resource extraction to capture debt sustainability risks from projected changes in resource revenues beyond the five-year horizon;
- Large debt amortizations beyond the five-year horizon.

The parameterization, testing, and guidelines for interpreting the tools will be developed in the Guidance Note.

71. Staff also suggests that teams report on the long-run public finance consequences of climate change. The natural disasters stress-test affords an opportunity to capture risks associated with one-off climate events over the medium-term. However, global warming and rising sea-levels will have gradual and, cumulatively, much more profound effects over the long-term, A few countries might face existential threats and the need to rethink their economic models; others may need to undertake substantial spending for adaptation (e.g., changing crop varieties and building higher dikes to guard against sea levels) and mitigation. Staff suggests that teams working on the first group of countries, and a subset of high-vulnerability countries from the second group should discuss the potential implications for key macroeconomic variables such as growth and public spending. Where feasible, the team could derive debt ratio implications of lower potential growth and higher climate change spending out over a period of 30 years.

[41]Long-term risks not covered by the modules would be reported by the teams in the staff commentary to the long-term risk section of the SRDSA. Where available, the results from FAD's public balance sheet tool could also be reported and incorporated in the SRDSA.

Table 6. Tools for Assessing Risks from Long-Term Factors

Module	Rationale	Trigger	Output
Population			
(i) pension and social security benefits	Unfunded liabilities for pension or social security programs could lead to higher indebtedness as demographic change proceeds.	Countries with a deficit in the social security program, either currently or in the next 10 years; or significant population aging.	Future pension fund reserves and/or yearly pension fund liabilities in periods t+6 to t+16, benchmarked against criteria.
(ii) pressures from health costs	Health care expenditures are also likely to be a cause of pressure from both aging as well as from non-demographic factors such as excess cost growth.	Countries with sizable public health expenditures that are likely to grow rapidly from demographic change or excess cost growth.	PV of health expenditures, net of health revenues and/or path for health expenditures under the various paths for excess cost growth in periods t+6 to t+10.
Scaling up/down of natural resources	This module considers the volume dimension of natural resource revenues. Future extraction volumes may be different from those seen in the past, either because of exhaustion, political choices, or projected scaling up.	Countries where extraction volumes over t+6 to t+15 deviate by more than one standard deviation from the historical average (calculated over the last 10 years).	Projected change in debt and/or primary balance at t+10 compared to t+5, assumption of constant natural resource revenues and non-interest expenditures.
Large debt amortizations	Large (above historic levels) amortizations may create liquidity pressure, if not timely addressed through debt management operations and/or	Debt amortizations in t+6 to t+25 above historical average over the last ten years plus one standard deviation.	Average yearly deposit over the period t+1 to maturity needed to meet the obligation, expressed in millions and in percent of GDP.
Climate change	To incorporate long-term macroeconomic implications of climate change.	Countries with existential or high vulnerability to climate change per exposure, susceptibility and adaptive capacity. 1/	Discuss impact on key macro-fiscal variables (such as potential growth, spending) and, where feasible, draw out 30-year implications for the debt, leveraging relevant analysis from Climate Change Policy Assessments (CCPA) or other country-specific work where relevant.

1/ The determination of high-risk countries could be informed by the Climate Change Exposure Index used in IMF (2016a; 2016b), and the World Risk Index in World Risk Report (2018). As of end-2018, these two indices identify 21 countries: Angola, Antigua and Barbuda, Bahamas, Belize, Chile, Costa Rica, Dominica, El Salvador, Guatemala, Fiji, Indonesia, Jamaica, Japan, Mauritius, Montenegro, Philippines, Swaziland, St Kitts and Nevis, Suriname, Trinidad and Tobago, and Vietnam.

DERIVING SUSTAINABILITY ASSESSMENTS

72. The proposed new tools, with slight adjustments, will be used to produce the probabilistic debt sustainability assessments required in Fund-supported programs (including under the Fund's Exceptional Access framework). As discussed earlier, the Board-endorsed definition is anchored in the notion of debt stabilization under feasible policies, with low rollover risk. Accordingly, under staff's proposed approach:

- Medium-term debt stabilization prospects will be quantified using the **debt fanchart** composite index;

- Assessments of rollover risk will be informed by the **GFN module,** where the analysis would take all components of program financing (including prospective Fund disbursements) into account;

- **A crisis prediction model calibrated on past episodes of unsustainable debt—rather than on sovereign stress—**will help to predict crises associated with unsustainable debt (including defaults and debt restructurings) rather than just stress events.

73. Outputs from the three modules will be aggregated, ultimately leading to a three-way mechanical signal on debt sustainability (sustainable with high probability; sustainable, but not with high probability; not sustainable). Team judgment and a robust review process will complement the mechanical signals, leading to a bottom-line assessment on debt sustainability. In general, a pattern of high risks from the tools would suggest concerns about sustainability, while a pattern of low risks would suggest sustainability with high probability. Anything in between would suggest sustainability, but not with high probability. The location of the three zones is calibrated based on past instances of unsustainable debt, aiming to limit the number of false alarms and missed crises, but also ensuring that (1) the signal "not sustainable" is associated with a probability of an unsustainable event of more than 50 percent; (2) the signal "sustainable with high probability" is associated with a probability of an unsustainable event of less than 20 percent. Since the methodology by which the Fund makes its sustainability (and thus lending) decisions is potentially market-sensitive, the precise aggregation method and the index cutoffs determining the three signals will remain confidential.

74. In *surveillance* cases, sustainability assessments would not be required. However, if staff chooses to provide such an assessment, they would need to generate additional scenarios to anchor it. In the absence of an active (program) baseline, to conclude that debt is sustainable, staff will either have to show that the sustainability metrics are satisfied using the assumed baseline; or are satisfied under an alternative scenario (developed by staff) that is politically and economically feasible. To conclude that debt is not sustainable, staff will need to show that the sustainability metrics are not satisfied even by policies staff considers "at the frontier" of feasibility.

75. In debt restructurings, the new tools can be readily used to set targets consistent with restoring sustainability. Because the medium-term tools link naturally with the sustainability definition, they do not require major modification. However, it would be generally appropriate to use a longer (10-year) horizon, which is common in restructurings. GFN targets, derived from the GFN module, are a convenient starting point to verify that the resulting financing needs after the restructuring are indeed manageable, including under adverse circumstances. Subsequently, a post-restructuring debt trajectory can be readily derived from the new debt structure that would attain the GFN targets. The new debt trajectory would then be analyzed through the debt fanchart module to assure that there is an appropriately high probability of debt stabilization. If this is the case, the debt trajectory can be used to set an additional target for the future debt level (along with the GFNs); while in the event that this is not the case, the debt relief envelope would need to be adjusted accordingly until both the GFN module and the fanchart module signal sustainability. The approach would maintain a role for judgment as well as for complementary targets to address

specific country vulnerabilities, when warranted. The targets derived according to this approach are in line with the Fund's usual role to define the needed envelope of debt relief to restore sustainability.[42] However, specific restructuring decisions will remain the responsibility of country authorities, in consultation with their legal and financial advisors.

REPORTING REQUIREMENTS

76. A sovereign risk analysis should be undertaken for all members both in a surveillance and a program context, while sustainability assessments are needed for arrangements involving GRA resources (including precautionary arrangements) as well as for the PCI. For non-program countries, a sovereign risk analysis is needed at the time of the Article IV consultation. For countries with Fund arrangements involving GRA resources, both a sovereign risk and a debt sustainability analysis need to be performed. The latter should be done at the time of program approval and subsequently once a year (unless developments in the outlook for public debt warrant a more frequent analysis), except for exceptional access cases, which require an updated DSA (with three-zone sustainability assessment) in every program review.

77. The information regularly reported in Staff Reports will depend on whether the analysis is performed in a program or surveillance context and the degree of disclosure desired by the Board.

78. In a program context, staff reports in which a DSA is required will need to include a sovereign risk analysis and potentially a three-zone sustainability assessment.

- With respect to risk of sovereign stress, staff reports will contain the full range of outputs for the medium and long-term, but not for the near-term because the latter is not informative (as a near-term risk event is already occurring). Reporting medium and long-term risk assessments beyond the program horizon is useful as medium and long-term risks are targeted and mitigated by the ongoing Fund-supported program. An overall risk assessment, which synthesizes the team's assessment of risks across horizons, will also be reported. In precautionary arrangements, reporting will be based on the status of the program (Box 5).

- With respect to debt sustainability assessments, two options for disclosure to the Board and/or the public could be considered:

  a. No change with respect to the current practice, in which a three-zone assessment is included in staff reports in exceptional access cases but not in normal access cases.

  b. Disclosure to the Board in both normal and exceptional access cases, but disclosure to the public only in exceptional access cases. Compared to the standard two-zone assessment

[42]In particular, this process is closely linked to Article V, Section 3(a) of the Articles of Agreement, which stipulates the requirement for the Fund to develop policies on the use of its general resources "that will assist its members to solve their balance of payments problems…and that will establish adequate safeguards for the temporary uses of the general resources of the Fund."

(sustainable/unsustainable), this would give the Board a more nuanced assessment of debt sustainability, and may help inform discussions around the balance between adjustment and financing and the appropriate level of access (for instance, the Board might accept a higher access level if it is advised that debt is assessed to be sustainable with a high probability). At the same time, it would preclude potentially adverse market reactions to the disclosure that a country's debt is sustainable but not with high probability (for example, if this is read as disqualifying the country from exceptional access in the future).

Importantly, option b would require a modification to the Transparency Policy (TP), as it would require a blanket provision permitting deletion of a particular set of information contained in normal-access staff reports, rather than application of the current rules permitting deletions only upon a case-by-case determination of market sensitivity.

79. In a surveillance context, undertaking and publishing a sustainability assessment would be optional, while two levels of disclosure of the sovereign risk analysis could be considered.

a. Full disclosure: The outputs of the risk assessment framework would be fully disclosed in staff reports. This would have several benefits, including protecting staff from political pressures; assuring evenhandedness and transparency (which external stakeholders have flagged as the existing framework's biggest shortcoming); and safeguarding the Fund's reputation by flagging risks in a timely manner. Furthermore, full disclosure would bring the Fund's debt sustainability framework for MACs in line with the publication requirements of the LIC DSF, which publishes both mechanical signals and staff's overall assessment. In line with the Transparency Policy, deletion of market sensitive signals would still be possible on a case by case basis.

b. Full disclosure to the Board, but limited disclosure to the public. For example, published staff reports could omit the near-term risk signal and assessment, where market sensitivity concerns are arguably higher than for the medium-term signals. Such an approach would avoid the need for case-by-case deletions (which might carry a risk of sending a signal in themselves). However, for similar reasons as option b in the previous paragraph, this option would require a modification to the Transparency Policy (TP).

Should the Board choose option b as described in paragraphs 78 or 79 (or both), staff will follow this paper with a targeted amendment to the Transparency Policy on a stand-alone, lapse-of-time basis.

80. In the event that the Board were to choose option b as described in paragraphs 78 or 79 with respect to disclosure, staff proposes that the level of disclosure to the public be reconsidered after 12 months. Following this period, the Board and country authorities will have gained familiarity with the new framework, and may be in a better position to decide whether full disclosure is warranted.

81. In both surveillance and program contexts, the SRDSA included in the staff report will be sharper than is the case today, reducing the need for a lengthy write-up. Box 6 discusses how results will be presented and Appendix I shows a mock SRDSA under a 'full disclosure' setting.

Box 5. Reporting in Precautionary Arrangements

Precautionary arrangements do not envisage drawings except in response to shocks. They do not imply that the member is in stress; rather, they aim to shield the member from incurring stress due to external factors. Hence, *risk assessments* reported for precautionary arrangements (including FCLs, PLLs, and precautionary programs) are informed by the baseline scenario (where no drawing occurs).

Sustainability assessments for precautionary arrangements, on the other hand, would be informed both by the baseline and, when appropriate, by an adverse scenario or stress scenario (where full drawing occurs). In particular, the latter would be appropriate in three settings:

- In exceptional access cases;
- if shocks that may trigger a drawing are not adequately captured by the medium-term (fan chart and GFN) tools; or
- when review departments have doubts about the realism of the baseline that cannot be resolved through discussions with the country team.

In such cases, running the sustainability assessment on a full-draw scenario (including Fund credit as part of public debt) would provide additional assurances that debt is sustainable in the event that Fund credit is drawn.

Program design

When a precautionary arrangement is under consideration, the near-term risk assessment can help inform program design: if near-term risk under the baseline is assessed to be high before/at the approval of an arrangement, a precautionary arrangement may not be the appropriate vehicle. High near-term risk signals a high probability that the country will enter stress and need to draw in the next two years. In such cases – and particularly if the medium-term tools also signal high risk – a financing arrangement may be more appropriate. This approach can also be adopted when the DSA at the final review of a precautionary program is used to consider a successor arrangement.

Near-term risk assessment

For approved precautionary arrangements, the near-term sovereign stress signal and assessment will be produced as long as the country does not draw on the IMF arrangement. Once the country makes a purchase, it is considered to be in stress, and no near-term stress signal would be produced, in line with the publication policy for regular drawing programs. If a drawing were imminent, the near-term signal/assessment would not be published, to avoid accelerating the drawing.

Medium-term risk assessment

Medium-term risk would be assessed based on the baseline scenario (where no drawing on the program occurs), with Fund financing not included in financing assumptions.

Box 6. New Reporting Format

Under the proposed SRDSF, published outputs will be sharper, virtually eliminating the need for lengthy DSA write-ups (see Appendix I for a mock SRDSA). In a normal case, the published DSA could be no more than 7–8 pages (shorter than the 10-page average at present). Moreover, to make it punchier and reader-friendly, the SRDSA would comprise the following standardized tables and charts, accompanied with staff commentaries:

i. An upfront table summarizing the risk and sustainability assessments and if and how judgment has been applied. This would be followed by a punchy SRDSA chapeau.

ii. A set of tables summarizing debt coverage disclosures and related information; and a figure on the composition of public debt by currency, holder, legal basis, level of public sector, and maturity.

iii. A table reporting debt and GFN projections and debt drivers, with the staff commentary highlighting any important trends or features of the projections.

iv. A figure reporting results of the realism tools with the staff commentary explaining any red flags;

v. A page each for the outputs of the near-, medium-, and long-term risk analyses and the derivation of the mechanical risk signals. Cross-country comparisons and stress-test results would also be reported here.

NEXT STEPS: IMPLEMENTATION TIMELINE AND ENGAGEMENT STRATEGY

82. The implementation of the new framework will be accompanied by close engagement with member country authorities, IMF country teams, and external stakeholders:

- **Developing materials to support implementation (expected to be available by the second half of 2021).** This will include preparation of the guidance note, software and Excel files underpinning the new framework, and settling remaining issues, including how to adapt some tools to country-specific circumstances, how to ensure evenhandedness across countries, and public communication. During this phase, there would be early engagement with a subset of country teams to test the new tools and templates, running them in parallel with the current framework to inform the staff guidance note and the design of the template.

- **Close engagement with country authorities and debt management offices** (Box 7). This step will allow staff to illustrate the features of the new framework ahead of the rollout so that country authorities can become familiar with it, a priority that management has also emphasized. In this context, after the formal Board approval of the new framework, staff will reach out to country officials, including during the Spring Meetings. In addition, outreach to debt management offices at an early stage will be critical in this step to ensure that the new files are easy to use, intuitive, and free from bugs.

- **Staff will also engage with external stakeholders that follow IMF debt policies and debt sustainability assessments closely**, including civil society organizations, think tanks and private creditors—for instance, during the 2021 Spring and Annual Meetings.

83. The current framework will continue to be used between Board approval and the rollout of the new framework, expected for Q4 2021 or Q1 2022. In cases where data availability issues pose challenges for the implementation of the new framework, an implementation team from SPR will provide support to country teams prior to the rollout date. As a last resort, this could include adjusting some tools of the new framework to accommodate data limitations.

84. In program cases, the transition between the old and the new framework will be carefully managed to ensure consistency. Specifically:

- For programs approved prior to the rollout date but expiring after the rollout date, SPR and area departments will work together to examine any differences between the results of the two frameworks. Staff will make necessary adjustments to arbitrate bottom-lines accordingly. Adjustments through judgment under the current framework would be appropriate if the new framework is deemed valid given new analyses or compelling conditions, while any issues with the tools or application of the new framework could be addressed in the guidance note.

- Programs that expire prior to the rollout date would not be affected by differences between the current and new framework. However, insofar as a successor program is expected to be approved after the rollout date, the successor program would be subject to assessment under the new framework and staff would need to manage communications with the authorities accordingly.

To the extent that experiences during the transition period raise concerns about a limitation of the new framework, the guidance note will be adjusted to address these concerns.

85. While the framework introduces new data requirements, the burden on Fund resources and member country officials will be contained by automation of data sources and gradual implementation supported by Fund TA, and may be offset by the new framework's substantially reduced, focused writeup requirements. Most of the needed data is already required to run the current framework; and where additional data is needed, it will be linked to existing centralized datasets as far as possible. The review team has already made significant use of these sources to run the new core elements of the framework on nearly every MAC, confirming the feasibility of operating the new framework. However, a limited set of new debt profile data will need to be collected by country teams in consultation with country authorities (Annex VIII). An implementation team from SPR will provide support to country teams when transitioning to the new framework, a practice that proved useful during the adoption of the LIC DSF. Staff will also work with authorities—including through TA—to prepare them for the additional data requirements. Where expanding debt coverage takes time, the contingent liability stress-test can be used to analyze risks while statistical capacity is strengthened. To address any special challenges created by the new framework in frontier LICs or recent PRGT graduates, staff will develop transitional solutions for these countries in the guidance note (for instance, the use of parameters calibrated on peer countries, or simplifications of the standard tools).

86. The rollout of the new framework will be preceded and complemented by comprehensive training—both internally and externally. The internal training will introduce the new framework to staff and leave them able to explain it to their counterpart country authorities. Additionally, staff will prepare training material for country officials, and potentially external stakeholders, including through onsite and online courses and seminars. In this respect, the successful rollout of the LIC DSF will serve as a model.

Box 7. Engagement with Country Authorities

Along with the development and testing, Fund staff will engage with country authorities in several ways. The outreach strategy will be informed by the experience of the successful rollout of the LIC DSF in 2017. The key elements of the strategy include:

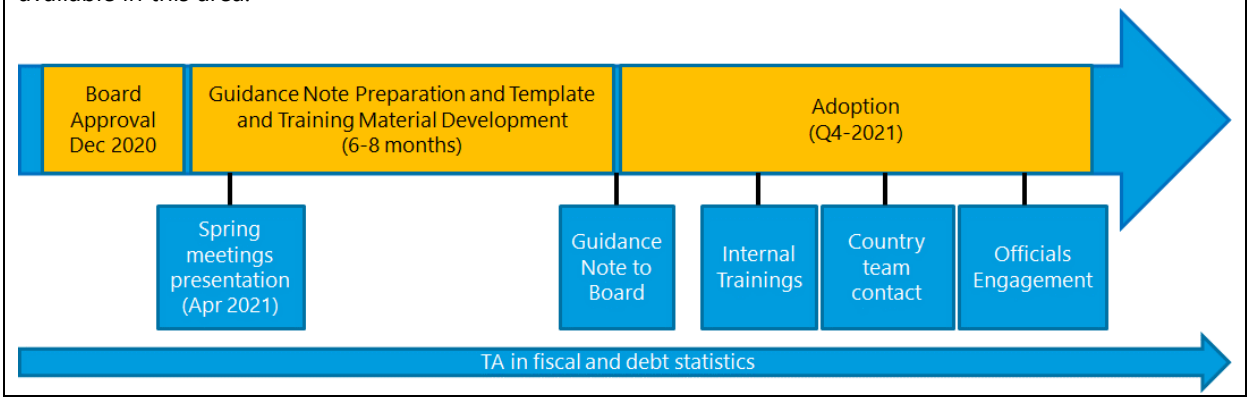
Guidance note: After the framework is formally approved, staff will begin developing the Guidance Note. It will cover all elements of the new framework and include critical information on interpreting the models correctly, handling special country cases, evaluating conflicting signals, and identifying key situations where staff judgment would be essential. Given the importance of these issues, staff will conduct additional outreach to Executive Directors and submit the completed draft for the Board's information. Under the usual process for Guidance Notes, Executive Directors will have 2 weeks to call a meeting to discuss it; thereafter the guidance note will become operational. This step will likely take [6-8] months.

Seminars at the Spring Meetings: Staff will organize events during the Meetings as an early opportunity to present the new tools, sensitize country officials to changes in the new frameworks, and take any questions stakeholders may have.

Contact with country teams: Upon rollout, authorities' close contact with country teams will be a natural way to ask technical questions, understand the tools, and apply them to their country. Internal training will ensure that country teams are fully abreast of the new framework.

Classes for country officials: Training courses, such as the ones that are currently done under the Debt Management Facility (DMF), will also be developed and delivered as the framework goes live. [Additional training events beyond the DMF may also be possible, though staff may need to identify financing.] Moreover, all training materials (e.g. presentations, an illustrative example for a fictitious country, and a user's manual) would also be posted to the Fund's public website and be available to all country officials.

Other capacity development: With the expectation that General Government will be the standard coverage for public debt in the MAC DSA, some countries will require support to develop the needed statistical capacity. The expectation is that this will be gradual, but consistent, and Fund TA will continue to be made available in this area.



ISSUES FOR DISCUSSION

- Do Directors support the continued application of the existing definition of debt sustainability (as previously adopted by the Board)?
- Do Directors endorse naming the proposed framework "Sovereign risk and debt sustainability framework" to capture the full range of its analysis?
- Do Directors agree with the enhanced debt coverage and the proposed 10-year horizon?
- Do Directors agree with the proposed realism tools and realism adjustments?
- Do Directors agree with the proposed horizon-based approach?
- Do Directors support the use of standardized tools to provide mechanical risk signals at each horizon, with judgment added to account for country-specific information not well captured by the tools?
- Do Directors agree that sovereign stress signals and assessments should be included in reports for program cases, except for near-term sovereign stress signals, which would only be included in reports for precautionary programs?
- With respect to the disclosure of three-zone debt sustainability assessments in program cases, which of the two options presented in paragraph 78 do Directors favor:
    a. continuing the current practice of including three-zone-assessments only in staff reports on exceptional access cases, or
    b. disclosure of three-zone assessments to the Board in both normal and exceptional access cases but to the public only in exceptional access cases, requiring a change in the Transparency Policy?
- With respect to the publication of sovereign stress signals in surveillance reports, which of the two options presented in paragraph 79 do Directors favor:
    a. full disclosure to both the Board and the public, or
    b. full disclosure to the Board but limited disclosure to the public, requiring a change in the Transparency Policy?
- Do Directors support the proposed use of the new tools to guide debt restructurings?
- Do Directors endorse the proposed DSA reporting format?
- Do Directors support the proposed timeline for the implementation of the framework?

References

Arslanalp, S., and Tsuda, T., 2012, "Tracking Global Demand for Advanced Economy Sovereign Debt," IMF Working Papers No. 12/284. (Washington: International Monetary Fund).

Arslanalp, S., and Tsuda, T., 2014. "Tracking Global Demand for Emerging Market Sovereign Debt," IMF Working Papers No. 14/39. (Washington: International Monetary Fund).

Bassanetti, A., Cottarelli, C., and Presbitero A., 2019, "Lost and Found: Market Access and Public Debt Dynamics," *Oxford Economic Papers*, Volume 71, Issue 2, April 2019, Pages 445–71.

Berg, M.A., Berkes, M.E., Pattillo, M.C.A., Presbitero, A. and Yakhshilikov, M.Y., 2014. "Assessing Bias and Accuracy in the World Bank-IMF's Debt Sustainability Framework for Low-Income Countries," (No. 14-48). (Washington: International Monetary Fund).

Bouabdallah, O., Checherita-Westphal, C., Warmedinger, T., de Stefani, R., Drudi, F., Setzer, R., Westphal. A., 2017, "Debt Sustainability Analysis for Euro Area Sovereigns: A Methodological Framework." *ECB Occasional Paper Series*, No. 185.

European Commission, 2016, *Debt Sustainability Monitor*, January 2017. (Brussels: European Commission)

Fall, F., and Fournier, J.M., 2015, "Macroeconomic uncertainties, prudent debt targets and fiscal rules," OECD Economics Department Working Papers No. 1230. (Paris: Organization for Economic Co-operation and Development).

Fitch Ratings: Sovereign Rating Criteria, March 27, 2020

Fournier, J.M., and Bétin, M., 2018, "Limits to Government Debt Sustainability in Middle-Income Countries," OECD Economics Department Working Papers No. 1493. (Paris: Organization for Economic Co-operation and Development).

International Monetary Fund, 2002, "Assessing Sustainability," (Washington: International Monetary Fund).

International Monetary Fund, 2005, "Information Note on Modifications to the Fund's Debt Sustainability Assessment Framework for Market Access Countries," (Washington: International Monetary Fund).

International Monetary Fund, 2011, "Modernizing the Framework for Fiscal Policy and Public Debt Sustainability Analysis." (Washington: International Monetary Fund).

International Monetary Fund, 2013, *Staff Guidance Note for Public Debt Sustainability Analysis in Market Access Countries*. (Washington).

International Monetary Fund, 2016a, "Small States' Resilience to Natural Disasters and Climate Change - Role for the IMF." (Washington: International Monetary Fund).

International Monetary Fund, 2016b, "After Paris: Fiscal, Macroeconomic, and Financial Implications of Climate Change," IMF Staff Discussion Note 16/01. (Washington: International Monetary Fund).

International Monetary Fund, 2017, "Guidance Note on the Bank-Fund Sustainability Framework for Low Income Countries." (Washington: International Monetary Fund).

Kaufmann, D., Kraay, A., and Mastruzzi, M., 2010, "The Worldwide Governance Indicators: Methodology and Analytical Issues," *World Bank Policy Research Working Paper* No. 5430. (Washington: The World Bank).

Laubach, T., 2009, "New Evidence on the Interest Rate Effects of Budget Deficits and Debt," *Journal of the European Economic Association*, MIT Press, vol. 7(4), pages 858-85, June.

Long, S., *1997*, "Regression Models for Categorical and Limited Dependent Variables," *Advanced Quantitative Techniques in the Social Sciences* Number 7. Sage Publications.

Moody's Investor Services. Rating Methodology: Sovereign Bond ratings, 2018

Muller, P., Bourque J., 2017 "Methodology for Assigning Credit Ratings to Sovereigns," Bank of Canada Staff Discussion Paper 2017-7. (Ottawa: Bank of Canada).

Obstfeld, M., Ostry, J.D. and Qureshi, M.S., 2019. "A tie that binds: Revisiting the trilemma in emerging market economies," *Review of Economics and Statistics*, 101(2), pp.279-293.

Rey, H., 2015. "Dilemma not trilemma: the global financial cycle and monetary policy independence," (No. w21162). National Bureau of Economic Research.

Samaké, I., and Spatafora N., 2012, "Commodity Price Shocks and Fiscal Outcomes," IMF Working Papers No. 12/112, (Washington: International Monetary Fund).

S&P Global: Sovereign Rating Methodology, December 18, 2017

World Economic Forum, 2018, "The Global Risks Report 2018," (Geneva: World Economic Forum)

Appendix. Ruritania—Sovereign Stress and Debt Sustainability Assessment[1]

Risk of Sovereign Stress			
	Mechanical signal	***Final assessment***	***Comments***
Overall		Moderate	Staff's assessment of "moderate" reflects the mitigating role of a large liquid asset buffer and the impact of scaling-up of natural resources.
Near term[1/]	High	Moderate	Near-term mechanical risk signal was "high". Staff's assessment of "moderate" reflects the mitigating role of a large liquid asset buffer.
Medium term	High	Moderate	Staff assesses medium-term risks as "Moderate" reflecting the mitigating role of a large liquid asset buffer and despite a mechanical medium-term signal that indicates "high" risk, largely attributable to the debt fanchart's width. Additionally, the commodity price stress-test did not generate a debt trajectory above the 75th percentile of the debt fanchart.
GFN:	Moderate		
Fanchart:	High		
Stress-test:	Yes (commodity price shock)		
Long term		Low	Triggered module to illustrate the scaling up/down of natural resources indicates lower debt and higher primary balances in the long term (after new oil and gas fields come on stream).
Sustainability Assessment:[2/] Not required as country is not in Fund-supported program.			
Note: The risk of sovereign stress is a broader concept than debt sustainability. Unsustainable debt can only be resolved through exceptional measures (such as debt restructuring). In contrast, a sovereign can face stress without its debt necessarily being unsustainable, and there can be various measures—that do not involve a debt restructuring—to remedy such a situation, such as fiscal adjustment and new financing. 1/ Reported only for surveillance cases. 2/ Required only for program cases; optional in surveillance.			

DSA Summary Assessment

Ruritania's overall risk of sovereign stress is assessed to be moderate. This is consistent with the a moderate near-, medium, and-long term final risk assessment. The near- and medium-term mechanical signals flag high risk driven by the large realized and projected increase in public debt to finance investments in the gas and oil sectors (corresponding to US$20 billion or 25 percent of nominal GDP) and the high sensitivity of debt dynamics to commodity price shocks (as illustrated by the fanchart width and the triggered stress-test). Judgment was applied to arrive at the final assessment in light of substantial liquid assets (70 percent of total debt), which represents an important mitigating factor not accounted for by the models, and relatively limited liquidity risks (moderate GFN mechanical signal). In addition, substantial additional revenue is expected from the new gas and oil fields coming on stream after t+5 (about US$40 billion at current gas prices).

[1]The Appendix illustrates the output of the SRDSF in a "full disclosure" setting. Results presented here are purely for illustrative purposes - portraying the fictitious country "Ruritania".

Table A1. Ruritania: Debt Coverage

1. Debt coverage used for the DSA: 1/	CG	GG	NFPS	PS	Comments on metadata

2. Subsectors included in the chosen coverage in (1) above:

		Subsectors captured in the baseline	
CPS / NFPS / GG (recommended) / CG	1	Budgetary central government	X
	2	Extra budgetary funds (EBFs)	Missing
	3	Social security funds (SSFs)	Missing
	4	State governments	Missing
	5	Local governments	Missing
	6	Public nonfinancial corporations	n.a.
	7	Public financial corporations	n.a.

Comments on metadata: Possible data gaps (State governments); Data unavailable (Local governments)

3. Instrument coverage:	Currency & deposits	Loans	Debt securities	Other accounts payable 2/	IPSGSs 3/

4. Accounting principles adopted:	Basis of recording		Valuation of debt stock		
	Non-cash basis 4/	Cash basis	Nominal value 5/	Face value 6/	Market value 7/

5. Debt consolidation across subsectors:	Consolidated	Non-Consolidated

Color code: : chosen coverage; : missing from recommended coverage; : not applicable.

1/ CG=Central Government; GG=General Government; NFPS=Nonfinancial Public Sector; PS=Public Sector.

2/ Stock of arrears could be used as a proxy in the absence of accrual data on other accounts payable (OAP).

3/ Insurance, Pension, and Standardized Guarantee Schemes, typically including government employee pension liabilities.

4/ Includes accrual recording, commitment basis, due for payment, etc.

5/ Nominal value at any moment in time is the amount the debtor owes to the creditor. It reflects the value of the instrument at creation and subsequent economic flows (such as transactions, exchange rate, and other valuation changes other than market price changes, and other volume changes).

6/ The face value of a debt instrument is the undiscounted amount of principal to be repaid at (or before) maturity.

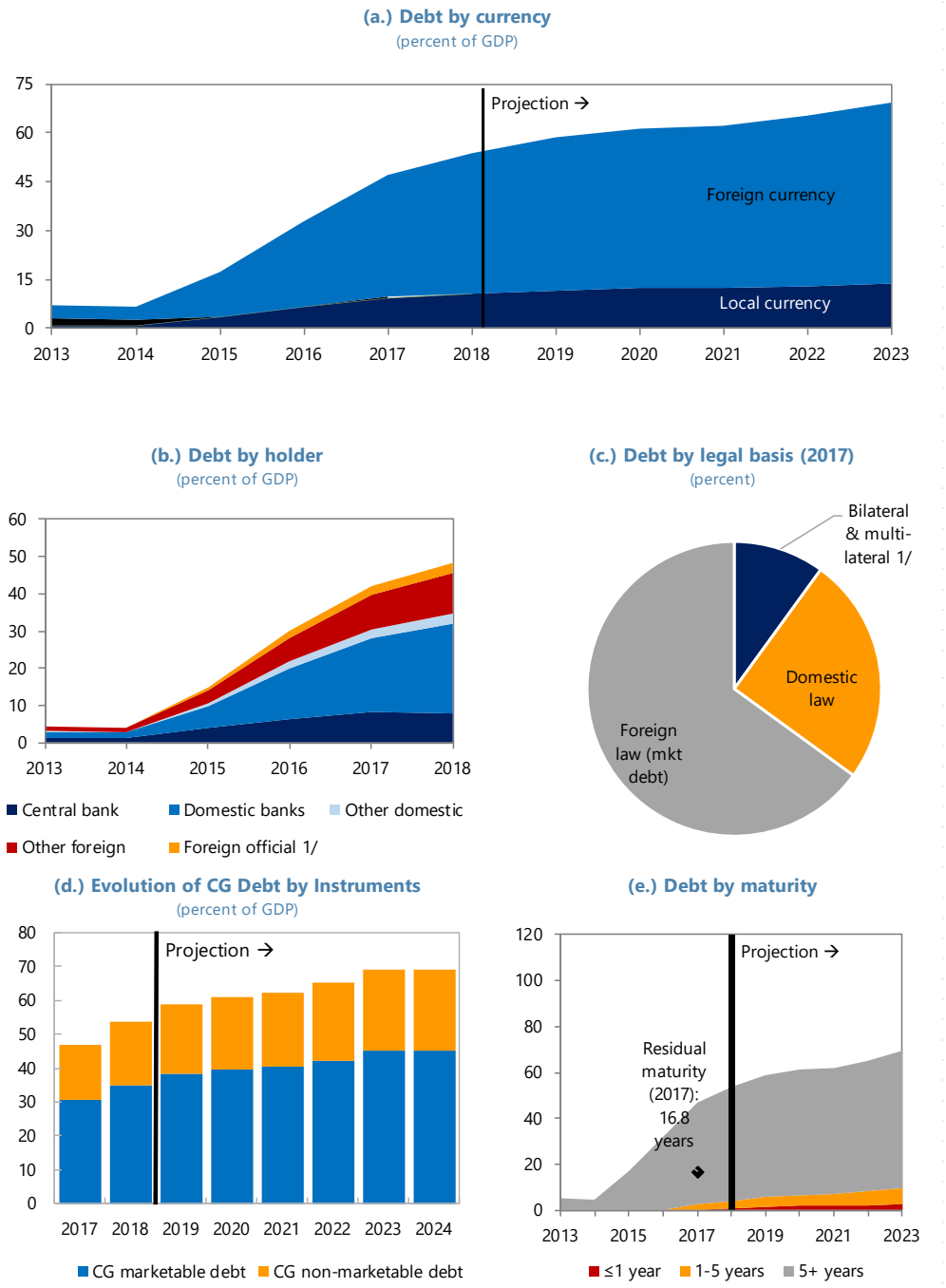
7/ Market value of debt instruments is the value as if they were acquired in market transactions on the balance sheet reporting date (reference date). Only traded debt securities have observed market values.

Reporting on Intra-Government Debt Holdings

		Issuing sector \ Holding sector	1 Budgetary central government	2 Extra budgetary funds (EBFs)	3 Social security funds	4 State/ provincial government	5 Local government	6 Public nonfinancial corporations	7 Public financial corporations (incl. central	Total
CPS / NFPS / GG / CG	1	Budgetary central government							5	5
	2	Extra budgetary funds (EBFs)								0
	3	Social security funds (SSFs)								0
	4	State/provincial government								0
	5	Local government								0
	6	Public nonfinancial corporations								0
	7	Public financial corporations (incl. central bank)								0
		Total	0	0	0	0	0	0	5	5

Staff commentary: *Debt coverage reported is Central Government. Coverage narrower than General Government is assessed to be appropriate given that the government functions are concentrated at the central government.*

Figure A1. Ruritania: Public Sector Debt Structure Indicators



1/ Includes SDR allocations, holdings by regional financing arrangements, and currency union central banks, where applicable.

Staff commentary: *Public debt at end-2018 was mostly denominated in foreign currency, of long-term maturity and issued under foreign law. Debt is mostly held by foreign banks. Over the medium-term, the share of foreign currency debt is projected to increase.*

Table A2. Ruritania: Baseline Scenario

(in percent of GDP unless otherwise indicated) [1/]

	Actual	Projections											
	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	Average
Nominal gross public debt	53.5	58.7	61.1	62.2	65.1	69.3	73.1	70.0	62.4	54.8	47.2	39.6	
Change in gross public sector debt	6.7	5.1	2.5	1.1	3.0	4.1	3.9	-3.1	-7.6	-7.6	-7.6	-7.6	-1.3
Identified debt-creating flows	7.7	8.3	5.0	3.7	5.2	5.8	6.1	-0.9	-5.4	-5.4	-5.4	-5.4	1.1
Primary deficit	7.2	5.0	4.9	3.3	4.1	4.7	4.9	-2.1	-6.6	-6.6	-6.6	-6.6	-0.1
Primary (noninterest) revenue and grant	37.1	37.2	36.3	37.6	36.7	35.6	34.7	41.7	46.2	46.2	46.2	46.2	40.4
Primary (noninterest) expenditure	44.3	42.2	41.2	40.9	40.8	40.3	39.6	39.6	39.6	39.6	39.6	39.6	40.3
Automatic debt dynamics 5/	0.8	3.2	0.1	0.4	1.0	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.2
Interest rate/growth differential 6/	0.8	3.2	0.1	0.4	1.0	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.2
Of which: real interest rate	1.8	3.4	3.5	1.7	1.6	2.0	2.3	2.3	2.3	2.3	2.3	2.3	2.4
Of which: real GDP growth	-1.0	-0.2	-3.4	-1.3	-0.5	-0.9	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1	-1.2
Exchange rate depreciation 7/	0.0	...	...	...	...	...	...	...	...	...	...	...	...
Other identified debt-creating flows	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Residual, including asset changes 8/	-1.0	-3.2	-2.6	-2.7	-2.2	-1.7	-2.2	-2.2	-2.2	-2.2	-2.2	-2.2	-2.3
Gross Financing Need	14.8	15.2	13.7	12.6	15.5	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.5
Memo items:													
Real GDP growth (in percent)	2.2	0.3	5.9	2.1	0.9	1.4	1.7	1.7	1.7	1.7	1.7	1.7	1.9
Inflation (GDP deflator, in percent)	0.9	-2.4	-2.3	1.2	1.8	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.1
Nominal GDP growth (in percent)	12.0	-2.1	3.5	3.3	2.7	3.4	3.8	3.8	3.8	3.8	3.8	3.8	3.0
Effective interest rate (in percent) [4/]	4.9	3.9	3.7	4.0	4.4	5.2	5.5	5.5	5.5	5.5	5.5	5.5	4.9

Source: IMF staff.

1/ Public sector is defined as general government.

2/ Based on available data.

3/ Long-term bond spread over German bonds.

4/ Defined as interest payments divided by debt stock (excluding guarantees) at the end of previous year.

5/ Derived as [(r - π(1+g) - g + ae(1+r)]/(1+g+π+gπ)) times previous period debt ratio, with r = interest rate; π = growth rate of GDP deflator; g = real GDP growth rate; a = share of foreign-currency denominated debt; and e = nominal exchange rate depreciation (measured by increase in local currency value of U.S. dollar).

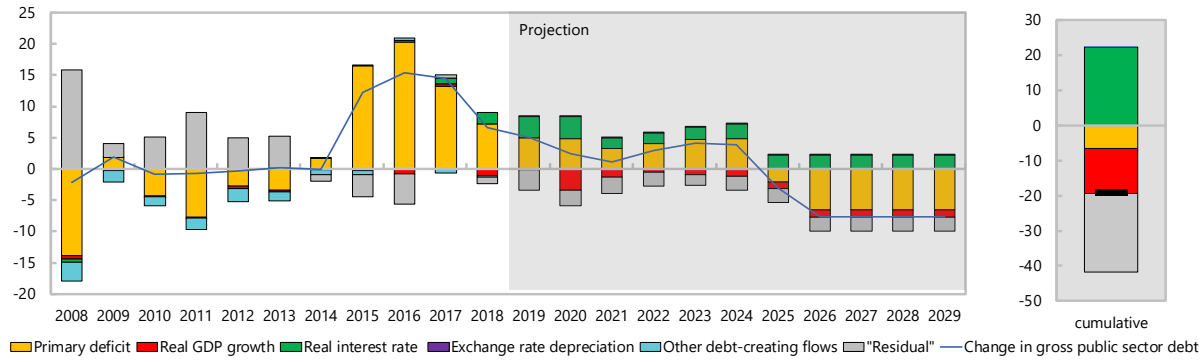
6/ The real interest rate contribution is derived from the numerator in footnote 5 as r - π (1+g) and the real growth contribution as -g.

7/ The exchange rate contribution is derived from the numerator in footnote 5 as ae(1+r).

8/ Includes asset changes and interest revenues (if any). For projections, includes exchange rate changes during the projection period.

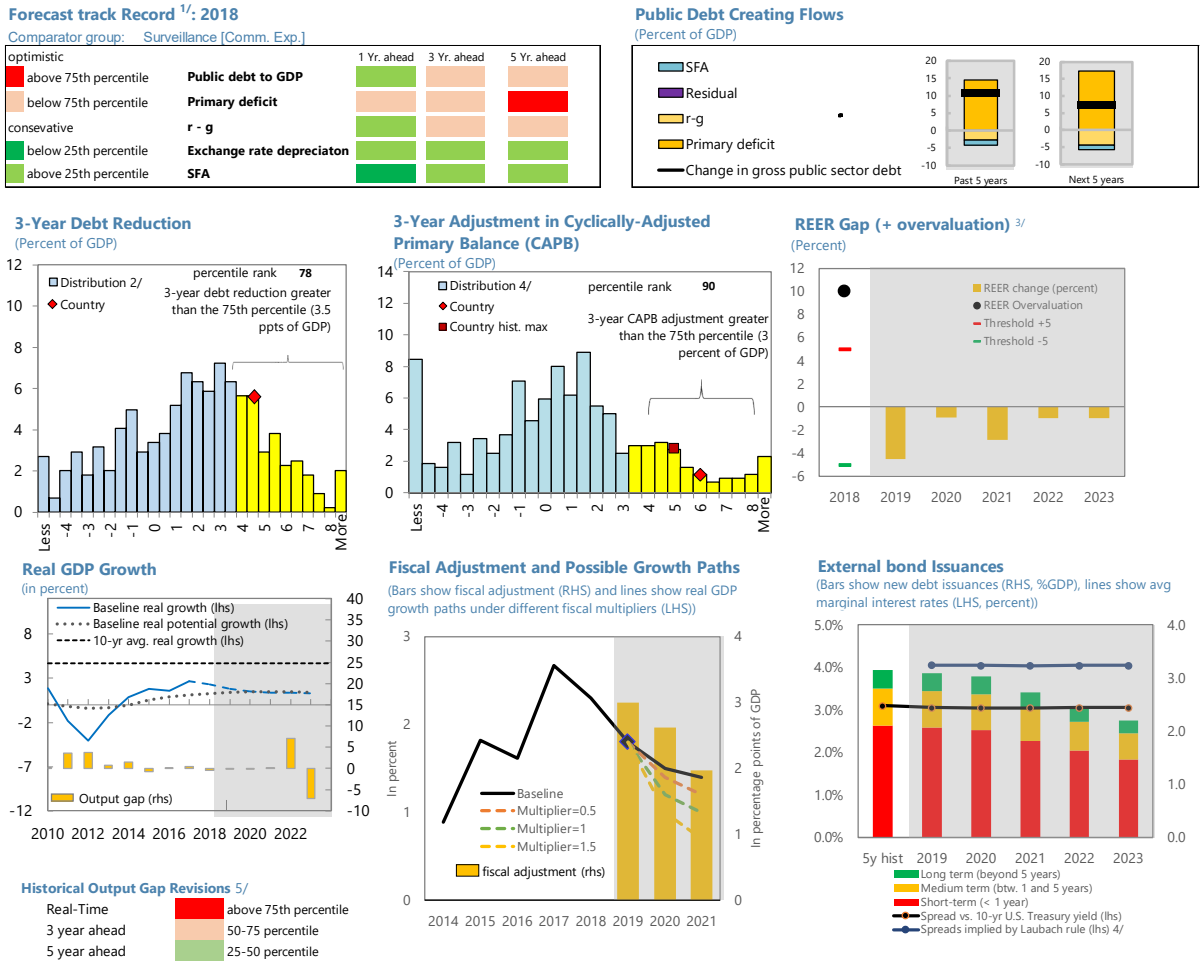
9/ Assumes that key variables (real GDP growth, real interest rate, and other identified debt-creating flows) remain at the level of the last projection year.

Debt-Creating Flows
(in percent of GDP)



Staff commentary: *Public debt at end-2017 stood at 53.5 percent of GDP and its projected trajectory is expected to largely follow developments of the oil and gas sectors. Over the next 5 years, debt would accumulate reflecting the associated investment expenditures. Then debt is expected to decline due to the additional revenue from projects in the oil and gas sectors coming on stream. As the project is profitable overall, cumulative primary balances over the next decade do not impart a net upward contribution to debt. However, higher borrowing costs imply that interest makes the principal positive contribution to debt ratios. GFNs remain relatively stable for the next 10 years.*

Figure A2. Ruritania: Realism of Baseline Assumptions



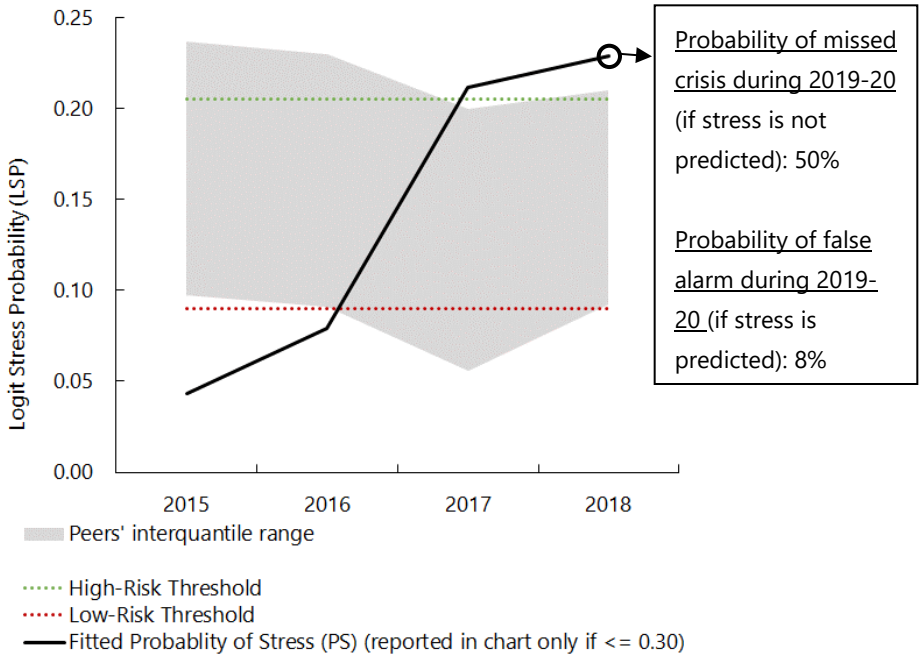
Source : IMF Staff.

1/ Projections made in the spring WEO vintage.
2/ Data cover annual obervations from 1990 to 2018 for MAC advanced and emerging economies. Percent of sample on vertical axis.
3/ Starting point reflects the team's assessment of the initial overvaluation from EBA (or EBA-Lite).
4/ The Laubach (2009) rule is a linear rule assuming bond spreads increase by about 4 bps in response to a 1 ppt increase in the projecetd debt-to-GDP ratio.
5/ Calculated as the percentile rank of the country's output gap revisions (defined as the difference between real time/period ahead estimates and final estimates in the latest October WEO) in the total distribution of revisions across the data sample.

Staff commentary: *Baseline assumptions of most debt drivers do not point to a systematic bias in the forecast track record, which is broadly in line with those observed in peer surveillance countries.*

Figure A3. Ruritania: Near-Term Risk Analysis[1]

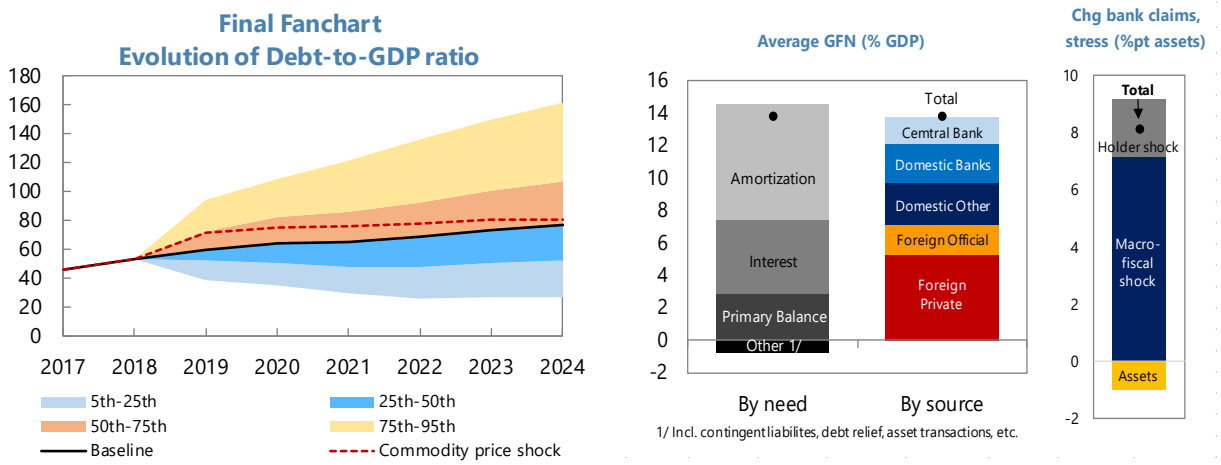
	2015	2016	2017	2018
To forecast stress in [t+1, t+2]	*2016-17*	*2017-18*	*2018-19*	*2019-20*
Logit Stress Probability (LSP)	0.043	0.079	0.212	0.229
Change in LSP	0.030	0.036	0.133	0.017
due to:				
Institutional quality	0.004	-0.005	0.015	0.016
Stress history	0.000	0.000	0.000	0.000
Cyclical	0.011	0.015	0.090	0.002
Debt burden	0.011	0.039	0.064	-0.037
Global	0.004	-0.013	-0.037	0.036



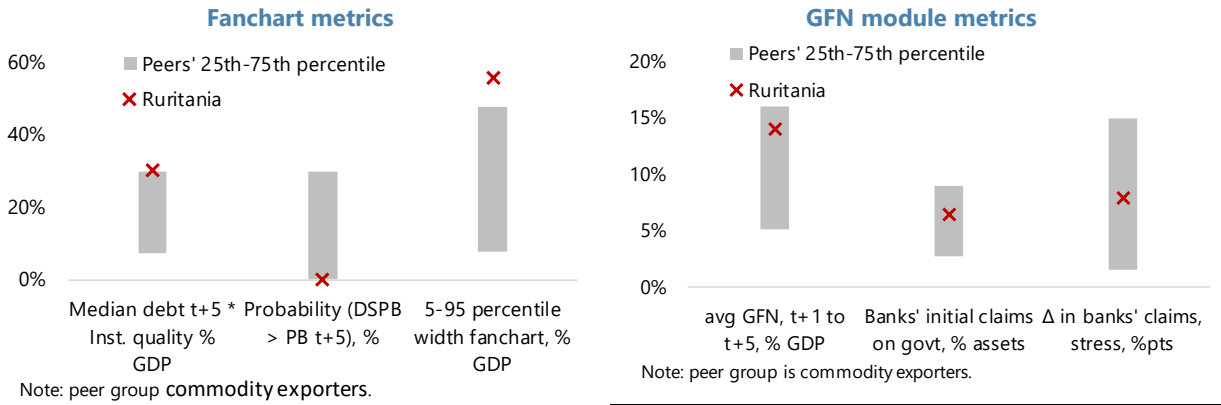
[1]Only reported for surveillance cases.

Staff commentary: *Results suggest a "high" probability of near-term sovereign stress. Main sources of stress appear to be the deterioration of debt burden indicators and cyclical indicators in previous years and a change in exogenous global factors (risk aversion) in 2018. Judgment has been applied to override the mechanical signal and upgrade the near-term risk assessment to moderate given the large liquid asset buffer (70 percent of debt) not accounted for in the total reserves used in the logit regression.*

Figure A4. Ruritania: Medium-Term Risk Analysis

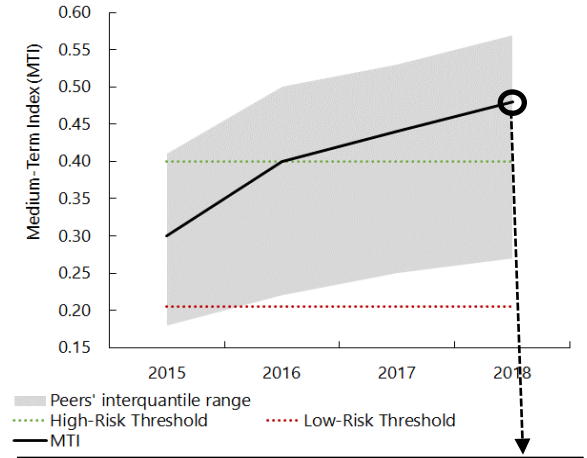


Comparison of Ruritania with Peers



Note: peer group commodity exporters.

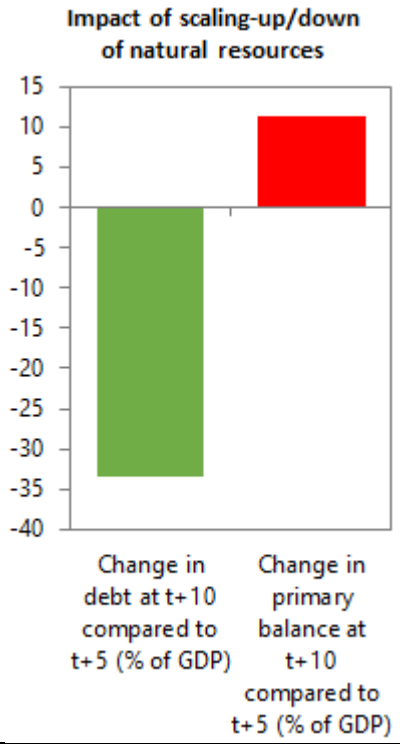
Note: peer group is commodity exporters.



Probability of missed crisis during 2019-23 (if stress is not predicted): 60%
Probability of false alarm during 2019-23 (if stress is predicted): 4%

Staff commentary: *The fanchart's mechanical signal is "high" owing to its relatively wide width and elevated debt level at t+5. The GFN module's mechanical signal is "moderate", in part due to banks' low initial sovereign exposure. Macro-fiscal shocks drive the rise in bank claims in the stress scenario, as contained foreign private financing limits the holder shock. The combined MT signal signal is therefore "high". The triggered commodity price stress-test indicates a t+5 debt/GDP ratio that is below the 75th percentile of the fanchart, although the RAM indicates a "high" likelihood of occuring. Staff's final MT rating is "moderate" due to large liquid asset buffers that mitigate liquidity pressures and support market confidence and the expectation that investment will increase potential growth (and revenue) going forward.*

Figure A5. Ruritania: Long-term Risk Analysis



Staff commentary: *The scaling-up/down of natural resources is triggered for Ruritania as the extraction volumes over t+6 to t+15 are expected to increase more than one standard deviation compared to the last 10 years. The outlook suggests a steady decumulation of debt and improvement in the primary balance.*

INTERNATIONAL MONETARY FUND

November 25, 2020

REVIEW OF THE DEBT SUSTAINABILITY FRAMEWORK FOR MARKET ACCESS COUNTRIES—ANNEXES

Approved By
Jeromin Zettelmeyer

Prepared by an IMF team led by S. Ali Abbas (MCD, formerly SPR) and Manrique Saenz, coordinated by Geoffrey Keim and Daniela Marchettini, and consisting of Tom Best, William Diao, Sandesh Dhungana, Amr Hosny, Plamen Iossifov, Maximilien Kaffo, Samuel LaRussa, Mike Li, Alexis Meyer-Cirkel, Monica Petrescu, Modeste Some, and Tim Willems, all SPR. Jean-Marc Fournier (FAD). This paper has benefitted from extensive discussions with an interdepartmental working group, consultations with external experts and stakeholders, and invaluable guidance by Mark Flanagan (EUR), Hugh Bredenkamp (formerly SPR), and Craig Beaumont (SPR).

CONTENTS

ANNEXES

Annex I. Backtesting Results for Current Framework

1. **This section provides additional results from backtesting the existing framework.** Staff's assessment of the existing framework is based on a rigorous interdepartmental process: two rounds of exchanges with external stakeholders (including academics, investors, and official sector institutions), as well as extensive backtesting. Staff has also benchmarked the existing framework against new, state-of-the-art sovereign risk and debt sustainability analytics (especially, the use of continuous, probabilistic methods), requirements associated with changes in Fund policy (particularly the 2016 reform of exceptional access policy that introduced three zones of debt sustainability), and the emergence of new debt instruments and databases. The results highlighted in this Annex relate to the framework's (i) coverage, (ii) discriminatory (predictive) capacity; and (iii) baseline realism and modeling of uncertainty.

Coverage

2. **The review found that coverage remains an area for further reform.** While most AEs report at least on a general government basis, with only 9 percent reporting on a central government basis, about two-fifths of EMs still restrict coverage to the central government (Table AI.1). Risks from narrow coverage are confirmed by the distribution of revisions to nominal debt levels by coverage level: revisions (percent deviation) were larger and more upward skewed where coverage was limited to the central government (Table AI.2).

Table AI.1 Debt Coverage Reported in MAC DSAs
(percent)

	Country's last DSA	
	EMs (78 countries)	AEs (35 countries)
Central government	34.6	8.6
General government	37.2	80.0
Nonfinancial public sector	11.5	0.0
Consolidated public sector	7.7	5.7
Other	9.0	5.7

Source: MAC DSA Database.

Table AI.2 Historic Debt Data Revisions by Coverage
(percent deviation)

	Debt (pct. deviation)			
	Mean	Median	75th percentile	Skew
Central government	11.0	2.1	4.5	2.8
General government	0.6	1.4	3.0	-2.3
Other definition	1.8	0.0	5.1	0.7

Source: MAC DSA Database.

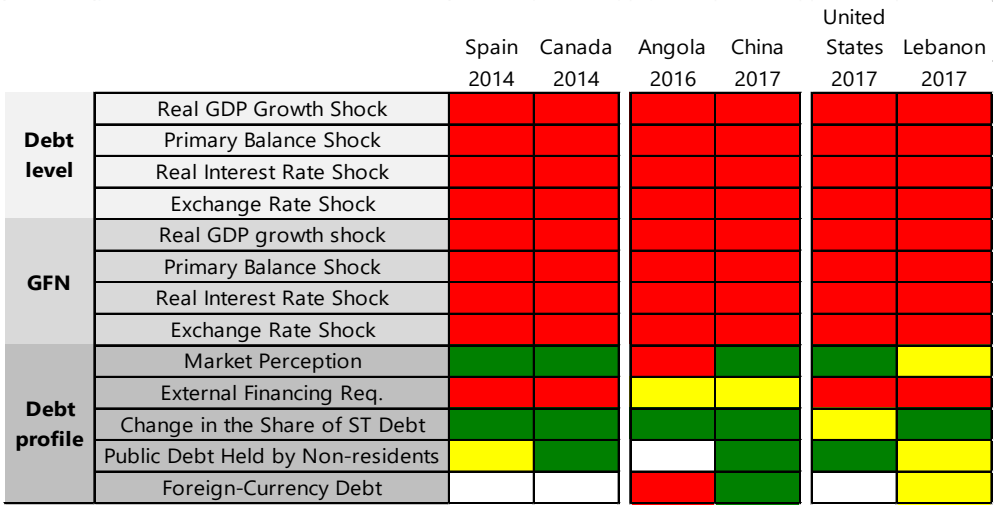
Discriminatory (Predictive) Capacity

3. **The predictive capacity of the threshold approach underlying the current framework has been weak (see Box AI.1).** Some of these limitations were already known at the time of the 2011–13 review. Annex 2 of the 2013 GN reports only the noise-to-signal (NTS) ratio corresponding to the individual thresholds, but the underlying rates of missed crises and false alarms were of the

same magnitude of those found in the backtesting exercise in Box AI.1.[1] This reveals a very high rate of missed crises even in sample (2007-13) associated with individual thresholds, e.g., around 70 percent for the debt and GFN thresholds (see table, Box AI.1). The rate of missed crises associated with individual thresholds is reduced to an average level of 12 percent if one were to consider an OR condition (i.e., call a crisis if any individual heatmap threshold is breached). However, in this case, the rate of false alarms rises to 68 percent. Box AI.1 also shows that the predictive power of the framework has further worsened over time. While missed crises rates for EMs declined slightly out of sample (2014-18), false alarm rates for debt (where the indicator flashed despite no subsequent crisis) rose in 2014–18 relative to 2007–13 for most indicators, both in AEs and EMs.

4. **The limited discriminatory capacity of the current framework implies that countries with very different risk profiles (and arguably, risks) can display very similar heatmaps.** Among advanced economies, in 2014, Spain, which was just beginning to emerge from stress, had almost the same heatmap as Canada. For emerging markets, Angola, a country in stress in 2016, had a very similar heatmap to China in 2017. Finally, across the AE/EM divide, we find similar heatmaps, for instance, for the U.S. and Lebanon in 2017 (Figure AI.1). The latter example highlights that the adoption of just two country buckets may not be sufficient to reflect the wide variation in debt carrying capacity across MAC DSA countries, which depends on differences in the strength of institutions, past history of crises, economic diversification, and the size of domestic investor base.

Figure AI.1 Examples for Poor Discriminatory Power of Heatmaps



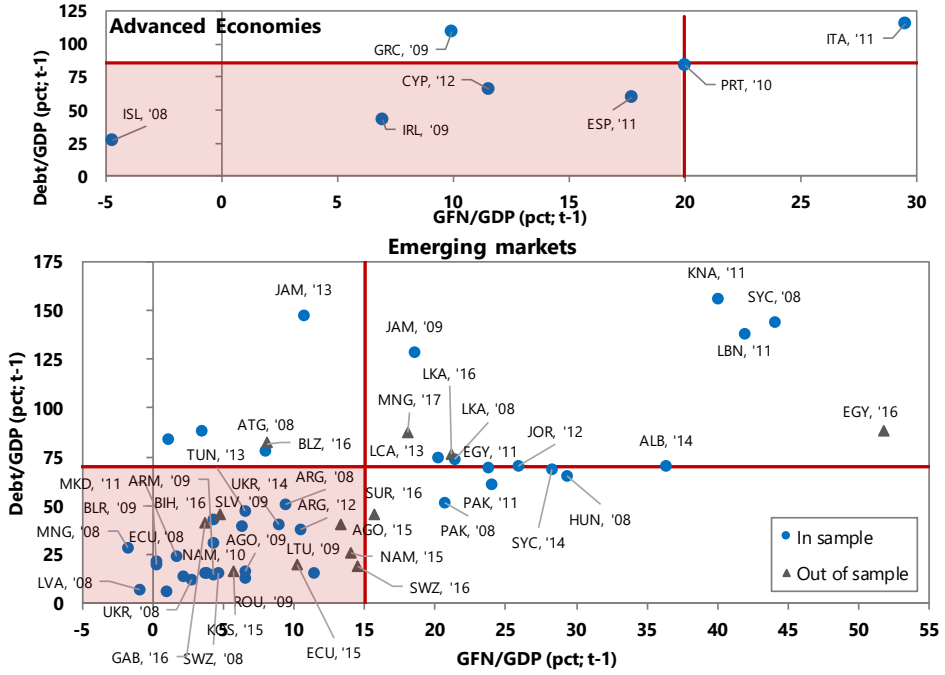
Source: MAC DSA Database.

[1]The criteria have been re-calibrated relative to 2011–13 following a validation exercise vis-à-vis true stress events, but this does not change the assessment on predictive performance of the 2013 framework.

Box AI.1 Predictive Performance of the Existing Framework

The predictive capacity of the approach underlying the current heatmap has been unreliable. Neither the debt nor the GFN ratio gave a signal in t-1 in 57 percent of stress cases in AEs over 2007–13 and in EMs over 2007–18 (using the updated definition of stress episodes; results are broadly similar if the previous definition is used). Missed crises where both indicators were below their respective thresholds are highlighted in the red quadrant of the figures below. While the share of missed crises fell over 2014–17 for EMs, false alarm rates rose in 2014–18 relative to 2007–13 for both groups of countries, in part reflecting rises in debt without the onset of stress. The performance of the five debt profile indicators was also mixed, although with somewhat better predictive power for external and FX debt indicators (e.g., external financing needs/GDP, share of FX debt and share of non-resident held debt).

Debt and GFN-to-GDP Ratios 1-Year Ahead of Crisis Episodes



Note: Red lines correspond to NTS thresholds for debt and GFNs as defined in the 2013 framework.
Source: MAC DSA database.

Share of Missed Crises and False Alarms Using t-1 Threshold Breaches, by Indicator

	Emerging markets					Advanced economies				
	2007-13 (in sample)		2014-18 (out-of-sample)		2007-18 (in sample)	2007-13 (in sample)		2014-18 (out-of-sample)		2007-18 (in sample)
	Missed crises (Type I)	False alarms (Type II)	Missed crises (Type I)	False alarms (Type II)	AUC	Missed crises (Type I)	False alarms (Type II)	Missed crises (Type I)	False alarms (Type II)	AUC
Public debt	70.7	7.5	68.8	15.7	0.66	71.4	18.5	n.a.	35.7	0.54
Gross financing needs	69.4	15.3	60.0	21.7	0.65	71.4	11.1	n.a.	14.3	0.71
External financing needs	50.0	28.4	60.0	32.1	0.60	n.a.	n.a.	n.a.	65.2	n.a.
Δ Share of short-term public debt	67.6	20.1	60.0	16.7	n.a	57.1	22.0	n.a.	8.0	n.a.
Share of FX debt	52.5	26.7	56.3	23.5	0.72	n.a.	n.a.	n.a.	n.a.	n.a.
Share of nonresident-held debt	n.a.	n.a.	27.3	45.0	n.a	57.1	48.5	n.a.	45.0	n.a.
Bond spread	65.2	7.7	72.7	5.2	n.a	100.0	0.0	n.a.	0.0	n.a.

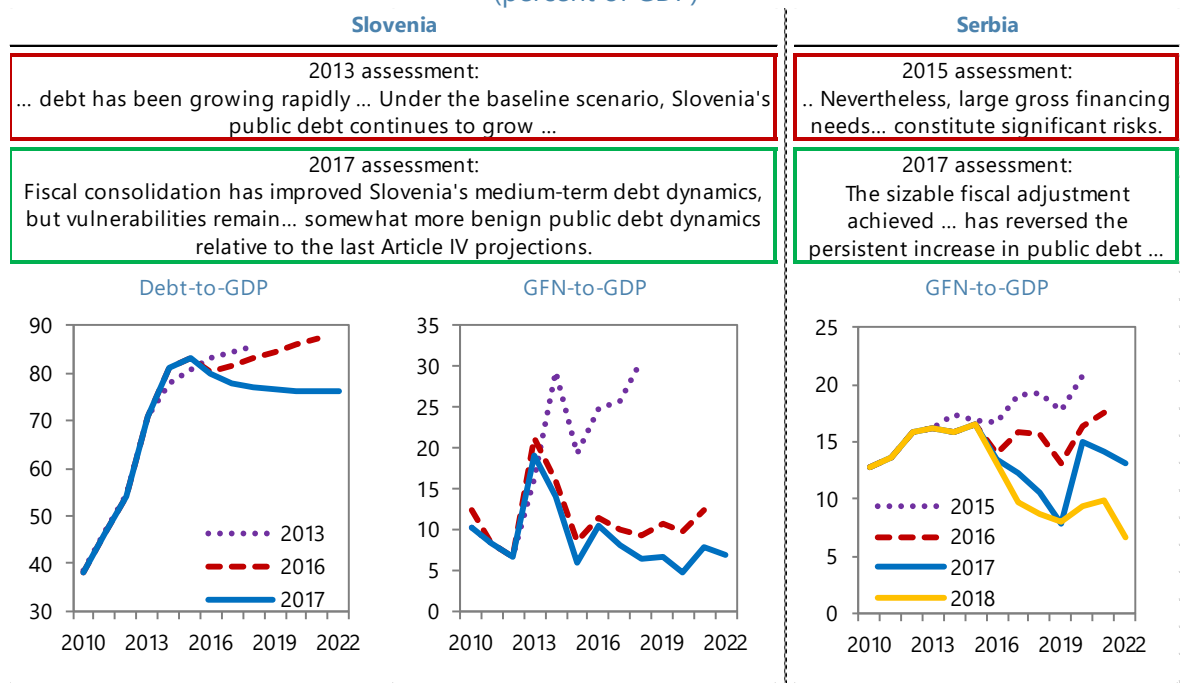
	NTS Ratio	Minimum total mis-specification error	of which: Missed crises	of which: False alarms
OR condition (2006-16 - Average AE/EM, 805 country-year observations)	*76%*	*79%*	*12%*	*68%*
AND condition (2006-16 - Average AE/EM, 805 country-year observations)	*NaN*	*100%*	*100%*	*0%*

Sources: MAC DSA database and Fund staff calculations.

5. **There is little evidence that false alarms resulted from policy action to avert crises in response to risk signals.** In principle, measured false alarm rates could be biased upward as a result of "policy endogeneity", i.e., due to the authorities' timely policy actions to avert crisis in the aftermath of a DSA flagging risks. However, staff analysis of individual cases found little support for this hypothesis. Teams rarely predicted explosive debt or GFN paths or made clear pronouncements on unsustainability. Staff was able to find only two examples, Slovenia and Serbia, where such a policy reaction may have occurred (Figure AI.2).

Figure AI.2 Policy Reactions to Restore Debt Sustainability

(percent of GDP)



Sources: MAC DSA Database; and IMF country reports 14/11; 17/125; 16/287; and 17/263.

6. **A comparison of indicator-based signals with teams' bottom-line assessments shows that team judgment has not been very successful in offsetting the noise generated by the mechanical framework.**

- Ahead of the 16 stress episodes that took place during 2013–17, the debt-to-GDP indicator flagged green in ten cases; while the GFN indicator flagged green in five cases and yellow in two (Figure AIII.3). In most of these cases—seven out of ten—team judgment did no better than these mechanical signals. Only in three cases (Albania 2014, Bosnia and Herzegovina 2016, and Suriname 2016) did team judgment predict greater risks that were picked up by the framework.

- Six of the 16 stress episodes during 2013-17 were correctly predicted by the mechanical framework in the sense that both indicators flash red before a stress episode. However, teams' bottom-line assessments flagged a major sustainability problem in only two. In the remaining four cases, teams provided a more sanguine assessment of risks than suggested by the heatmap. In two instances, debt was ruled sustainable even though debt and GFN indicators both flagged red (Figure AI.3).

- Taken together, these results imply that for the 16 cases shown in Figure AIII.3, team judgment was about in line with the mechanical signal in seven cases (twice correctly and in five instances incorrectly), did worse than the mechanical signal in six cases, and did better than the mechanical signal in just three cases. Based on 2018 stress events, there is little evidence that these patterns have changed (Figure AI.4).[2,3]

- In false alarm cases,[4] comprehensive analysis of DSA chapeaux reveals that teams mainly acknowledged debt and GFN risks already highlighted in the heatmap, with discussion of relevant mitigating factors included in less than a quarter of cases (interestingly, mitigating factors associated with indicators *not* included in the heatmap were more likely to be mentioned). Moreover, references to red flags for debt profile risks were generally uncommon (Table AI.3).

[2]Argentina, Barbados, Pakistan constitute stress events because of their program requests. Turkey satisfies the inflation criterion and Lebanon is exhibiting high spreads. This analysis excludes stress events that began before 2018 (e.g., Angola).

[3]However, in Argentina's case, the stress tests and team judgment corresponded to the shocks that triggered the crisis.

[4]The cases examined are DSAs that (i) contain red flags, (ii) are subject to the high-scrutiny reporting requirements, and (iii) where there was no sovereign stress.

Figure AI.3 Heatmaps Ahead of Stress Episodes

	Country	Year	Heatmap signal: Debt	Heatmap signal: GFN	Team assessment 1/	Stress criterion
Both green	Albania	2014			… concerns about public debt sustainability could undermine the government's capacity to rollover its debt… 2/	Program, LMA
	Ecuador	2015			… the medium-term debt trajectory is on a sustainable path … 3/ 4/	Spreads
	Kosovo	2015			In the past two years, Kosovo has restored a sustainable fiscal stance. 2/	Program
	Bosnia and Herzegovina	2016			… debt will continue on a downward path and debt servicing obligations will be manageable. However, debt indicators could deteriorate rapidly to unsustainable levels in case of sustained adverse shocks… 3/	Program
	Suriname	2016			Suriname's public debt sustainability risks have risen significantly.	Program
GFN stress test identifies risk/debt green	Angola	2015		Real GDP, Real int. rate	Angola's public and external debts are rising but remain sustainable. … The projected path of Angola's public debt is sustainable despite vulnerabilities.	Arrears
	Gabon	2016		Real interest rate	While Gabon's public and external debt remain at moderate levels, they have considerably increased … Under a baseline … debt is projected to increase rapidly only temporarily …	Program
Only GFN red	Ukraine	2014		t+3, t+5	Strengthening public finances in a durable manner remains an overarching policy objective. The authorities agreed that the rapid increase in public debt in recent years is a key vulnerability. 2/ 3/	Rest., Prgm., Arrears, Inflation,
	Namibia	2016		t+3	Though Namibia's public debt level remains low, continuous rise in public debt and increasingly high gross financing needs raise concerns. 4/	LMA
	Swaziland	2016		t+4 - t+5	Though Swaziland's public debt is low, large gross financing needs raise concerns.	Arrears
Both red	Seychelles	2014	t - t+2	t - t+3	… debt dynamics demonstrate elevated sensitivity to shocks …	Program
	Belize	2016	t - t+5	t+5	Belize's public debt will remain high and unsustainable…	Spreads
	Egypt	2016	t - t+5	t - t+5	Public debt sustainability risks remain significant although mitigated by an ambitious fiscal adjustment plan and a friendly domestic investor base.	Program
	Iraq	2016	t+1	t - t+5	… debt remains sustainable over the medium-term, given the projected fiscal path … 4/	Program
	Sri Lanka	2016	t - t+2	t - t+5	While still relatively elevated, public and external debt remains on a sustainable trajectory.	Program
	Mongolia	2017	t - t+5	t+2	… this debt sustainability analysis (DSA) concludes that Mongolia is at high risk of public debt distress … 2/	Program

Source: Fund staff analysis of DSA writeups.
Note: Assessment from two-year ahead DSA unless otherwise noted.
1/ The team assessment is green (red) if the report notes that debt is sustainable (unsustainable) under the baseline; it is yellow otherwise, including when the writeup highlights vulnerabilities and/or mitigating factors.
2/ Results are from two-year MAC DSA prepared under old template for countries that were MAC at the time, or LIC DSF for subsequent PRGT graduates.
3/ Assessment from main text of staff report as there was no DSA writeup; only the baseline could be simulated.
4/ One-year ahead DSA used.

Figure AI.4 Heatmaps Ahead of Countries Exhibiting Vulnerability in 2018

	Country	Heatmap signal: Debt	Heatmap signal: GFN	Team assessment 1/
Both green	Ecuador			While Ecuador's current level of public debt—at 31.3 percent of GDP in 2014—is low by international standards, it has grown rapidly in recent years... Medium-term risks remain manageable, ... 2/
	Turkey			The DSA suggests that Turkey's government debt is sustainable even under different shock scenarios.
GFN yellow	Argentina		Real GDP, ER	Risks to solvency are modest but there are vulnerabilities from the high share of external debt and sizable gross public gross financing needs.
Both red	Barbados	t - t+5	t - t+5	The financing needs generated under the baseline scenario are large and keep growing and, hence, the debt-to-GDP ratio does not stabilize in the next 5 years.
	Lebanon	t - t+5	t - t+5	...risks to public debt sustainability are increasingly significant. Under the baseline scenario, debt and financing needs will continue to rise as a share of GDP.
	Pakistan	t	t - t+5	To improve public debt sustainability and build sufficient fiscal buffers, sustained fiscal consolidation is needed. 3/

Source: Fund staff analysis of DSA writeups.
1/ The team's assessment is green if the report notes that debt is sustainable and red if it notes that debt is not sustainable or fails to stabilize. It is yellow otherwise, even if the writeup mentions other vulnerabilites or mitigating factors.
2/ Three-year ahead DSA from 2015 Article IV Consultation.
3/ One-year ahead DSA.

Table AI.3 Interpretation of Risk Signals in High-Scrutiny False Alarm Cases

	Debt	GFN	Spread	EFN	ΔST	Nonres	FX
Panel 1: False alarms (% red flags)							
Total	61.2	58.6	24.0	57.7	38.3	48.0	29.6
AE	80.0	90.0	0.0	69.2	44.4	48.9	0.0
EM	45.9	49.0	25.0	50.0	36.8	47.6	30.9
Panel 2: Risk signal acknowledged (% false alarms)							
Total	78.0	80.0	33.3	23.4	38.9	41.1	66.7
AE	75.0	88.9	...	8.9	25.0	21.7	...
EM	82.4	75.0	33.3	36.7	42.9	50.0	66.7
Panel 3: Mitigating factor for risk signal (% risks acknowledged)							
Total	1.6	23.3	0.0	13.6	14.3	23.3	7.1
AE	2.8	29.2	...	25.0	0.0	60.0	...
EM	0.0	19.4	0.0	11.1	16.7	16.0	7.1
Panel 4: Other mitigating factors (% false alarms)							
Total	29.3	28.0	33.3	33.0	27.8	26.0	33.3
AE	22.9	25.9	...	22.2	0.0	17.4	...
EM	38.2	29.2	33.3	42.9	35.7	30.0	33.3

Source: Fund staff analysis of chapeaus in DSA writeups.
Note: Panel 1 indicates the percentage of false alarms in the sample of high-scrutiny, non-stress DSAs. Panel 2 indicates the percentage of false alarms where the team acknowledged the risk signal in the chapeau. Panel 3 indicates the percentage of times teams provided mitigating factors in their acknowledgement of a risk signal. Finally, panel 4 indicates the percentage of times where there was a false alarm, but teams supplied a mitigating factor.

7. **An analysis of the text of DSAs reveals that teams considered a wide array of factors when assessing debt sustainability, including factors not covered by the standard framework (Figure AI.5).**

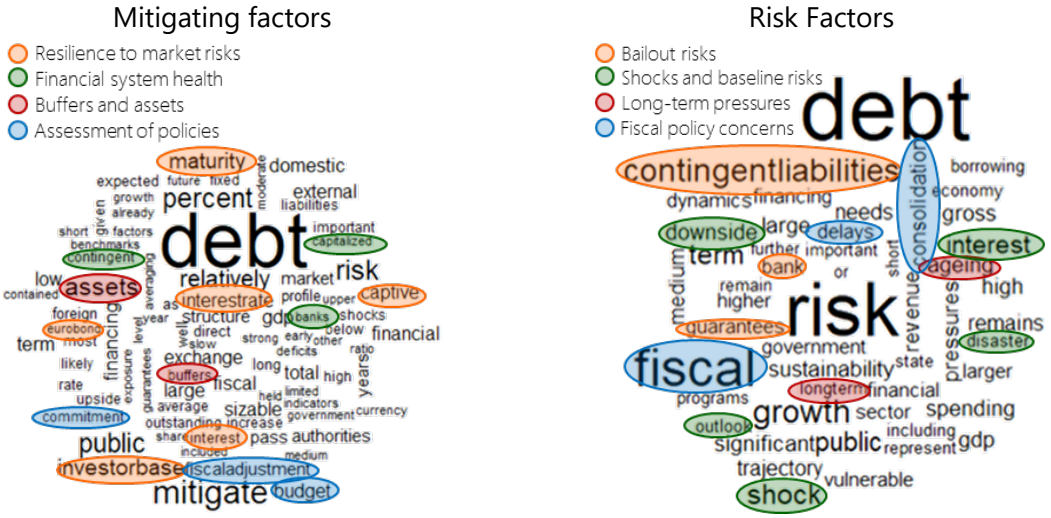
- *Resilience to market risks*. Teams often cited the existence of assets or buffers as a mitigating factor. The holder profile of debt was also referenced, at times as a mitigating factor (e.g., a stable or captive investor base) and at other times as a risk factor (e.g., vulnerability so sudden stops international capital flows).

- *Long-term factors*. The most commonly cited long-term factors were long-term fiscal costs associated with old age benefit and/or health programs.

- *Tail risks*. Teams sometimes mentioned bailout risks, for example from SOEs. They also often mentioned the strong health of the banking system as a mitigating factor against contingent liabilities. Additionally, DSAs for many small states featured discussions of natural disaster risks.

- *Authorities' intentions*. Some teams cited a commitment to strong policies as a mitigating factor, conversely, others raised doubts about the authorities' abilities to deliver needed reforms.

Figure AI.5 Commonly Used Words to Describe Mitigating Factors and Risks in DSA Chapeaus

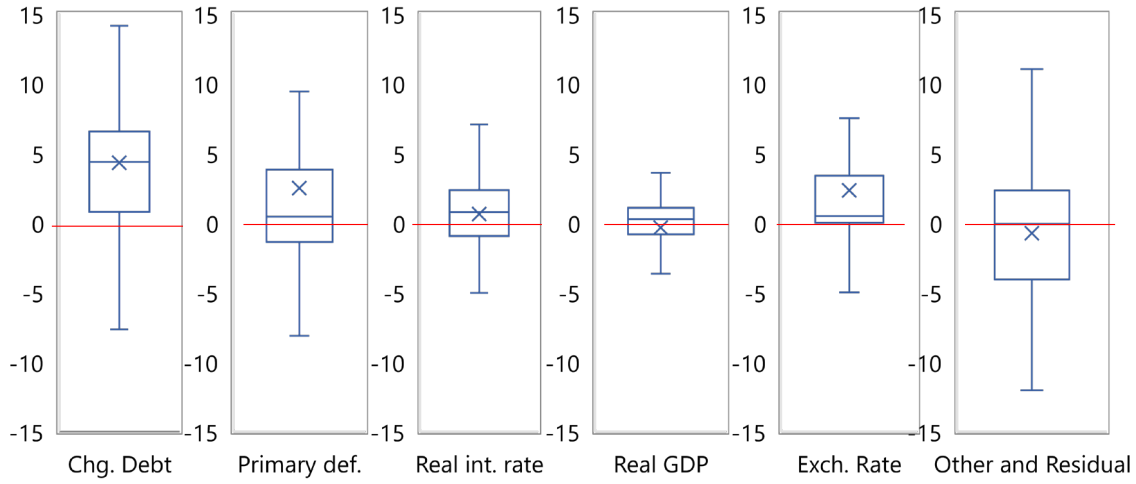


Source: Fund staff analysis of DSA writeups.
Note: These word clouds illustrate commonly used words in clauses that describe mitigating factors (left chart) and clauses that describe risk factors (right chart). In each chart, the size of the word corresponds to the number of times that it was used; a word that appears large appears more often. Both sets of clauses were obtained by staff's analysis of chapeaus from 242 writeups between 2013 and 2018.

Baseline Realism and Modeling of Uncertainty

8. **The introduction of visual realism tools in the 2013 framework appears to have helped reduce optimism in baseline projections for some debt drivers.** On average, projections errors for debt drivers covered by the realism tools—primary balance, growth rate—were somewhat smaller than for debt drivers not covered by the tools—e.g., exchange rate and interest rate. However, the average three-year change in debt/GDP outturn in post-2013 DSAs was about 5 percent of GDP higher than forecast, with an interquartile range of 1–7 percent of GDP. A decomposition of the errors reveals that higher than expected exchange rate depreciations and interest rates seem to have been important factors (Figure AI.6). These debt drivers are not covered by the existing realism toolkit. Risks to the debt path from forecast optimism remain highly relevant—in the latest MAC DSA vintages, 78 percent of country teams projected debt stabilization by year t+5, despite only 34 percent of MACs achieving this since 2011.

Figure AI.6 Decomposition of 3-Year Forecast Errors in the Debt/GDP Ratio (2013–17)



Source: MAC DSA database.
The figure shows the difference between the outturns and the teams' forecast. x is the average difference between outturns and forecasts, the horizontal line is the median difference, the box shows the interquartile range; and the whiskers are upper and lower limits within 1.5 times the 1st and 3rd quartile levels.

9. **Forecast errors with respect to changes in debt/GDP projected in DSAs suggest a continued bias toward optimism.** Since 2013, forecast errors were largest for EMs, especially for the commodity producers among them. Negative debt forecast errors in AEs (as in Ireland) were an exception. Small states also exhibited a high propensity for adverse debt surprises. Several post-crisis advanced economies had large adverse debt surprises, often reflecting major liability management operations. Forecast errors were generally smaller for program than for surveillance countries. Tests for statistical biases are shown in Table AI.4. These tests involve regressing the cumulative 3-year forecast error (from current year to t+2) on a constant; if the constant is statistically significant, there is evidence of a bias. A bias is detected for debt/GDP projections when the test is run on the full sample, but not when the test is performed on program cases only. When the forecast error with respect to debt/GDP is decomposed into the various debt drivers, there is some evidence of bias in real interest rate and real exchange rate forecasts, for both the full sample and the subsample of program cases.

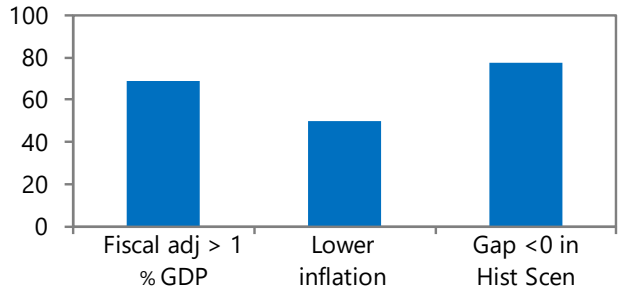
Table AI.4 Tests for Projection Biases

	Baselines		Prgm Scenarios	
	All	Prgm	Hist.	Const PB
Debt/GDP	7.841***	-0.149	-1.255	-4.42
	(2.854)	(6.075)	(6.539)	(6.317)
GFN/GDP	1.381	0.067	-1.49	-3.407
	(1.633)	(3.048)	(3.277)	(3.183)
Real growth	-0.592	1.406	0.572	1.406
	(0.662)	(2.006)	(2.119)	(2.006)
Real IR	2.174**	1.363*	1.780**	1.485*
	(0.917)	(0.76)	(0.793)	(0.784)
Prim. Def.	1.481*	-0.569	3.669	4.899
	(0.843)	(1.218)	(4.474)	(4.819)
ER (contrib)	2.416***	2.003*	...	...
	(0.69)	(1.039)	...	...

Note: Estimates from OLS regressions. The dependent variable is the 3-year (current year to t+2) forecast error (actual-forecast) for the variable listed in the left most column. The independent variable is only a constant. Standard errors in parenthesis. ***, **, * denote sign-ificance at 1, 5, 10 percent levels,
Source: Fund staff calculations.

10. **Optimism bias also exists for output gap estimates.** The text chart below shows features of projections for high-scrutiny DSAs where an initial output gap was negative but closed by the end of the forecast horizon. In many cases, above-potential growth in the baseline was observed in countries experiencing fiscal adjustments and lower inflation. Additionally, if growth evolved according to the historical scenario (based on a historic average), the output gap would not have closed by the end of the projection period.

Figure AI.7. Signs of Potential Optimism
(percent of DSAs 1/)



Source: MAC DSA database.
1/ High-scrutiny DSAs with negative output gaps at t that close to +/-½ percent by t+5.

Annex II. Additional Details on Debt Coverage

This annex describes important debt coverage issues in greater detail, including customizations for liquid assets, consolidation of central banks and the possible inclusion of central bank liabilities in the definition of public debt.

A. Liquid Assets

1. **The new tools introduce specific customizations for liquid assets.**[1] These will include accounting for FX reserves in the near-term risk module and the use of liquid assets as a first defense against rollover shocks in the GFN module's stress scenario. The near-term risks module will allow for the inclusion of readily available liquid assets (e.g. large foreign sovereign wealth funds (SWF)) in the model's 'FX reserves' variable. The GFN module will allow for the use of liquid assets in the stress scenario before extra debt is issued to be absorbed by the domestic banking sector.

2. **While such customization is not feasible for the debt fanchart tool due to data limitations, it could be substituted by staff judgment in specific cases.** For example, for the very few countries with SWF assets in excess of both 100 percent of gross debt and 100 percent of GDP, staff considers a low risk fanchart signal appropriate, as it can be reasonably expected that the government would neutralize an explosive debt path by tapping its large assets. In other countries where such assets are significant but below these thresholds, the mechanical fanchart signal would continue to be based on gross debt, but the overall medium-term risk assessment could be adjusted, as appropriate, based on country teams' judgment informed by the liquidity and availability of these assets. Details on operationalization will be fleshed out in the Guidance Note.

B. Central Bank Consolidation

3. **The new framework proposes central bank consolidations only in cases of central banks with large negative capital positions and/or where the country team considers the central bank to be involved in significant direct monetary financing of the budget and/or quasi-fiscal activities.**[2] Consolidation is appropriate in these cases to fully capture the public debt burden and debt risks. In addition, when the member country's own debt reporting focuses on a consolidated concept, consolidation could benefit the policy dialogue.

4. **In case of central banks with healthy balance sheets, the framework will incorporate the mitigating characteristics of central bank holdings, without consolidation.** From a solvency perspective, substantial central bank-holdings of government debt could represent a mitigating factor when the net worth of the central bank (incorporating the expected value of its future

[1]The definition of liquid assets will refer to government financial assets, including those in SWF, as defined in the Fiscal Monitor, which typically includes currency and deposits, loans and debt securities. This approach helps to ensure cross-country comparability and consistency with statistical principles. However, upon implementation, teams will have the ability to adjust this measure if they see fit (validated by the review process), to reflect information about readily available assets not captured by standardized cross-country databases.

[2]Such consolidation would imply that (i) central bank claims on the government are netted out *and* (ii) central bank debt liabilities (excluding currency and deposits held by residents) are added.

seigniorage profits) is substantially positive—for example, where these holdings reflect a natural expansion of the monetary base. From a liquidity perspective, financing risks associated with central bank holdings of government debt are mitigated by the fact that central banks can typically be counted on to continue funding the government in periods of stress to the extent that this does not aggravate macro instability. These factors can be addressed through incorporation of future seigniorage revenues into the fiscal projections and by accounting for their impact on the government's financing risks. The fact that central bank purchases of government debt rarely exacerbate sovereign financing pressures is embedded in the GFN module, which does not consider these flows as being at risk of a sudden stop.

C. Central Bank Liabilities

5. **Staff proposes a risk-based approach for including two specific types of central bank liabilities in the definition of public debt in countries where the central bank is not consolidated with government accounts for public sector reporting:**

- **Liquidity papers** that are issued solely for monetary policy purposes would normally be excluded from the debt definition used for the DSA, provided (i) no financing to the government can be provided through their issuance; (ii) the government is not *de facto* responsible for paying debt service thereon;[3] and (iii) the securities do not represent a material fiscal risk (as indicated, for example, by a track record of central bank independence and monetary stability). Where one or more of these conditions is *not* met, liquidity papers would be included in public debt and GFNs for DSA purposes unless their outstanding stock can be deemed *de minimis*.[4]
- **Bilateral FX swap liabilities (CBFXS)** will, similarly, not be included in the definition of public debt used for the DSA so long as: (i) they represent normal central bank monetary or liquidity operations (as opposed to sovereign-to-sovereign medium-term balance of payments support), and (ii) the central bank is expected to be able to extinguish the swap position without actions detrimental to government debt levels (e.g. outright government foreign borrowing to pay off the swap). If either of these conditions is not met, the drawn amount of the FX swap should generally be included in the DSA, unless deemed *de minimis*.[5]

  When drawn, swaps reflecting normal central bank liquidity operations are associated with the accumulation of a short-term FX claim on the banks by the central bank. When those claims are repaid, the central bank can unwind the swap. This FX claim on the central bank balance sheet could hence be a feature distinguishing swaps for liquidity purposes from swaps for BOP support purposes. The matching of short-term FX asset and liability would signal the monetary/liquidity nature of these swaps.

[3]In particular cases where the central bank issues Treasury securities in the primary market, solely for monetary policy purposes, these securities would normally be excluded from the debt definition used for the DSA (even though they are a liability of the Central Government), provided (i) funds collected as counterpart for the issuance of those securities will be kept in a blocked account in the books of the central bank that can only be debited for repayment of the said securities; and (ii) the securities do not represent a material fiscal risk.

[4]Further direction as to when claims could be considered de minimis would be included in the Guidance Note.

[5]Idem.

Annex III. Additional Details on the Realism Tools

1. **Staff is proposing to refine and expand the existing realism tools.** The full set of realism tools (Figure AIII.1) could include the following:

- A color-coded table showing the track record for forecast of all debt drivers and public debt at one-, three-, and five-year horizons vis-à-vis a relevant comparator group. The scale shown in the table ranges from green (pessimism) to red (optimism). *If a table reported many red cells, it would be an indication of persistent forecast optimism, warranting discussion or revisions).*

- A decomposition of past and projected drivers of debt dynamics allowing users to identify and scrutinize large changes in debt drivers between the past 5 years and the projection period (next 5 years). This tool is already included in the LIC DSF. *Large shifts in debt drivers (e.g., a drop in the contribution from the real growth-interest differential) would flag risks to projections.*

- A distribution of observed changes in debt-to-GDP ratios over a three-year horizon, with which a country's projected change in debt-to-GDP ratio would be compared. *Projections of a debt reduction that are large in a cross-country context would suggest potential over-optimism.*

- A distribution of fiscal adjustments (three-year change in cyclically adjusted primary balance, as in the current framework), with which a country's projected adjustment would be compared. *The tool would signal an issue if the projected adjustment were large relative to a country's own history or in a cross-country context.*

- A figure showing the evolution of the real effective exchange rate (REER) gap. As in the current framework, the users would be requested to provide an estimate of initial REER missalignment and the template would extrapolate a path using baseline projections of the REER and assuming no change in the equilibrium REER. *An initial over- or under-valuation that was not unwound (i.e. gap that exceeds ±5 percent) would trigger greater scrutiny of exchange rate assumptions.*

- A chart showing how real GDP growth projections compare with potential growth projections and output gap. Signs of optimism (that would merit an explanation) would arise if the output gap without fiscal stimulus is positive at the end of the projection period or there is a significant increase in real growth over the projection period relative to the historical average.

- For countries for which output gap projections have been available since 2010, the SR will also report a color-coded table showing the track record for revisions of real-time, three- and five-year ahead output gap projections,[1] defined as the difference between output gap estimates as of the latest WEO October vintage and the projections. *The scale shown in the table would range from green (cases where output gap revisions are least positive, i.e. below the 25 percentile of the distribution of peer countries) to red (cases where the output gap revisions are most positive, i.e. above the 75 percentile of the distribution). Red cells would indicate negative bias in output gap projections.*

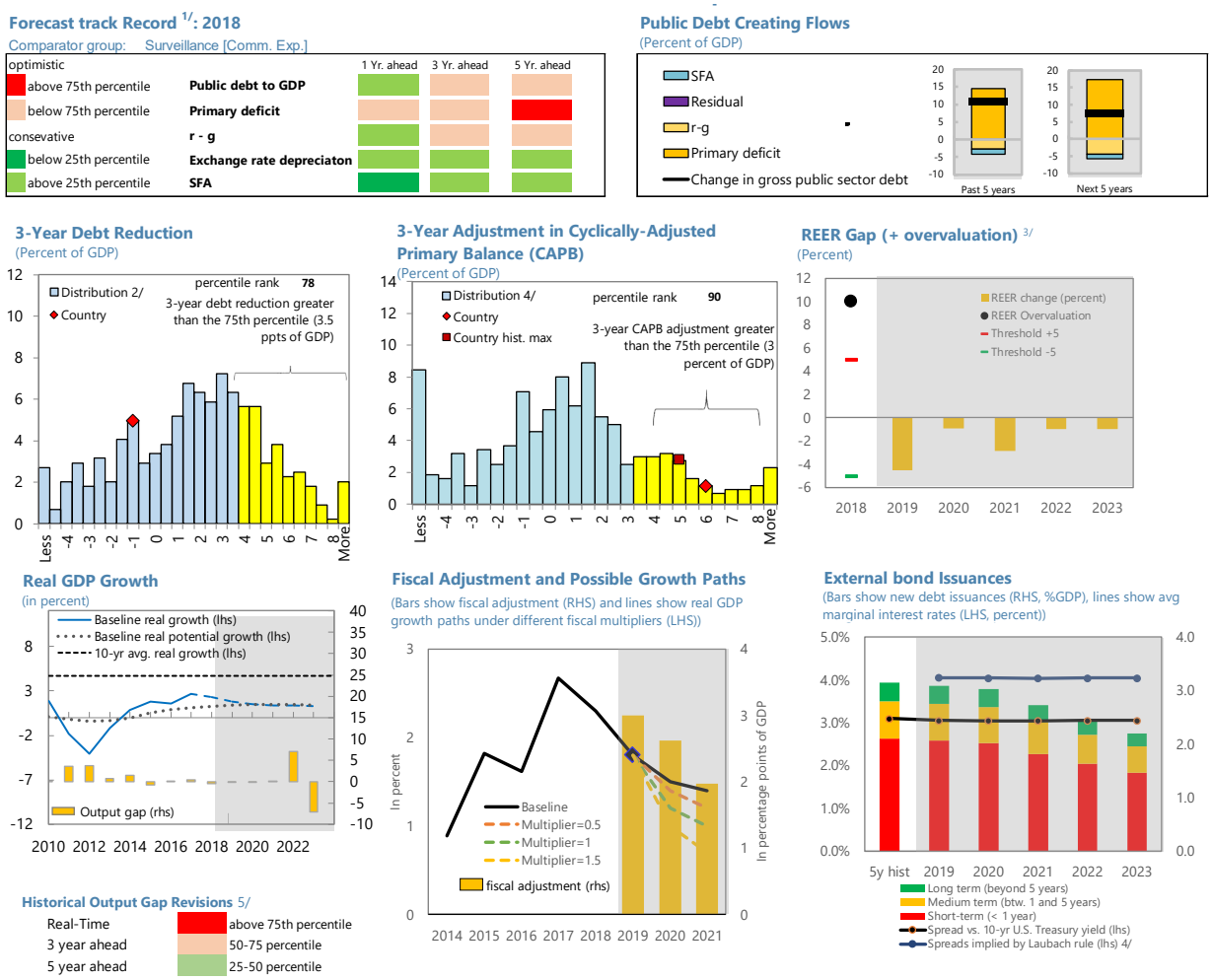
[1]This tool is based on Kangur et. al. (2019) and staff analysis showing the existence of real-time output gap biases for a majority of market access countries.

- A consistency check between fiscal adjustment and growth assumptions. This tool, which is included in the LIC DSF, would compare the impact of the planned fiscal adjustment on growth under a range of plausible fiscal multipliers and persistence parameters with the baseline projected growth path. *Large discrepancies between the baseline and growth implied by fiscal adjustment paths (e.g., a growth pickup during a consolidation) should be explained.*

- A tool assessing new private borrowing and financing terms in terms of maturity composition and spreads under the baseline versus those implied by the Laubach rule.[2] *A shift toward long maturities or a compression in spreads during a debt accumulation would flag a realism problem.*

[2]The Laubach (2009) rule states that bond spreads increase linearly by about 4 bps in response to a 1 ppt increase in the projected debt-to-GDP ratio.

Figure AIII.1. Proposed Realism Tools



Source : IMF Staff.

1/ Projections made in the spring WEO vintage.
2/ Data cover annual obervations from 1990 to 2018 for MAC advanced and emerging economies. Percent of sample on vertical axis.
3/ Starting point reflects the team's assessment of the initial overvaluation from EBA (or EBA-Lite).
4/ The Lauback (2009) rule is a linear rule assuming bond spreads increase by about 4 bps in response to a 1 ppt increase in the projecetd debt-to-GDP ratio.
5/ Calculated as the percentile rank of the country's output gap revisions (defined as the difference between real time/period ahead estimates and final estimates in the latest October WEO) in the total distribution of revisions across the data sample.

Note: The tools in the top row (from left) analyze forecast record for debt drivers vis-à-vis a relevant comparator group (red cells indicating forecast optimism) and compare past and projected drivers of debt dynamics to check for large shifts. The tw left charts in the middle row compare the projected three-year debt reduction and increase in the cyclically adjusted primary balance with the past distribution of such changes (changes corresponding to the yellow shaded portions of the distribution are unusual and may signal overoptimism). The REER gap chart indicates whether an initial overvaluation is expected to be unwound. Finally, charts in the bottom row check whether the output gap closes by the end of projection period, output gap optimism based on the track record on past output gap revisions, check consistency between fiscal adjustment and growth assumptions using plausible multipliers, and assess the realism of new external issuance assumptions based on the history of issuance in the last five years and by comparing assumed spreads with those implied by the Laubach (2009) rule.

Annex IV. Definition of Stress Events

1. **The MAC DSA review utilizes a refined and broad set of criteria to identify the stress events used to calibrate the tools.**[1] The new definitions broadly maintain the stress selection criteria used in the last review.[2] Changes have been introduced to place the definitions on stronger conceptual footings, to ensure alignment with true stress episodes. Additionally, to better capture strains that were not captured under the prior definitions, several criteria have been broadened (e.g. inclusion of large official financing from non-IMF sources; extension of high inflation and spreads from AEs to the full sample).

2. **The mechanical criteria for identifying stress events are as follows.**

i. Episodes associated with large IMF programs (data from the IMF Finance Department and the MONA database) and exceptional financing from other IFIs and donors. Conditions for stress event:
   - IMF Program size equal or greater than 100 percent of quota AND positive disbursement during the first year of the program. Years after the first are considered stress years if there are continuing positive disbursements;
   - Other IFI arrangements above 5 percent of GDP, and positive disbursements in the years classified as stress;
   - Exceptional donor disbursement above 5 percent of external debt.

ii. Episodes associated with default. Conditions for stress event:
   - External arrears equal or greater than 5 percent of public external debt AND increasing at least 10 percent in nominal terms (from the BoC-BoE Sovereign Default Database);
   - Domestic defaults. List from Erce and Mallucci (2018).

iii. Episodes associated with restructuring episodes. Conditions for stress event:
   - List from Das et al. (2012), complemented with Guscina et al. (2017).

iv. Episodes associated with hyperinflation. Conditions for stress event:
   - Doubling of inflation rate compared to the year before AND inflation rate equal or greater than 25 percent OR inflation above 100 percent.

v. Episodes flagged by market-related indicators.
   - For AE. Conditions for stress event:
     - Spreads (for EU countries computed in nominal terms against corresponding German Bund maturity, for other countries computed in nominal terms against corresponding US Treasury maturity as in Baldacci et al., 2011,) equal or greater than 1.5 standard deviations above 10-year mean AND above 150bp, OR spreads above 500bp.
   - For EM. Condition for stress event:

[1]Countries enter the MAC sample only when they graduate from the PRGT status. For instance, Armenia enters the sample in 2013, Bosnia and Herzegovina in 2011, etc.

[2]See IMF (2013), Annex 2.

- 100 percent increase or more in EMBIG spreads compared to the year before AND EMBIG equal or greater than 500bp OR, if EMBIG spreads not available, 100 percent increase in real domestic interest rate compared to the year before AND real domestic interest rates equal or greater than 10 percent

- Loss of market access. Conditions for stress event:
  - List from Medas et al. (2018) and Guscina et al. (2017).

vi. Financial repression. Conditions for stress events:
- Central Bank claims on Central Government (from IFS) greater than 4 percent of GDP AND annual growth greater than 100 percent;
- Commercial Banks' claims on Central Government (from IFS) greater than 9.1 percent of GDP AND growth greater than 100 percent;
- T-bill rate increase (IFS Database) above 4.5ppts y/y (if rate less than 11 percent) OR above 50 percent y/y (if rate equal or above 11 percent)
- List selected individually from the Money and Capital Market Department of the IMF, based on TA reports and FSAPs.

3. **The list of stress events derived with the mechanical criteria underwent an extensive validation process.**

- Members of the MAC DSA team verified the validity of the individual stress country-years derived with the mechanical signals, as well as additional potential stress country-years not flagged by the mechanical criteria, by using IMF staff reports, articles, working papers, newspapers, and additional databases (Paris club, World Bank, Central Banks, etc.). For restructuring episodes it was verified (i) whether the debt treatment was referring to a preemptive or rather a post-default operation and (ii) whether the episode was a part of a larger operation or was an isolated treatment. For preemptive debt treatments, the date of the stress episode was set coincident with the restructuring operation. For post-default episodes, the start date of the stress episode was set coincident with the default and the period between the default and the restructuring operation was considered as continuation of stress only if the country continued to accumulate external arrears (proxied by the increase in the stock of external arrears). Analogously, for debt treatments split in different operations, the period between the different operations and the operations after the first were considered continuation of stress only if the country continued to accumulate external arrears. As a cross check, these stress events were validated by IMF country teams.

- Where two stress episodes are separated only by one year, they were considered the same episode and the intermediate year was considered a stress year even if not flagged by the mechanical criteria. For instance, Jamaica 2012 was considered a stress-country year, even if not identified by mechanical criteria, because Jamaica 2011 and Jamaica 2013 are stress country-years, based on mechanical criteria (iii) and (i), respectively.

- An audit team from the IMF Research Department and the Institute for Capacity Development further reviewed the list in July 2020, resulting in some final minor corrections. [3]

4. **This process allowed to identify 486 stress country-years, corresponding to 139 distinct "stress episodes".**

- Table AIV.1 lists the stress country-years with blue, green, yellow and red color codes for, respectively, stress country-years identified by mechanical criteria, single country-years separating two stress episodes identified by mechanical criteria, country-years inserted by applying judgement and country-years added post-audit. Table AIV.2 provides details on the country-years that were added exercising judgement.

- Among the stress episodes, defaults (37 percent) and market stress (32 percent) were the most common "triggers", in the sense that they occurred more often in the first years of stress episodes.

[3]These revisions regarded stress events identified by the mechanical criteria that were incorrectly dropped out of the sample.

Table AIV.1 Stress Country-Years Used for the Calibration of the Tools					
Albania	2014	Argentina	2000	Belize	2012
Albania	2015	Argentina	2001	Belize	2013
Albania	2016	Argentina	2002	Belize	2016
Algeria	1991	Argentina	2003	Belize	2017
Algeria	1992	Argentina	2004	Bosnia&Herzegovina	2012
Algeria	1993	Argentina	2005	Bosnia&Herzegovina	2013
Algeria	1994	Argentina	2006	Bosnia&Herzegovina	2016
Algeria	1995	Argentina	2007	Bosnia&Herzegovina	2017
Algeria	1996	Argentina	2008	Brazil	1990
Algeria	1997	Argentina	2009	Brazil	1991
Algeria	1998	Argentina	2010	Brazil	1992
Angola	2010	Argentina	2011	Brazil	1993
Angola	2011	Argentina	2012	Brazil	1994
Angola	2015	Argentina	2013	Brazil	1997
Angola	2016	Argentina	2014	Brazil	1998
Angola	2017	Armenia	2014	Bulgaria	1991
Antigua & Barbuda	1996	Armenia	2015	Bulgaria	1992
Antigua & Barbuda	1997	Armenia	2016	Bulgaria	1993
Antigua & Barbuda	1998	Barbados	2014	Bulgaria	1994
Antigua & Barbuda	1999	Barbados	2015	Bulgaria	1995
Antigua & Barbuda	2000	Barbados	2016	Bulgaria	1996
Antigua & Barbuda	2003	Barbados	2017	Bulgaria	1997
Antigua & Barbuda	2008	Belarus	1992	Bulgaria	1998
Antigua & Barbuda	2009	Belarus	1993	Bulgaria	1999
Antigua & Barbuda	2010	Belarus	1994	Bulgaria	2000
Antigua & Barbuda	2011	Belarus	1995	Chile	1990
Antigua & Barbuda	2012	Belarus	1999	Colombia	1998
Antigua & Barbuda	2013	Belarus	2000	Colombia	1999
Antigua & Barbuda	2016	Belarus	2009	Costa Rica	1990
Antigua & Barbuda	2017	Belarus	2010	Costa Rica	1991
Argentina	1990	Belarus	2011	Costa Rica	1993
Argentina	1991	Belgium	2011	Costa Rica	1994
Argentina	1992	Belize	2006	Croatia	1992
Argentina	1993	Belize	2007	Croatia	1993
Argentina	1994	Belize	2008	Croatia	1994
Argentina	1995	Belize	2009	Croatia	1995
Argentina	1998	Belize	2010	Croatia	1996
Argentina	1999	Belize	2011	Croatia	1997

Legend:

	Stress country-year identified by mechanical criteria
	Country-Year separating two stress country-years identified by mechanical criteria
	Stress Country-Year identified by judgment
	Stress Country-Year added post-audit

Table AIV.1 Stress Country-Years (continued)

Croatia	1998	Egypt	2016	Greece	2015
Croatia	1999	Egypt	2017	Greece	2016
Cyprus	2011	El Salvador	1990	Greece	2017
Cyprus	2012	El Salvador	1991	Guatemala	1990
Cyprus	2013	El Salvador	2009	Guatemala	1993
Cyprus	2014	Equatorial Guinea	1991	Hungary	1991
Cyprus	2015	Equatorial Guinea	1992	Hungary	1992
Dominican Republic	1990	Equatorial Guinea	1993	Hungary	2008
Dominican Republic	1991	Equatorial Guinea	1994	Hungary	2009
Dominican Republic	1992	Equatorial Guinea	1996	Iceland	2008
Dominican Republic	1993	Equatorial Guinea	2015	Iceland	2009
Dominican Republic	1994	Equatorial Guinea	2016	Iceland	2010
Dominican Republic	2003	Gabon	1990	Iceland	2011
Dominican Republic	2004	Gabon	1991	Indonesia	1997
Dominican Republic	2005	Gabon	1992	Indonesia	1998
Dominican Republic	2006	Gabon	1993	Indonesia	1999
Dominican Republic	2007	Gabon	1994	Indonesia	2000
Dominican Republic	2008	Gabon	1995	Indonesia	2001
Dominican Republic	2009	Gabon	1996	Indonesia	2002
Dominican Republic	2010	Gabon	1997	Indonesia	2003
Ecuador	1990	Gabon	1998	Indonesia	2004
Ecuador	1991	Gabon	1999	Indonesia	2005
Ecuador	1992	Gabon	2000	Iran, I. Rep. Of	1993
Ecuador	1993	Gabon	2001	Ireland	2009
Ecuador	1994	Gabon	2002	Ireland	2010
Ecuador	1995	Gabon	2003	Ireland	2011
Ecuador	1996	Gabon	2004	Ireland	2012
Ecuador	1997	Gabon	2005	Ireland	2013
Ecuador	1998	Gabon	2006	Italy	2011
Ecuador	1999	Gabon	2007	Italy	2012
Ecuador	2000	Gabon	2016	Jamaica	1990
Ecuador	2003	Gabon	2017	Jamaica	1991
Ecuador	2004	Greece	2009	Jamaica	1992
Ecuador	2008	Greece	2010	Jamaica	1993
Ecuador	2009	Greece	2011	Jamaica	1997
Ecuador	2015	Greece	2012	Jamaica	2009
Egypt	2011	Greece	2013	Jamaica	2010
		Greece	2014	Jamaica	2011

Legend:

	Stress Country-Year identified by mechanical criteria
	Country-Year separating two stress country-years identified by mechanical criteria
	Stress Country-Year identified by judgment
	Stress Country-Year added post-audit

Table AIV.1 Stress Country-Years (continued)					
Jamaica	2012	Latvia	2010	Pakistan	2012
Jamaica	2013	Lebanon	2001	Pakistan	2013
Jamaica	2014	Lebanon	2002	Pakistan	2014
Jamaica	2015	Lebanon	2007	Pakistan	2015
Jamaica	2016	Lebanon	2011	Pakistan	2016
Jordan	1990	Lithuania	1991	Panama	1990
Jordan	1991	Lithuania	1992	Panama	1991
Jordan	1992	Lithuania	1993	Panama	1993
Jordan	1993	Lithuania	1994	Paraguay	1990
Jordan	1994	Lithuania	1995	Paraguay	1991
Jordan	1995	Lithuania	1996	Paraguay	1992
Jordan	1996	Lithuania	1997	Paraguay	1993
Jordan	1997	Lithuania	1998	Paraguay	2002
Jordan	1998	Lithuania	1999	Paraguay	2003
Jordan	1999	Lithuania	2000	Peru	1990
Jordan	2002	Lithuania	2009	Peru	1991
Jordan	2012	Macedonia	2011	Peru	1992
Jordan	2013	Macedonia	2012	Peru	1993
Jordan	2014	Macedonia	2013	Peru	1994
Jordan	2015	Malaysia	1997	Peru	1995
Jordan	2016	Malaysia	1998	Peru	1996
Kazakhstan	1992	Malta	2011	Peru	1997
Kazakhstan	1993	Malta	2012	Peru	2001
Kazakhstan	1994	Mexico	1990	Peru	2002
Kazakhstan	1995	Mexico	1995	Philippines	1990
Kazakhstan	2008	Mexico	1998	Philippines	1991
Korea, Republic of	1997	Mexico	1999	Philippines	1998
Korea, Republic of	1998	Mongolia	2017	Philippines	1999
Kosovo	2010	Morocco	1990	Philippines	2000
Kosovo	2011	Morocco	1991	Poland	1990
Kosovo	2012	Morocco	1992	Poland	1991
Kosovo	2015	Namibia	2010	Poland	1994
Kosovo	2016	Namibia	2016	Portugal	2010
Kuwait	1990	Namibia	2017	Portugal	2011
Latvia	1992	Pakistan	2008	Portugal	2012
Latvia	1993	Pakistan	2009	Portugal	2013
Latvia	2008	Pakistan	2010	Romania	1990
Latvia	2009	Pakistan	2011	Romania	1991

Legend:

	Stress Country-Year identified by mechanical criteria
	Country-Year separating two stress country-years identified by mechanical criteria
	Stress Country-Year identified by judgment
	Stress Country-Year added post-audit

Table AIV.1 Stress Country-Years (continued)					
Romania	1992	Seychelles	2016	Swaziland	2016
Romania	1993	Seychelles	2017	Thailand	1997
Romania	1994	Slovak Republic	2012	Thailand	1998
Romania	1997	Slovenia	2012	Thailand	1999
Romania	1998	Slovenia	2013	Trinidad & Tobago	1990
Romania	1999	South Africa	1990	Tunisia	2013
Romania	2009	South Africa	1993	Tunisia	2014
Romania	2010	Spain	2011	Tunisia	2015
Russian Federation	1991	Spain	2012	Tunisia	2016
Russian Federation	1992	Spain	2013	Tunisia	2017
Russian Federation	1993	Sri Lanka	2011	Turkey	1994
Russian Federation	1994	Sri Lanka	2012	Turkey	1998
Russian Federation	1995	Sri Lanka	2016	Turkey	1999
Russian Federation	1996	Sri Lanka	2017	Turkey	2000
Russian Federation	1997	St. Kitts and Nevis	2011	Turkey	2001
Russian Federation	1998	St. Kitts and Nevis	2012	Turkey	2002
Russian Federation	1999	St. Lucia	2013	Turkey	2003
Russian Federation	2000	Suriname	1993	Turkey	2004
Serbia	2009	Suriname	1994	Turkey	2005
Serbia	2010	Suriname	1998	Turkey	2006
Serbia	2011	Suriname	1999	Turkey	2007
Seychelles	1990	Suriname	2000	Turkey	2008
Seychelles	1991	Suriname	2001	Turkmenistan	1993
Seychelles	1994	Suriname	2004	Turkmenistan	1994
Seychelles	1997	Suriname	2005	Turkmenistan	1995
Seychelles	2000	Suriname	2009	Turkmenistan	1996
Seychelles	2001	Suriname	2010	Ukraine	1992
Seychelles	2002	Suriname	2016	Ukraine	1993
Seychelles	2004	Suriname	2017	Ukraine	1994
Seychelles	2005	Swaziland	2003	Ukraine	1995
Seychelles	2008	Swaziland	2004	Ukraine	1998
Seychelles	2009	Swaziland	2005	Ukraine	1999
Seychelles	2010	Swaziland	2006	Ukraine	2000
Seychelles	2011	Swaziland	2007	Ukraine	2001
Seychelles	2012	Swaziland	2008	Ukraine	2008
Seychelles	2013	Swaziland	2009	Ukraine	2009
Seychelles	2014	Swaziland	2010	Ukraine	2010
Seychelles	2015	Swaziland	2011	Ukraine	2014

Legend:

	Stress Country-Year identified by mechanical criteria
	Country-Year separating two stress country-years identified by mechanical criteria
	Stress Country-Year identified by judgment
	Stress Country-Year added post-audit

Table AIV.1 Stress Country-Years (concluded)			
Ukraine	2015	Venezuela	1990
Ukraine	2016	Venezuela	1994
Ukraine	2017	Venezuela	1995
Uruguay	1990	Venezuela	1998
Uruguay	1991	Venezuela	1999
Uruguay	2002	Venezuela	2002
Uruguay	2003	Venezuela	2008
Uruguay	2004	Venezuela	2009
Uruguay	2005	Venezuela	2010
Uruguay	2006	Venezuela	2011
		Venezuela	2012
		Venezuela	2013
		Venezuela	2014
		Venezuela	2015
		Venezuela	2016
		Venezuela	2017
		Venezuela	1990

Legend:

	Stress Country-Year identified by mechanical criteria
	Country-Year separating two stress country-years identified by mechanical criteria
	Stress Country-Year identified by judgment
	Stress Country-Year added post-audit

Table AIV.2 Stress Country-Years Included through the Exercise of Judgment		
Argentina	2006-07 2010-11	Limited or no access to international capital markets, the central government heavily relied on the Central Bank balance sheet to finance its deficit (IMF Country Report No. 16/69).
Armenia	2014-16	IMF program for 89.4 percent of quota (US$ 0.1 billion) + financing from Eurasian Fund for Stabilization and Development (US$ 0.3 billion) (IMF Policy Paper "Collaboration between Regional Financing Arrangements and the IMF", 2017).
Barbados	2014-17	Large accumulation of domestic arrears estimated at 4 percent of GDP in 2015 (IMF Press Release No. 15/342). In 2016 Moody's downgraded Barbados to Caa1.
Equatorial Guinea	2015-16	Large accumulation of domestic arrears (information from IMF country team).
Lebanon	2006-07	Financing needs satisfied through donor conference (US$7.6 billion) (see IMF WP/08/17)
Malaysia	1997	Large capital outflows (52 percent decline in the Stock Exchange composite index), sharp cut in government spending (-17 percent), 35 percent exchange rate depreciation at end-1997 (see IMF Public Information Notice 99/88).
Namibia	2016-17	Persistent under-subscriptions on government securities in auction across all maturities. Shortfall satisfied by the Government Pension Institution Fund through a private placement. (Information from IMF country team).
St. Lucia	2013	Government unable to sell in auction about 2/3 of total (info from IMF country team).

Annex V. Technical Notes on the Near-Term Risk Tool

The near-term risk module consists of a multivariate logit model whose regressors characterize domestic institutions, stress history, cyclical variables, debt burden, and global conditions. This annex explains how the regressors and the estimation methodology were selected and describes the model's predictive capacity both in- and out-of-sample, robustness checks, and customization options.

A. Selection of Regressors and Choice of the Methodology

1. **The selection of regressors and the choice of the methodology for the MAC DSA EWS was guided by considerations of robustness, statistical forecasting power, and ease of interpretation and reproducibility.**

The model was selected based on a four-step procedure: i) selection of regressors; ii) selection of the estimation methodology; and iii) internal and external consultations on the specification derived in the first two steps.[1]

A. 1. Selection of Regressors

2. **Initially, staff identified a large selection of four types of variables: (a) structural indicators; (b) cyclical indicators; (c) debt and buffer indicators; and (d) global variables.** The indicators in group (b) are potential early warning indicators (EWI) because they provide information on a country's accumulation of imbalances and are associated with the position in the business/financial cycle. As such, they help to predict the timing of a crisis. Indicators in groups (a) and (c), instead, are structural indicators or stock variables, and hence exhibit little variability over time. However, structural indicators can capture the country's ability to react to and recover from shocks, and hence "debt carrying capacity", while debt and buffer indicators provide information on the debt burden (and its composition) and on the risk mitigating effect of buffers. Finally, indicators in group (d) provide information on changes in global economic/financial conditions that may trigger sovereign stress. Staff identified more than 150 variables (and their transformations) that could be included in the four categories.

3. **The selection of regressors from this set was guided by two statistical analyses.**

i. The first analysis aimed at identifying individual *cyclical* indicators (group (b) above) that have strong early-warning proprieties and satisfy dynamic forecasting requirements such as timeliness and stability of the signal (Drehmann and Juselius, 2014). In light of the heterogeneity of the MAC sample, the analysis was performed separately on advanced economies (AE) and emerging markets (EM). The predictive performance of individual indicators was tested in each sub-group at five different (pointwise)

[1]The sample used to estimate the near-term tool covers the period 1990-2017 and includes most Market Access Countries (MACs). MACs refers to advanced economies and emerging markets that principally receive financing through market-based instruments and on non-concessional terms.

projection horizons through a signal detection approach applied to pooled data.[2] This analysis revealed that, while there are some differences in which variables matter, and how much, for AEs vs. for EMs, there are several common early warning indicators for both groups, including debt dynamics and the current account balance (see Figure AV.1). Accordingly, staff opted for a single model for all MACs.

ii. The second statistical analysis employed a Bayesian logit methodology to select early warning indicators (EWI) of sovereign stress together with structural and debt burden indicators. Unlike the first analysis, this methodology accounts for variables interaction; therefore, EWI that may be weak predictors when analyzed in isolation can become relevant when considered in combination with other variables. The methodology can also handle high dimensionality (i.e. the estimation of many regressors, their transformations, and interaction terms at the same time) in the presence of a limited number of observations,[3] and produces a ranking of covariates by their importance.

The outcome of this preliminary two-step analysis highlighted the importance of financial and external imbalances, in addition to fiscal misalignments, as sources of sovereign stress. The analysis revealed also that these factors are more likely to generate sovereign stress when the country is characterized by structural vulnerabilities, revenue volatility, and a debt structure exposed to currency risk.

A.2. Selection of the Methodology

4. **Using the highest performing indicators identified in Step 1, staff estimated a logistic regression (logit) model.** The selection of a logit for the final MAC DSA near-term tool reflected considerations of robustness, high statistical forecasting power, and ease of interpretation and reproducibility. It reflected a trade-off between more sophisticated techniques (e.g., Bayesian approaches, machine learning), which frequently outperform logit models but produce results that are difficult to communicate and reproduce. Compared to the probit approach, the logit methodology is simpler and easier to interpret.[4] Logit models have been widely used in crisis prediction both in literature and institutional contexts [See Manasse, Roubini, and Schimmelpfennig

[2]The assessment of the performance of each indicator at each horizon is performed using the area under the receiver operating characteristic curve (AUC). A completely uninformative indicator has an AUC of 0.5 (corresponding to a ROC curve that equals the 45 line for every threshold), indicating that for any positive signal the probability that the event of interest will materialize in the forecast horizon is equal to the probability of a false alarm. Indicators that are expected to increase (decrease) ahead of the stress episode have higher predictive performance the higher is the distance of the AUC from 0.5 and the closer to 1 (0). The significance of AUC estimates was derived non-parametrically through bootstrap resampling to calculate point-wise confidence intervals.

[3]The methodology uses shrinkage priors to induce sparsity in the coefficient vector. Staff adopted a horseshoe prior that has superior shrinkage properties in sparse signal contexts. The corresponding distribution has an infinite tall spike at 0 and heavy tails, which helps minimize noise and maximize signal (Carvalho et al., 2008). The computation were carried out by Markov Chain Monte Carlo methods (Gibbs Sampler). Thinning (i.e. using only the n^{th} step of the MCMC sample) was used to reduce autocorrelation of MCMC samples and produce a more precise estimate of the posterior. Finally, variables were standardized to improve the efficiency of MCMC sampling (i.e., to reduce autocorrelation in the chains), particularly in presence of interaction terms. The estimates were derived in Matlab with the bayesreg package (Makalic and Schmidt, 2016).

[4]The inverse linearizing transformation for the logit model is directly interpretable as a log-odds, while the inverse transformation of the probit does not have a direct interpretation.

(2003), Pamies, Sumner and Berti (2017), Cerovic et al. (2018)]. The resulting specification is reported in Table AV.1.

Table. AV.1. Preliminary Specification of Multivariate Logit Model

Bucket	Regressor	Coeff	St Coeff
Structural Indicators	Composite Governance Index	-0.77***	-4.47
	Country's Crisis History	0.33***	2.93
	Volatility in Government Revenue (% GDP)	0.10***	2.70
Cyclical Indicators	Current Account (% GDP)	-0.03**	-2.62
	REER Gap to 3Y Average	0.02*	1.64
	Credit to Private Sector Gap (% GDP) (Lag)	0.05***	3.96
Debt-Burden and Buffer Indicators	GG Debt (%GDP) (FD)	0.07***	3.62
	GG Debt Dummy (80% of GDP Cutoff)	0.74**	2.28
	Foreign Currency Governement Debt (% GDP)	0.03***	4.56
	Total external debt (% GDP) (FD)	0.01*	3.47
	International Reserves(% GDP)	-0.03***	-3.05
Global Indicators	US long-term yield	0.71***	4.76
	Commodity Price Index	0.01***	9.03

Source: Fund staff calculations.

Note: Stars indicate statistical significance at the 1 percent (***), 5 percent (**), and 10 percent (*) levels. Standardized coefficients are scaled by variable standard deviations, thus providing a measure of relative importance (see full standardization in Long, 1997).

A.3. Consultations on the Specification

5. **Staff consulted internally and externally on the specification obtained in Step 2, which resulted in some additional improvements.**

- Suggestions from these consultations were tested and endorsed when supported by statistical evidence, yielding to the final specification of the model (Table AV.2).[5] Staff checked the robustness of results to outliers. Removing potential outliers did not have a significant effect on the coefficient estimates and the predictive performance of the model but reduced the statistical significance level of some variables.[6] However, an examination of the most extreme observations showed that the outliers correspond to countries that experienced severe stress events, and, consequently, should not be considered statistical abnormalities as they provide important information on sovereign risk. They were hence maintained in the sample.

- Out-of-sample performance was tested using both temporal cutoffs (by training the model on a certain time period and then testing on the remaining time period) and cross-validation on

[5]External consultation included discussions on the model with several experts, including: C. Reinhart and K. Rogoff (Harvard), E. Duggar (Moody's), L. Giorgianni (Tudor) and S. Pamies (EC).

[6]To identify outliers staff used Stata's ldfbeta command.

country-samples (by training the model on a certain group of countries and then testing on the remaining countries) (see Section AV.4). In addition, the performance of the final specification was compared to that of a benchmark fiscal crisis prediction model based on machine learning. As expected, the machine learning approach led to an improvement in out-of-sample predictive performance, but this was limited (see Section AV.D), and in Staff's view is offset by the greater transparency and economic interpretability of the model shown table AV.2.

- The final specification is intuitive and captures *structural* (institutional quality and stress history, see Box AV.1), *cyclical* (current account balance/GDP, 3-year real effective exchange rate appreciation, credit/GDP gap), *debt burden/buffers* (change in public debt/GDP, public debt/revenue, foreign currency public debt/GDP, and FX reserves/GDP), and *global* (change in VIX, see Box AV.2) factors that may contribute to or mitigate sovereign stress. Moreover, the two-year forecast window (t+1, t+2) should accommodate uncertainty over the exact timing of a crisis; as well as a window that allows time for corrective action (thus, a signal of stress would not mean a stress episode cannot be averted).[7]

Table. AV.2. Specification of Multivariate Logit Model

Bucket	Regressor	Coeff.	Std. Coeff.
Institutional Quality		-1.073 ***	-0.377
Stress History		0.514 ***	0.1006
Cyclical	Current account balance/GDP	-0.024 **	-0.095
	REER (3-year change)	0.013 **	0.070
	Credit/GDP gap (t-1) (if + ve)	0.086 ***	0.258
Debt burden and buffers	Δ(Public debt/GDP)	0.052 ***	0.1182
	Public debt/revenue	0.002 ***	0.1213
	FX public debt/GDP	0.024 ***	0.1601
	International reserves/GDP	-0.034 ***	-0.2348
Global	ΔVIX	0.015 ***	0.1373
Number of Observations			1,579
LR chi2			246.70
Pseudo R2			0.25

Note: Stars indicate statistical significance at the 1 percent (***) and 5 percent (**) levels. Standardized coefficients are scaled by variable standard deviations, thus providing a measure of relative importance (see full standardization in Long, 1997). For instance, the standardized coefficient for the FX public debt to GDP is about 1.4 times the magnitude of the coefficient for the change in public debt-to-GDP. This implies that ceteris paribus, a 1 standard deviation higher FX public debt-to-GDP ratio (about 16.8 percent of GDP, see Table AV.5) would have roughly the same effect on the stress probability as a 1.4 standard deviation increase in change in public debt-to-GDP (approximately 7.5 percent of GDP, see Table AV.5).Source: Fund staff calculations.

6. **The variables included in the final specification are widely used in the literature, albeit not in one single model (also due to data constraints that staff has worked hard to overcome).** Bassanetti, Cottarelli, and Presbitero (2019) highlight the importance of debt dynamics in the lead up to sovereign stress. Kumar and Woo (2010); Cecchetti et al. (2011); Cyclical changes and global indicators are well-established regressors in models of sovereign stress (e.g. Pamies Sumner and Berti, 2017; and Medas et. al., 2018). Finally, structural variables feature prominently in Reinhart et al. (2003); Kraay and Nehru (2006); Manasse and Roubini (2009); and Fournier and Bétin (2018).

[7]For purposes of coding the left-hand side (stress/non-stress) variable, cases where two stress episodes were separated only by only one year, were considered a single episode.

7. **The specification underwent a technical audit conducted by an independent team of economists from the IMF's Research Department and its Institute for Capacity Development. Its main results and recommendations are synthesized below:**

i. Estimates proved to be broadly robust to sample selection. In particular, the audit team performed two analyses:

- First, the team checked whether the estimated coefficients deviate from the baseline estimates using 15- and 20-year windows and running all feasible rolling regressions. In 85 percent of cases (161 out of 190), the estimated coefficients remain within the 2-standard-error bands of the baseline coefficients. When the rolling-regression coefficients deviate beyond the bands, the deviation is small, and the sign is preserved. As far as statistical significance is concerned, in 15 percent of cases (18 out of 120) significance is lost using a 15-year window, due to the shorter sample size. With 20-year windows, statistical significance is preserved at least at a 10 percent level in 99 percent of cases (69 out of 70). Exceptions are the coefficients attached to public debt to revenue and current account balance to GDP, which lose significance in one subsample.
- Second, the team removed from the baseline regression specification the observations of one country at a time and checked the extent to which the coefficients attached to the remaining explanatory variables deviated from their baseline values, and whether they remained significant. In all cases the sign of the coefficients remained unchanged. In more than 95 percent of cases the estimated coefficients remained within the 2-standard-error bands of the baseline coefficients, and in over 99 percent of cases the coefficients remained statistically significant at least at a 10 percent level. Among the coefficients that become insignificant when a particular country is removed, the current account balance to GDP and the change in the REER were the least robust.

The results suggest that that the specification is robust and stable at a comfortable statistical level. To decide whether the comparatively less robust variables (the current account balance and the REER change) should remain in the specification, the MAC DSA team performed an *out-of-sample* validation (over the period 2016-17 to predict stress in 2017-19) to check whether removing the two variables would affect predictive performance.[8] This analysis led to an out-sample AUC of 0.9737 when the two variables are included against 0.9698 when they are excluded. While the difference is minor (which is likely partly related to the small out-sample size), the comparison supports the inclusion of the two variables in the final model. It must also be noted that external and sovereign crises are frequently correlated, and some studies use a definition of external crisis that is very close to the sovereign crisis definition (for instance Catao and Milesi-

[8]Regression metrics such as R2, F-statistics, and p-values are all in-sample metrics: they are applied to the same data that is used to fit the model. However, a good fit does not necessarily lead to a good forecast. For example, overfit models typically have very small in-sample errors and low p-values but perform poorly in forecasting.

Ferretti, 2014). This further supports the inclusion of external sector variables in a model aimed at predicting sovereign stress.

ii. The audit team recommended investigating whether the use of fixed effects (FE) could further improve the forecasting performance of the model. The MAC DSA team considered this option but decided against it for both conceptual and statistical reasons:

- The use of FE estimated over the period 1990-2015 would penalize countries that have improved their debt carrying capacity over time, particularly post-2015, either by implementing reforms to strengthen their institutions, or undergoing structural transformations (for instance through discovery of natural resources) or experiencing debt restructuring/relief. The use of slow-moving structural variables accounts for this evolution while still providing relevant information on debt-carrying capacity. In addition, the use of country fixed effects is politically sensitive and difficult to communicate to the authorities and the public, as it suggests that some countries suffer from inherent unidentified structural characteristics that make them more vulnerable to crises and are not amenable to reform, even in the long run.
- While the predictive capacity of the model (measured by the AUC) seems to improve when country fixed effects are added to the baseline model (0.91 AUC vs 0.88), this effect turns out to be driven by a change in the sample, rather than a genuine improvement. Introducing FE more than halves the size of the sample (675 observations for 52 countries against 1,675 for the pooled logit), because the fixed effect can only be computed for countries that experienced stress over the estimation period and, consequently, have variability in the dependent variable. This implies that most advanced economies drop out of the FE sample. As a result, the fixed effect approach would make it impossible to apply the model to advanced countries, as coefficient estimates of the fixed effect would not exist for such countries.

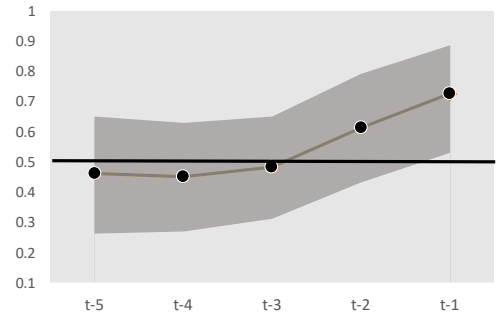
While the option of using fixed effects was dismissed for the reasons above, the analysis provided an additional robustness test for the estimates. The significance and sign of the coefficients remains broadly stable when fixed effects are estimated, except for the coefficient of "stress history", which switches from a positive to a negative sign, and for the coefficients of FX public debt/GDP and International reserves/GDP, which lose statistical significance. In both cases, this is likely due to the fact that the estimation is performed only on countries that experienced stress and that the variables that lose significance are slow moving and hence likely to be captured by the fixed effect.

iii. The audit team also recommended using standard errors corrected for heteroscedasticity and within-country correlation. The MAC DSA team followed this suggestion and adopted robust standard errors. All coefficients remain statistically significant except for the current account variable; however, this is maintained in the regression for the reasons explained in point i above.

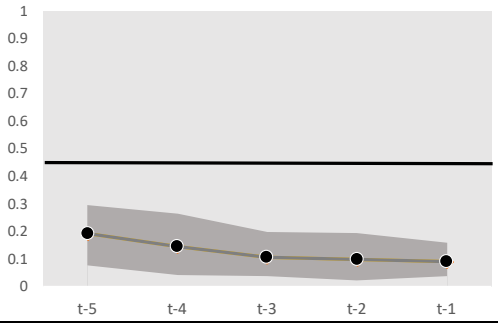
Figure AV.1 Predictive Performance (in terms of AUC) of Individual EWI at *t*+1 to *t*+5

ADVANCED ECONOMIES

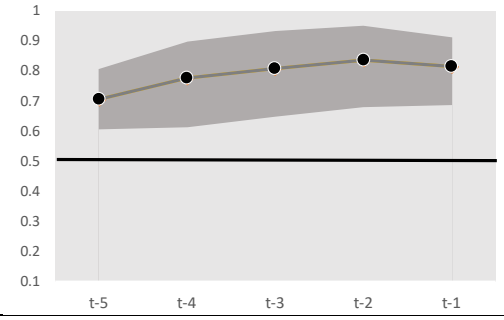
Government Debt Gap (HP Filter)



Current Account Balance



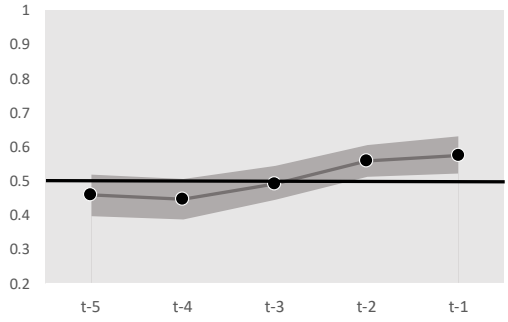
Credit to GDP Gap (HP Filter)



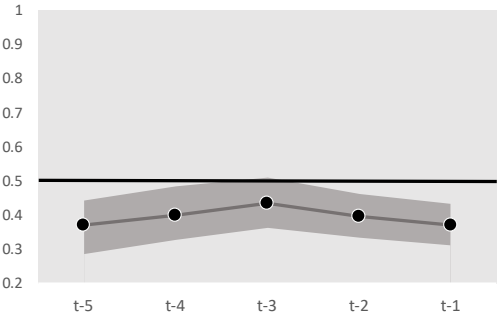
The Figure shows that, for AE, debt increases (in the picture measured as gap from a trend) are significant predictors of sovereign stress 1-2 years before the stress materializes. Cyclical indicators of external of financial imbalances are significant predictors over the full 5-year projection horizon.

EMERGING MARKETS

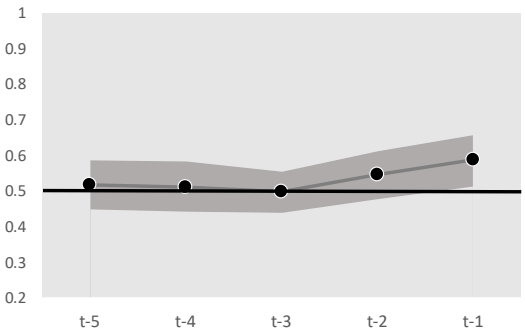
Government Debt Growth



Current Account Balance



REER Percent Change



For EM, debt increases (in the picture measured as YoY debt growth) are significant predictors of sovereign stress 1-2 years before the stress materializes. Cyclical indicators of external imbalances are also significant predictors at least 1-2 years ahead. It can be noted that the predictive performance of individual indicators, even if significant, is lower than for advanced economies, which support the idea of interacting cyclical variables with structural indicator and stock variables.

95 percent confidence interval Median AUC

Box AV.1 Capturing Country Heterogeneity in the Logit Regression

To capture country heterogeneity in a granular continuous way, Staff examined several slow-moving variables which, reflecting structural characteristics, could inform on countries' inner "debt carrying capacity". 1/

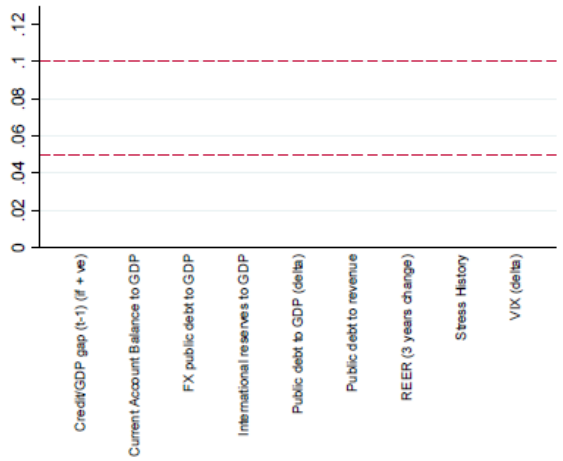
Estimation results (see Table) suggest that the WGI-based variable and stress history have strong predictive power and deliver the best statistical properties relative to the other candidates. In particular, the WGI-based variable (the "quality of institution" index in the logit regression) significantly outperform other variables in terms of statistical significance, coefficient magnitude and robustness to different specifications. For instance, the audit team found that significance of the variable remains intact under different specifications and country samples (Figure).

Table. Alternative Logit Specifications Including Different Structural Variables

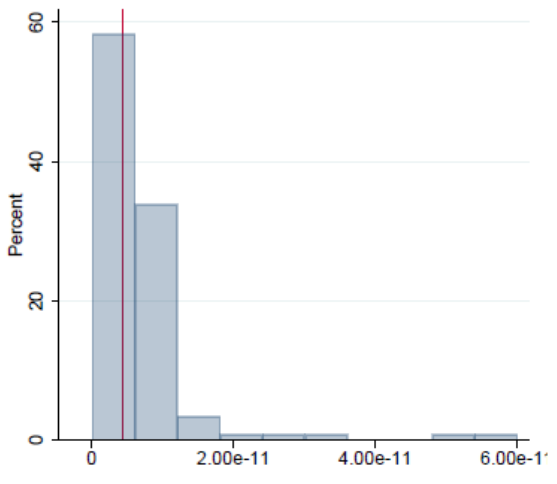
	Coeff	P--value		Coeff	P--value		Coeff	P--value
Institutional Quality	-0.95	0.00	Institutional Quality	-1.04	0.00	Institutional Quality	-1.01	0.00
Stress History	0.51	0.01	Stress History	0.46	0.01	Stress History	0.54	0.00
Current account balance/GDP	-0.02	0.07	Current account balance/GDP	-0.02	0.11	Current account balance/GDP	-0.02	0.06
REER (3-year change)	0.01	0.05	REER (3-year change)	0.01	0.03	REER (3-year change)	0.01	0.04
Credit/GDP gap (t-1) (if + ve)	0.09	0.00	Credit/GDP gap (t-1) (if + ve)	0.09	0.00	Credit/GDP gap (t-1) (if + ve)	0.09	0.00
International reserves/GDP	-0.03	0.00	International reserves/GDP	-0.03	0.00	International reserves/GDP	-0.04	0.00
Δ(Public debt/GDP)	0.05	0.00	Δ(Public debt/GDP)	0.05	0.00	Δ(Public debt/GDP)	0.05	0.00
Public debt/revenue	0.00	0.02	Public debt/revenue	0.00	0.01	Public debt/revenue	0.00	0.00
FX public debt/GDP	0.02	0.00	FX public debt/GDP	0.03	0.00	FX public debt/GDP	0.02	0.01
ΔVIX	0.01	0.00	ΔVIX	0.02	0.00	ΔVIX	0.01	0.00
PPP GDP per capita	0.00	0.06	Age dependency ratio	0.01	0.26	GDP Share (% of world GDP)	-0.25	0.08
	Coeff	**P--vaue**		**Coeff**	**P--vaue**		**Coeff**	**P--vaue**
Institutional Quality	-1.05	0.00	Institutional Quality	-1.20	0.00	Institutional Quality	-1.45	0.00
Stress History	0.50	0.01	Stress History	0.42	0.04	Stress History	0.15	0.52
Current account balance/GDP	-0.02	0.06	Current account balance/GDP	-0.04	0.01	Current account balance/GDP	-0.09	0.00
REER (3-year change)	0.01	0.04	REER (3-year change)	0.01	0.06	REER (3-year change)	0.01	0.14
Credit/GDP gap (t-1) (if + ve)	0.09	0.00	Credit/GDP gap (t-1) (if + ve)	0.09	0.00	Credit/GDP gap (t-1) (if + ve)	0.09	0.00
International reserves/GDP	-0.03	0.00	International reserves/GDP	-0.03	0.00	International reserves/GDP	-0.05	0.00
Δ(Public debt/GDP)	0.05	0.00	Δ(Public debt/GDP)	0.06	0.00	Δ(Public debt/GDP)	0.02	0.32
Public debt/revenue	0.00	0.01	Public debt/revenue	0.00	0.01	Public debt/revenue	0.00	0.09
FX public debt/GDP	0.02	0.00	FX public debt/GDP	0.03	0.00	FX public debt/GDP	0.01	0.33
ΔVIX	0.01	0.00	ΔVIX	0.01	0.00	ΔVIX	0.02	0.00
r-g volatility	0.00	0.56	UN Human Development Inde	0.24	0.88	ICRG Compposite Index	-0.05	0.10

Figure. P-values of the Institutional Quality Index under Different Specifications and Country Samples

Taking out one explanatory variable at a time



Taking out one country at a time



Box AV.1 Capturing Country Heterogeneity in the Logit Regression (Concluded)

While WGI are perception-based indicators, they are considered good proxies for institutional quality (see for instance Faria, A. and Mauro, P., 2009), as they are a summary measure of the largest set available of such indicators, based on several hundred individual variables measuring perceptions of governance, drawn from 31 separate data sources constructed by 25 different organizations, ranging from think-tanks to governments, multilateral organizations and commercial firms.

In addition, the use of the institutional quality index is in line with the use of the CPIA index (not available for MACs) in the composite index of LIC DSF.

In cases where teams assess the WGI-based institutional quality variable to be a poor proxy for the true institutional quality of the country, and the variable is deemed to have a disproportionate effect on the mechanical signal from the logit, teams would be able to incorporate this into their judgement when arriving at the final risk assessment.

1/ While Staff considered the WB Doing Business indicators, the historical series is too short (starting in 2003) to support a robust regression with an adequate number of crises.

2/ Only two of the six WGIs are used in the quality of institution index: Government Effectiveness and Regulatory Quality.

Box AV.2 Capturing Regional Spillovers in the Logit Regression

In some crises, spillover risks are poorly proxied by the VIX, because contagion is of a regional rather than global nature (for example, the VIX was negative during the euro area crisis).

To capture non-global dimensions of spillovers, staff tried several variables: the share of AE or EM countries in stress, the share of countries in stress in each region, the share of countries with strong trade linkages or cross-border flows. However, in all cases the corresponding variable was not statistically significant. In contrast, the coefficient on the share of currency union (CU) members in stress turned out to be highly significant (see Table), consistent with both the experience during the euro area sovereign debt crises (see performance in individual countries in Figure AV.4), and stress episodes in CEMAC witnessed in the wake of the 2014-15 oil price drop.

Table: Specification of Multivariate Logit Model with CU variable

Bucket	Regressor	Coeff.	Std. Coeff.
Institutional Quality		-1.168***	-0.402
Stress History		0.610***	-0.116
Cyclical	Current Account Balance/GDP	-0.024**	-0.093
	REER (3-year change)	0.014**	0.076
	Credit/GDP gap (t-1) (if + ve)	0.090***	0.259
	International reserves/GDP	-0.032***	-0.215
Debt Burden	Δ (Public debt/GDP)	0.049***	0.109
	Public debt/revenue	0.002***	0.124
	FX public debt/GDP	0.024***	0.160
Global	ΔVIX	0.016***	0.147
	Share of currency union MACs in Stress	7.465***	0.146
Number of Observations			1581
LR chi2			264
Pseudo R2			0.266

Source: Fund staff calculations.
Note: Stars indicate statistical significance at the 1 percent (***) and 5 percent (**) levels. Standardized coefficients are scaled by variable standard deviations, thus providing a measure of relative importance (see full standardization in Long, 1997).

Acknowledging that the ongoing transformations in the governance of some currency unions (e.g. the eurozone) may address these risks, the default setting of the logit model mutes the CU variable. However, this can be switched on if country teams consider spillover risks within a CU a material risk.

B. In-Sample Performance of the Logit Model

8. **The overall in-sample performance of the model is very good, and a significant improvement compared to the heatmap in the existing framework.**

- The model's overall in-sample predictive capacity of stress/non-stress episodes is high, as illustrated by the fact that the distributions of fitted probabilities for stress and non-stress cases have limited overlap (Figure AV.6). Quantitatively, this discriminatory capacity is reflected in a high value of the Area Under the receiver operating characteristic Curve (AUC), 0.88, and a low minimum total misspecification error (TME, equal to the sum of missed crises and false alarms) of 37 percent, corresponding to a 9 percent probability of stress (the vertical blue line).[9]

- The improvement over the existing framework is substantial. For instance, the minimum TME of 37 percent reflects a missed crisis rate of 10 percent and a false alarm rate of 27 percent. In contrast, using an OR rule to combine the signals from the heatmap (crisis signaled when at least one of the heatmap indicators breaches its threshold), the existing framework has about the same missed crisis rate as the new framework (9 percent for EMs and 14 percent for AE crises) but a much higher false alarm rate (63 and 72 for EMs and AEs, respectively, implying TMEs of 72 percent and 86 percent, respectively).

C. In-Sample Performance in Individual Countries

9. **In-sample performance is very good in individual countries (Figure AV.2-5).**

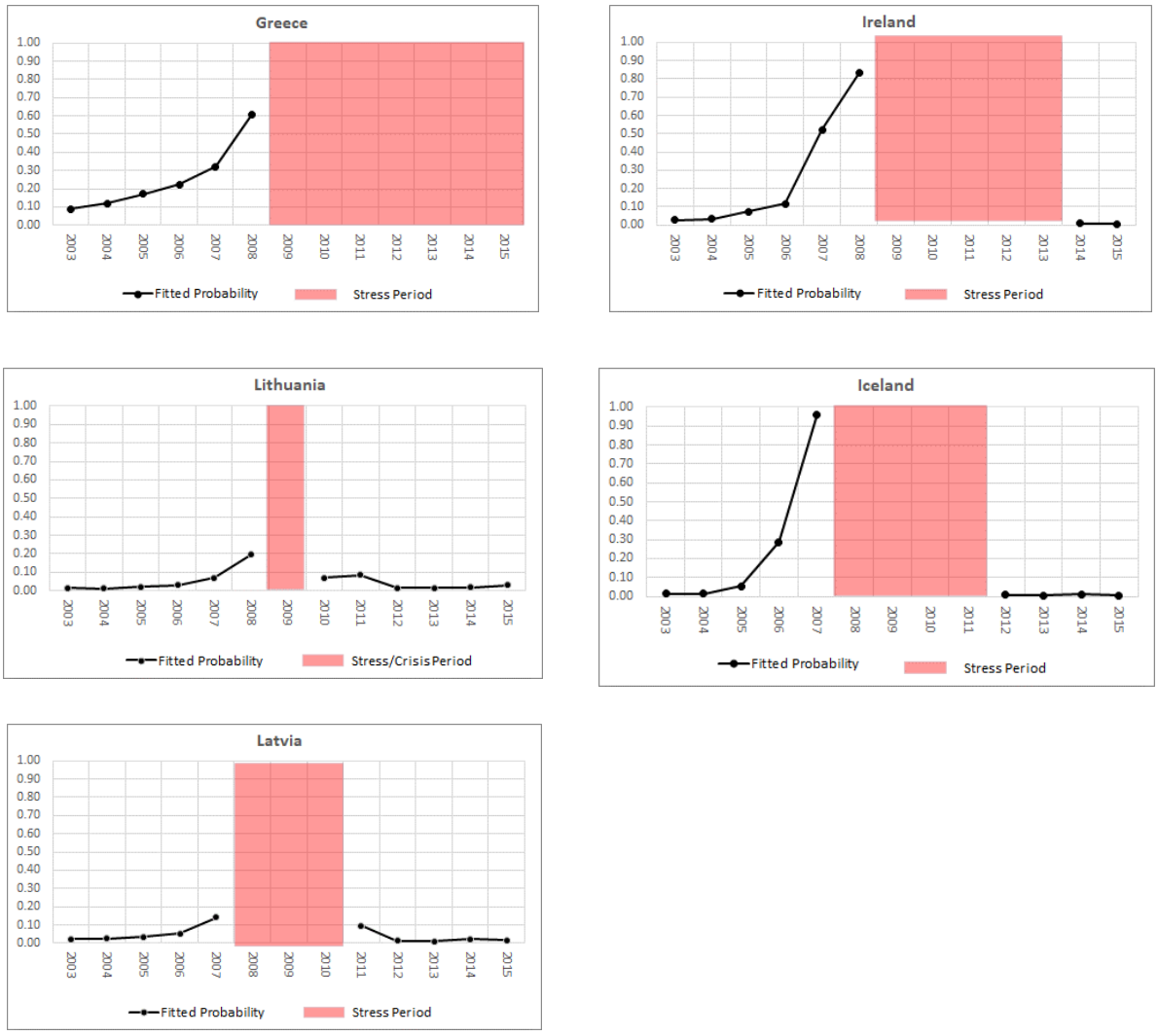
Predictive performance in countries that experienced stress due to regional spillovers is higher for the logit specification which includes the share of CU MACs in stress (Figure AV.3).

Predictive performance is weaker in countries that experienced sovereign stress due to episodes of political instability, which is hard to predict and is not captured by any of the regressors of the model, such as in MCD countries in years 2010-12 due to the Arab Spring or in Ukraine in 2014 due to the political crisis/revolution. This confirms the importance of judgement in the final near-term risk assessment (Figure IV.5).

[9]When using sufficiently long training periods, the performance of both models was found to be broadly comparable. Based on shorter training periods, the performance of the logit was weaker than that of the VE fiscal module, but still strong. Both models captured recent stress episodes well. As an additional consistency check, the estimated risk rankings (based on 2018 data) from the two models were compared and revealed a 0.83 correlation.

Figure AV.2. In-Sample Performance in Selected Countries which Experienced Stress

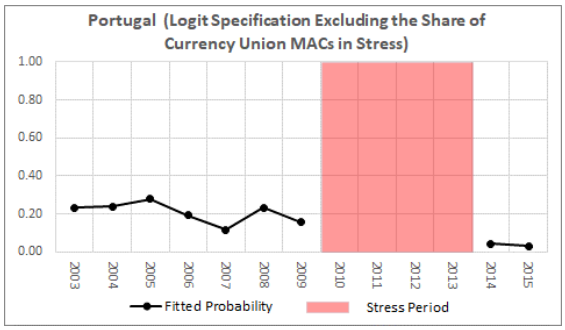
Stress Associated with GFC/post-GFC Fund Programs in Advanced Economies

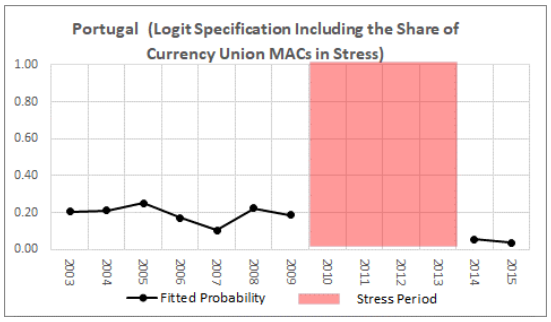


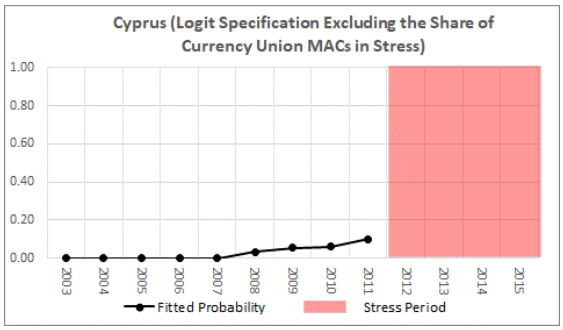
Source: Fund staff estimates.

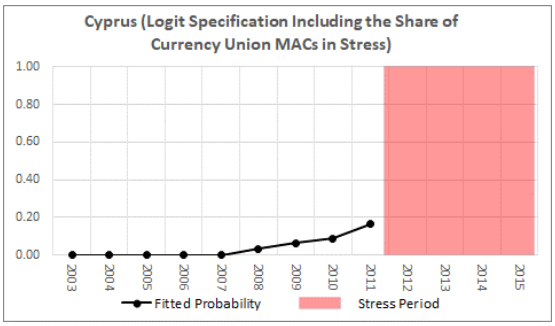
Figure AV.3. In-Sample Performance in Selected Countries which Experienced Stress

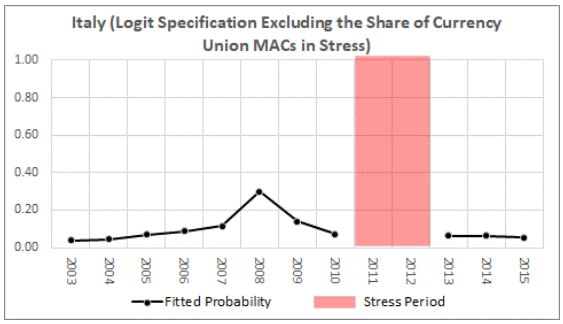
Stress Associated with GFC/post-GFC in Advanced Economies Exposed to Regional Contagion

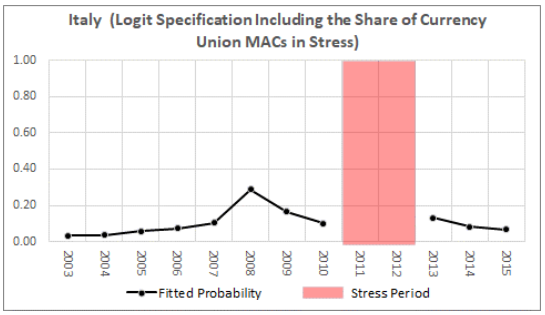


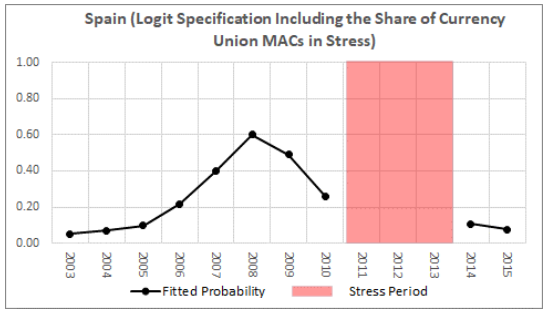


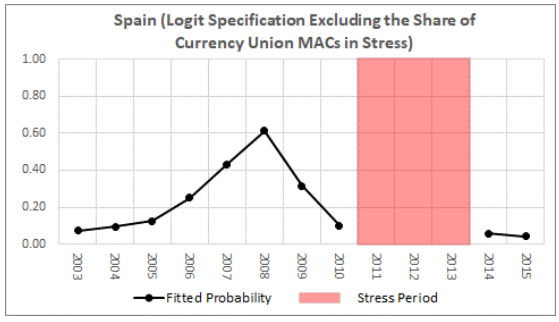






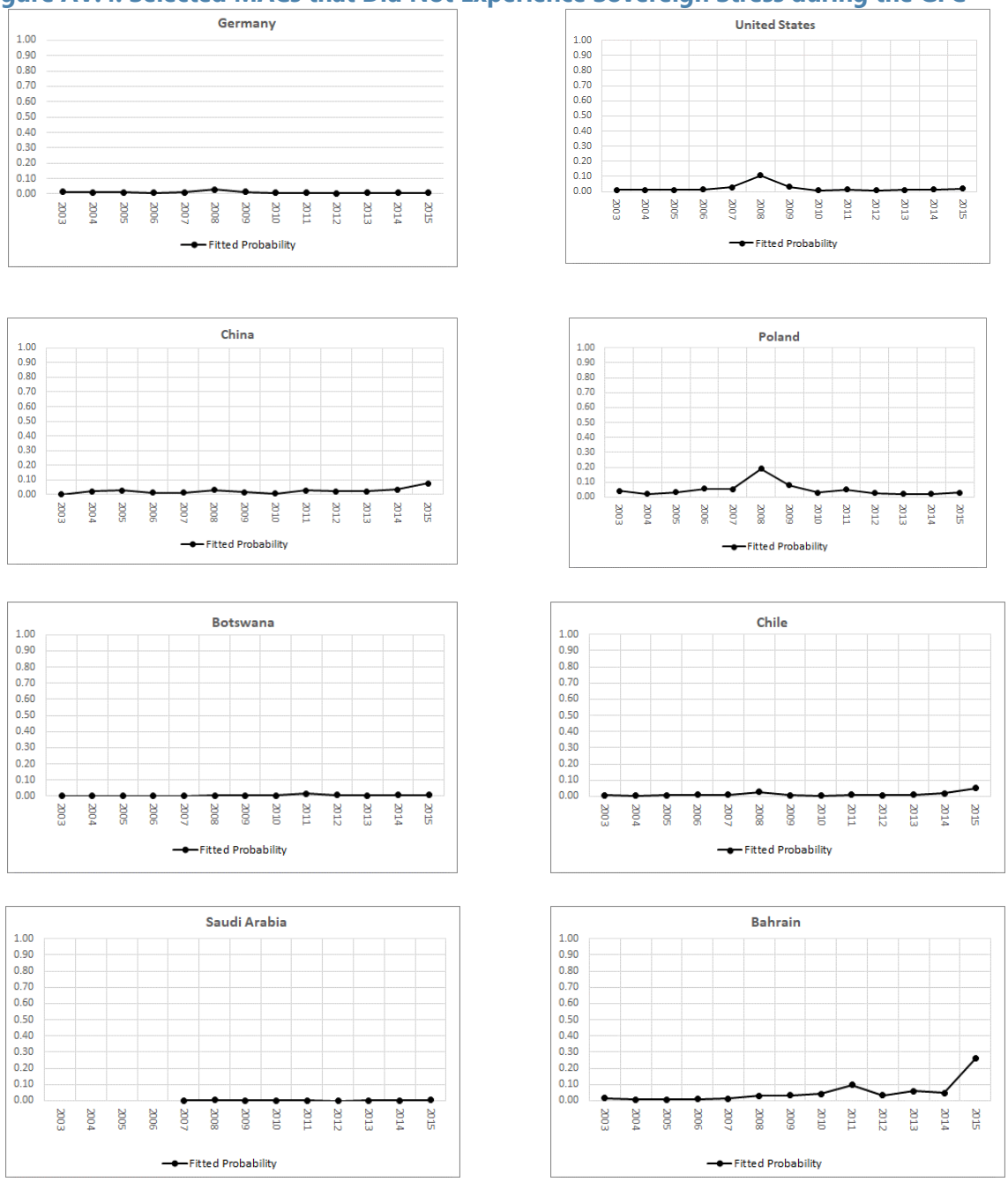






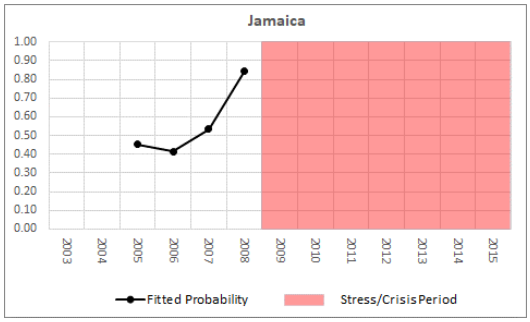
Source: Fund staff estimates.

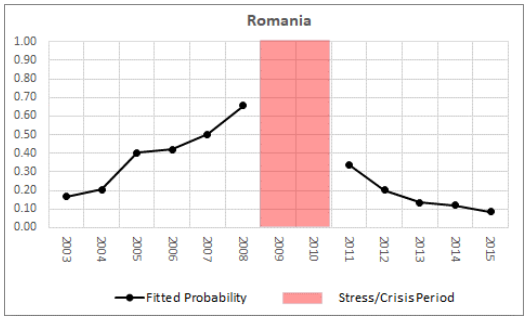
Figure AV.4. Selected MACs that Did Not Experience Sovereign Stress during the GFC

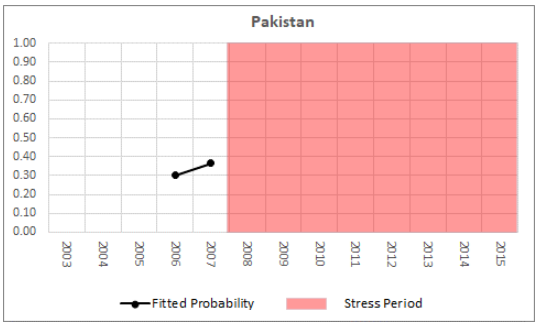


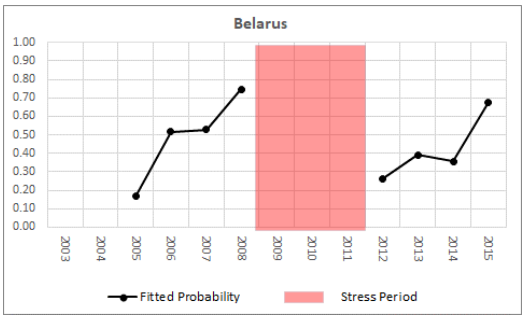
Source: Fund staff estimates.

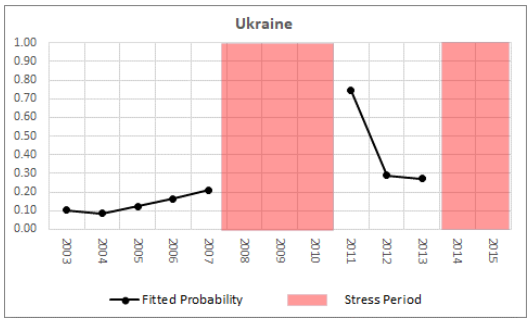
Figure AV.5. Selected Recent Stress Episodes in Emerging Markets

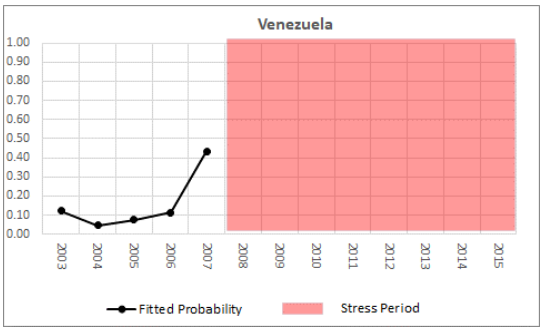


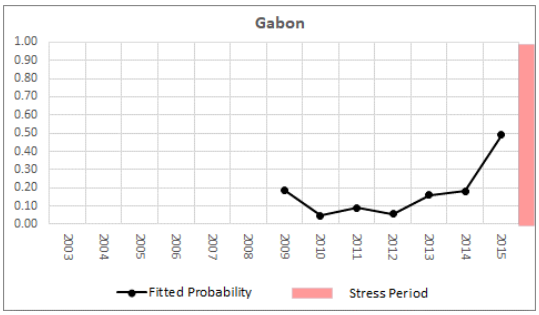












Source: Fund staff estimates.

D. Pseudo-out-of-Sample Performance and Robustness Checks[10]

10. **The revised specification also performs well pseudo-out-of-sample tests, using different temporal cutoffs.**

Testing predictive performance out of sample requires "training" (estimating) the model on a certain time period and then testing it on the remaining time period. Two alternative training (estimation) samples were chosen: i) from 2000 to 2015, and ii) from 1990 to 2012; with corresponding test (i.e., "out") samples 1990–99 and 2013–15, respectively. The selected time cutoffs shed light on whether the specification does a good job in predicting the earliest and latest stress episodes in the sample (e.g., Asian crisis in the 1990s and stress in commodity exporters after 2014). The period of the GFC was included in both training samples because this is the only period when AEs faced stress, thus containing unique information not available in other parts of the sample. Performance in terms of missed crises and false alarm rates and minimum total misspecification error is robust in the test (out-sample) periods under both cutoffs (Table AV.3).[11]

Table AV.3. Pseudo out-sample Performance under Different Training Samples

Training Sample	Test Sample	AUC	Loss Function Minimization			# of Stress Episodes
			Minimum Total Misspecifcation Error (TME)	*of which: Missed Crisis Rate*	*of which: False Alarm Rate*	
2000--2015	1990--1999	0.86	0.38	*0.19*	*0.19*	*12*
1990--2012	2013--2015	0.88	0.39	*0.12*	*0.27*	*15*

Source: Fund staff estimates.

Note: The model, based on the baseline specification in Box AV.2 (i.e. including the CU variable), is re-estimated on the training sample and, then, its performance is verified in the test sample in terms of AUC and minimum Total Misspecification Error (TME) (and corresponding missed crisis and false alarm rates). The TME is the sum of the probabilities of type I and type II errors. The minimum TME provides information on the discriminatory capacity of the corresponding tools based on a single threshold that divides the space of possible results in two zones (high risk, predicting a crisis; and low risk, predicting no crisis).

11. **As an additional test, staff compared the out of sample performance of the logit with the performance of the Fiscal Module of the IMF's Vulnerability Exercise (VEFM), which uses a**

[10]The difference between out-of-sample and pseudo-out-of-sample analyses rests on the fact that in a pseudo out-of-sample exercise a model is first specified using the entire sample (in this case, 1990-2017) and then re-estimated on a sub-sample (the "training sample") in order to evaluate its predictive performance in the remaining sample (the "test sample"). In contrast, in a pure out-of-sample exercise, the training sample is used to both specify and estimate the model before of its out-of-sample predictive performance is examined.

[11]To check the robustness of the specification, staff has estimated the model exclusively on EMs to see if estimating the model on the full (including AE) sample biases results for the EM subgroup. The coefficients of all variables maintain the same sign and magnitude in an EM-only sample; only the current account coefficient loses significance, as many non-commodity EMs entered periods of stress when the external imbalances, recorded for many years before the stress episode, were actually correcting

sovereign stress prediction model based on machine learning.[12] The VEFM delivers even better out-of-sample predictive performance than the logit, particularly when estimated over shorter sample periods. However, in Staff's view, this is offset by the greater transparency and easier economic interpretability of the logit model (as shown table AV.2):

- Using long estimation periods, the performance of the logit model was found to be almost as good as that of VEFM: when "trained" over a 1990-2012 period, the AUC for the logit was 0.88 compared with 0.90 for the VEFM (trained over 1980–2012). The difference in predictive performance rises when both models are "trained" over shorter periods. Estimating the logit on the 1990–2005 period leads to an AUC of 0.73 for the logit compared with 0.82 for the VE model (estimated over 1980-2005). Both models captured recent stress episodes well.

- The proposed logit is simple and easy to communicate. By comparison, the output of the VEFM, based on a "Random Forest" (RF) model, is less amenable to policy discussions, as it is based on a very large number of variables (above 100) including interaction effects that may not be straightforward to explain/interpret.

Although the logit will be the main workhorse for near-term risk analysis, the VEFM—due to its high predictive performance, and possible complimentary insights—would be made available to teams to inform their final judgment-based assessment on near-term risks.

12. **The data was checked carefully for outliers.** Large regressor values (for example the very large surplus in the CA of Gulf countries, or the very large credit-to-GDP gap in countries that experienced a financial crisis) were all cross-checked and validated in the data. In addition, staff ran the specification with the top and bottom 1 percentile removed (263 observations). The results of this analysis confirmed the magnitude and signs of estimated coefficients.

E. Performance of the Proposed Mechanical Signals

13. **The logit stress probability (LSP) predicted by the model is divided into three risk zones (high, moderate, low) based on the probability of missed crises and false alarms** (see ¶31, 32, 48 and Box 3 of the main paper). Low- and high-risk cutoffs are calibrated to keep the rate of missed crises and false alarms at 10 percent, respectively (Figure AV.6). The corresponding stress probability cutoffs are 9 percent (at the threshold between the low and moderate risk signal) and 20.5 percent (at the threshold between the moderate and high risk signal), respectively. The *average* stress probability based on the historical sample is 40 percent for a country whose fitted probability signal is "high", compared to 16 percent and 2 percent for "moderate" and "low" risk countries, respectively.

14. **As a plausibility check of the model's predictive performance, the risk signals generated by the model ahead of selected well-known stress episodes are reported in Table AV.4.**

[12]See IMF, 2020, *How to Assess Country Risk: Vulnerability Exercise Approach Using Machine Learning.*

With only one exception—stress in Jordan in 2012, associated with political uncertainty connected to the Arab Spring—the model flagged risks in advance in the form of a moderate-risk or high-risk signal.

Figure AV.6. Distribution of Stress and Non-Stress Outcomes

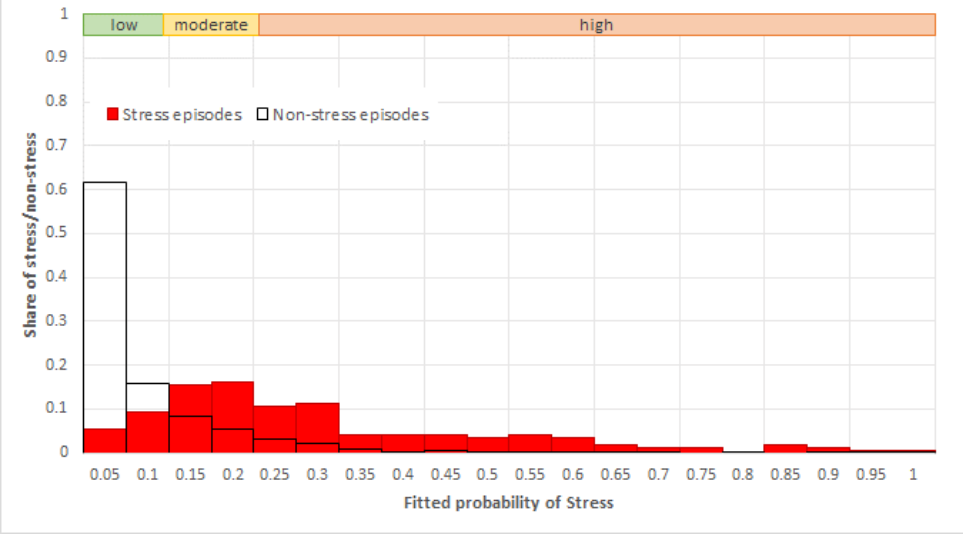


Table AV.4. Signal Derived with the Proposed Decision Rule in Selected Stress Episodes

Country	*Onset of Stress Episode*	*Signal 1 year before stress*	*Onset of Stress Episode*	*Signal 1 year before stress*
Italy	2011	Moderate		
Portugal	2010	Moderate		
Spain	2011	High [1]		
Cyprus	2011	Moderate		
Greece	2009	High		
Iceland	2008	High		
Ireland	2008	High		
Latvia	2008	Moderate		
Lithuania	2009	Moderate		
Egypt	2011	Moderate	2016	High
Lebanon	2007	High	2011	Moderate
Ecuador	2008	High	2015	High
Antigua and Barbuda	2008	Moderate	2016	High
Belarus	2009	High		
Hungary	2008	High		
Jamaica	2009	High		
Jordan	2012	Low		
Romania	2009	High		
Tunisia	2013	Moderate		
Ukraine	2008	High	2014	High
Venezuela	2008	High		
Angola	2015	High		

Source: Fund staff estimates.
[1]Risk signal generated by the specification that includes the currency union variable (see Box AV.2). If the variable is excluded, the risk signal drop to "moderate".

F. Customization of the Logit Tool in Special Cases

15. **Guidance will be provided to address some special cases.**

- In commodity exporters, where GDP is more volatile, large increases in the credit-to-GDP gap could be due to GDP shrinking rather than to credit to the private sector increasing, thus introducing noise in the signal issued by this regressor. In those cases, it could be warranted to use the credit to non-oil to GDP ratio to compute the gap.

- In countries with large foreign assets in a SWF, a customized approach would allow for the inclusion of the share of those assets that are liquid and readily available in case of stress in the model's 'FX reserves' variable. Guidance will discuss how to handle situations where a clean accounting of liquid assets is not available.

Some countries (e.g. safe havens, or countries with very low near-term external financing needs) may be less vulnerable to changes in global risk appetite, proxied in the model by changes in the VIX.[13] Guidance will be provided to deal with situations where the VIX movements (positive or negative) alone are seen to drive a change in the mechanical risk signal for such countries.

[13]The impact will not be nil, as changes in the VIX can also provide a signal on expected real economic activity, which can affect countries via real (rather than purely financial) channels, such as changes in trade and foreign direct investment.

Table AV.5. Logit Regressors' Summary Statistics
(this excludes variable values observed during stress episodes)

1557 observations	Institutional Quality	Stress History	Current Account Balance/GDP (percent of GDP)	REER (3Y change), percent	Credit to Private Sector Gap Lag (only positive), percent of GDP	Total international reserves (percent of GDP)	GG Debt (Change), percent of GDP	GG Debt, percent of government revenue	Foreign Currency Public Debt, percent of GDP	VIX, Index 2010=100, Annual, Change	Share of currency union MACs in Stress
min	-1.60	0.00	-90.32	-73.13	0.00	0.18	-79.10	1.24	0.00	-39.60	0.00
p1	-1.38	0.00	-23.29	-27.54	0.00	0.93	-13.66	7.55	0.00	-39.60	0.00
p10	-0.45	0.00	-7.83	-11.24	0.00	3.76	-4.74	44.32	0.00	-28.40	0.00
p25	0.01	0.00	-3.95	-4.66	0.00	6.65	-2.22	82.78	0.00	-15.83	0.00
p50	0.62	0.00	-0.67	0.62	1.22	13.82	-0.09	144.59	3.60	-3.76	0.00
p75	1.21	0.21	2.61	5.27	5.40	20.07	1.89	200.79	14.05	10.80	0.00
p90	1.75	0.90	9.87	14.87	13.09	34.24	5.63	311.11	31.71	24.61	0.00
p99	2.03	1.98	31.84	40.97	31.32	91.32	14.24	669.36	72.78	67.22	0.29
max	2.25	3.60	45.46	95.86	88.60	118.21	25.51	783.05	136.90	67.22	0.35
sd	0.84	0.47	9.20	12.66	6.97	17.23	5.37	123.83	16.18	22.38	0.05
mean	0.64	0.25	0.28	1.46	4.49	17.94	0.11	168.42	11.12	-0.45	0.01

Annex VI. Technical Notes on the Debt Fanchart

This annex describes the two-step procedure used to generate the new debt fanchart and discusses the three metrics that are derived from it. It also describes how the overall index was defined and backtested.

A. Fanchart Methodology

1. **Staff propose a two-step procedure to derive an improved debt fanchart that would replace both the current fancharts and the standardized macro-fiscal stress tests.** The new procedure applies a high-level realism check and imposes a "realism-adjustment" when risks to the debt projections appear to be heavily skewed. This addresses a major shortcoming of the current fancharts—namely, that their direction depends entirely on the baseline. Even when the baseline passes the realism check, fancharts no longer assume a normal distribution around the baseline. Instead, they are constructed based on the historical shocks of the debt drivers, resulting in a fanchart that is generally asymmetric.

2. **In the first step, the team's baseline would be compared with a "historical fanchart".** The latter is generated by drawing stochastic realizations of the debt drivers from their joint empirical distribution (to capture the correlations across debt drivers). To capture the inter-temporal dependence in the data, the stochastic realizations of the debt-drivers are drawn using a "block-bootstrap" approach, in which draws from the historical distribution are taken for consecutive two-year "blocks". [1] The historical fanchart produces a stochastic version of the existing historical scenario. [2] Since it is independent from the team's baseline, this historical fanchart can be used to diagnose baseline realism. When the team's baseline debt path falls *below* the 20th percentile of the historical debt fanchart, the baseline would be assessed as unlikely to represent an adequate balance of risks and further scrutiny would be required.[3]

3. **The second step produces the final fanchart, based on the results of the first step:**

   i. If the team's baseline debt path does not fall below the 20th percentile of the historical debt fanchart, the second step generates a "standard" fanchart (Figure AVI.2, country 1).

[1]Specifically, a specific two-year "block"— that is, two consecutive annual realizations of the debt drivers (growth, the primary balance, interest, etc.) is randomly drawn from the 1990-2018 sample period. The first annual realization of the drivers is substituted into the debt stock-flow equation to generate a predicted debt ratio at time t, conditional on debt at time t-1 (the most recent realization). Conditional on the debt ratio at t, the second annual realization of debt drivers from the block is used to compute debt at t+1. Debt at t+2 and t+3 are computed similarly, based on a newly drawn two-year block. Finally, debt at t+4 and t+5 are computed based on a third draw. This process generates one debt path between t and t+5. To "populate" the fan chart, the process is repeated 10,000 times.

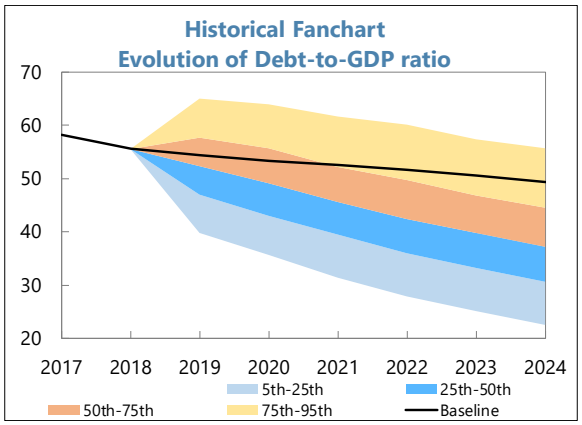
[2]Uncertainty about the *initial* level of the debt-to-GDP ratio is also incorporated by appealing to the historical WEO debt data revisions for the country. This adjustment for base effect risk was proposed in place of the initial proposal of using stock-flow-adjustment (SFA) shocks, which was dropped in light of concerns about SFA data quality and the perceived challenges of calibrating appropriate shock SFA distributions. Note that risks from potential contingent liabilities are now addressed in the triggered stress testing module.

[3]The same consideration could apply for debt paths above the 80th percentile, although evidence on forecasts suggests this is a rarer occurrence.

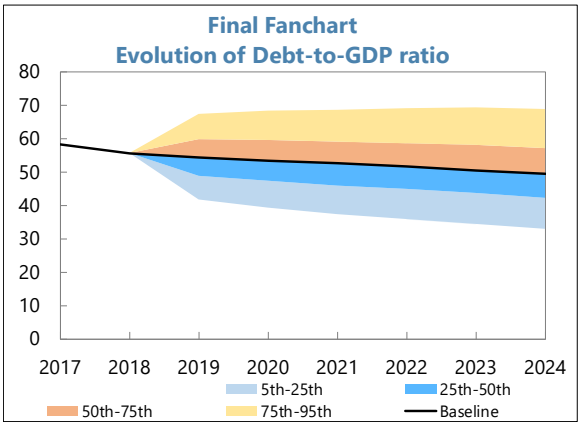
In this case, the team's baseline would be assessed as sufficiently realistic and representing an adequate balance of risks. The forward-looking information included in the baseline fully determines the (upward/horizontal/downward) "direction" of the fanchart; while its width and skew is determined by that of the historical fanchart.

Figure AVI.2. Application of the New Fanchart Methodology

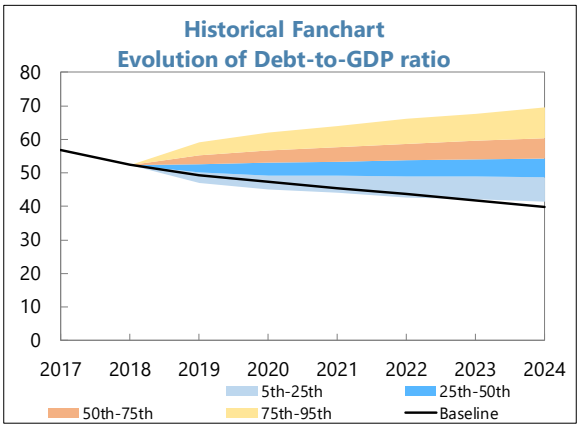
County 1 (2019): Baseline above 20th percentile of historical fanchart



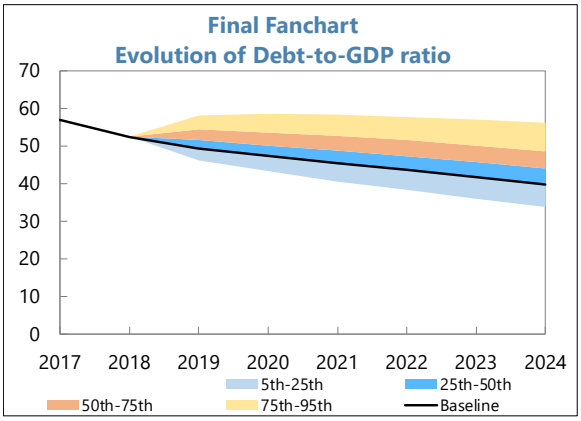
Country 1 (2019): Final fanchart obtained by applying de-meaned shocks to baseline (standard fanchart)



Country 2 (2019): Baseline below 20th percentile of historical fanchart



Country 2 (2019): Final fanchart obtained by applying skewed shocks to baseline (asymmetric fanchart)



Source: Fund staff estimates.

ii. If, even after further scrutiny, the team's baseline continues to fall *below* the 20th percentile of the historical fan in any projection year (Figure AVI.2, country 2), the deviation between the team's baseline projection for debt and the level implied by historical trends would be compared with the historical cross-country distribution of this metric (Box AVI.1) for relevant peers,[4] to determine the country's percentile. A final, "realism adjusted" fanchart would then be constructed by adding skewed shocks to the

[4]Countries are grouped into three groups for this peer-based analysis: Advanced Economies, EM commodity exporters, and EM non-commodity exporters.

underlying debt drivers, moving the distribution to the right until the (fixed) team's baseline falls just as far on the lower tale (same percentile) of the fanchart distribution as it does in the cross-country distribution.[5]

Box AVI.1. Assessing the Need for Adjustment in the Central Projection of the Fanchart

If the team's baseline debt projections fall below the 20th percentile of the historical fanchart, then the final fanchart would generally not be centered on the team's baseline, as this suggests baseline optimism compared with historical trends. Additional scrutiny would be applied by comparing the projected deviation of the team's baseline from the historical trend with the historical distribution of such deviations for all MACs. The central tendency of the fanchart would then be adjusted so that the team's baseline falls just as far in the lower tail (same percentile) of the fanchart distribution as it does in the distribution of deviations from historical trends for all MACs.

Formally, the template will compute for each projection horizon *j* (with *j=0,1, …,5*) the following distance:

$$d^{p}_{x,j} = debt^{p}_{x,t+j} - \overline{debt_{x,t+j}}, \quad \forall\, j \in \{0,1,\dots,5\}$$

where $debt^{p}_{x,t+j}$ is the team's debt projection for country *x* at time *t+j* and $\overline{debt_{x,t+j}}$ is the debt projection (or historical trend) derived by using the debt dynamic equation with debt drivers set equal to their 10 year average at time *t*. The largest distance d^{max}_{x}=max($d^{p}_{x,0}$, $d^{p}_{x,1}$,…., $d^{p}_{x,5}$) will then be compared to the distribution of actual departures from historical trends at the corresponding projection horizon to derive the percentile ($\bar{p}$) corresponding to d^{max}_{x}.

The distributions of departures from historical trends at projection horizon *j* (with $\in \{0,1,\dots,5\}$), in turn, will be derived using historical data for the period 2010-2019, computing for each country *c* and year $t \in \{2010,2011,\dots 2019\}$ the following distance:*

$$d^{a}_{c,j} = debt^{a}_{c,t+j} - \overline{debt_{c,t+j}}, \quad \begin{cases} \forall\ t \in \{2010,2011,\dots 2018\} \\ \forall\ c \text{ in the MAC sample} \end{cases}$$

where $debt^{a}_{c,t+j}$ is the actual debt realization for country *x* at time *t+j* and $\overline{debt_{c,t+j}}$ is the debt projection derived by using the debt dynamic equation with debt drivers set equal to their 10 year average at time *t*.

It is worth noting that, while for country *x*, under assessment, the distances $d^{p}_{x,j}$ are computed using the team's debt projections, the distribution of departures from the historical trend is derived using the distances $d^{a}_{c,j}$ computed using the actual debt realizations in the MAC sample.

The central projection of the fanchart would then be shifted upward until the team's baseline lies exactly at the $\bar{p}$ percentile of the fanchart distribution. This is achieved by adding skewed shocks to the underlying debt drivers until the team's baseline coincides with the $\bar{p}$ percentile of the fanchart distribution.

4. **In special cases, when there is a strong reason to believe that past dynamics are less relevant, an exit clause from the "asymmetric fanchart" would be introduced.** To avoid excess discretion, staff will provide clear guidance on when to apply the escape clause. Possible situations for this exemption would be rare and could include restructuring cases, and the deviation from the standard methodology in these cases would need to be clearly explained.

[5]The minimum setting is the 10th percentile, to avoid cases where the team's baseline falls outside the fanchart.

5. **A modified version of the historical fanchart is proposed for the recovery phases of the Covid-19 pandemic (2021-22) to limit the number of instances where the optimism correction mechanism is incorrectly triggered.**

- Since many country teams will project a significant (atypical) decline in debt-to-GDP during the recovery phase, the historical fanchart, which relies completely on past data (where large debt reductions were rare), may incorrectly flag baseline optimism in many cases. This issue will be particularly relevant for the first few years following the approval of the proposed framework (2021-22).

- To address this issue, staff has considered a modified historical fanchart which uses team's baseline debt projections for the first two years as its central tendency and the standard historical fanchart data generating process after that point. By giving credit to baseline debt projections during the first two years (recovery years), the modified historical fanchart would limit the number of instances where the optimism correction mechanism is incorrectly triggered.

- Staff has performed a test and generated 2021 fancharts as if we were already in 2021 and found that the realism correction would be applied in only 9 cases when using a modified historical scenario which uses baseline debt projections for 2021-22 as its central tendency.[6] This contrasts with 22 cases when using the standard historical scenario.

6. **The new fanchart methodology maintains some of the simplifying assumptions underlying the current methodology.** In particular, we continue to assume (i) no feedback between the debt drivers and the level of debt (ii) that interest rates on domestic- and foreign-currency debt face the same shock distribution (calibrated based on the past behavior of average effective interest rates), (iii) that the foreign currency debt shares are non-stochastic (fixed at the baseline projections). The first and third of these assumptions imply that the uncertainty expressed in the fanchart understates the true uncertainty (i.e. the fancharts will be too narrow). In particular, the upper percentiles of the fanchart will be missing some explosive debt paths, in which higher debt and rapidly widening spreads create a snowball effect.

7. **For the purposes of this review, these assumptions are justified for two reasons.** First, addressing these points would add an additional layer of complication to an already very ambitious reform. In particular, a proper modeling of the feedback between debt and interest rates is beyond the present research frontier. While DSAs at the Fund and elsewhere have sometimes used simple linear feedback rules, these offer only a modest improvement over ignoring the feedback altogether, as they do not capture the sharply non-linear rises in borrowing spreads when markets begin to view debt as unsustainably high. Second, while the fancharts understate the true uncertainty, this does not affect the predictive capacity of the fanchart tool. As explained in the next two paragraphs, a risk signal is derived by combining several fanchart-based metrics, including the width of the fanchart, into an index, and comparing index values with low- and high-risk thresholds based on the

[6]Since WEO projections end in 2025, the 2026 values for the debt drivers in this exercise were set at their 2025 levels for simplicity and the corresponding debt levels were obtained by feeding these drivers into the debt dynamics equation.

probabilities of missed crises and false alarms associated with each index value. While a wider fanchart would lead to higher index values, it would also lead to higher thresholds.

B. Fanchart Metrics and Predictive Performance

8. **Staff has analyzed the discriminatory (predictive) power of various candidate metrics using the 2010–15 fans.**[7] Four broad categories of metric were considered, reflecting: (i) probability of debt stabilization over the projection horizon; (ii) probability of long-term debt stabilization; (iii) uncertainty around the debt projection; and (iv) the projected level of debt. To assess potential discriminatory power, staff assessed the ability of indicators (constructed from the 2010–15 'real-time' fancharts) to 'predict' episodes of sovereign stress occurring in subsequent years (1–5 years ahead). Three metrics have both a strong intuitive appeal and demonstrated encouraging performance, as illustrated by the "receiver operating characteristic" (ROC) curve,[8] over the backtesting period (Figure AVI.2): the probability that the debt does not stabilize in the medium-term;[9] fanchart width; and debt level at t+5, interacted with an index of institutional quality, as a way of capturing debt-carrying capacity.

9. **While each of these metrics can be used to predict sovereign stress individually, their discriminatory power is even greater when used in combination (Figure AVI.2).** Consequently, the three metrics will be aggregated into a composite Debt Fanchart Index (or "DFI") that weights each metric by its predictive power and can be used to classify countries into risk groups. The distribution of this aggregate index differs markedly for crisis and non-crisis episodes (see Figure AVI.3), indicating a strong discriminatory capacity. Quantitatively, the aggregate index has an AUC of 0.82 and a minimum TME of 43 percent, corresponding to an index value of 0.32 (vertical blue line). Following a similar approach as for the logit model, three risk zones (low-, medium-, and high-) can be derived by calibrating a low- and a high-risk threshold for the fanchart index such that the latter is associated with a 10 percent missed crisis rate and the former with a 10 percent false alarm rate.

10. **The relationship between these risk ratings and the likelihood of stress can be examined by estimating posterior stress probabilities.** While the level of the DFI is not a direct estimate of the probability of a "stress" event, estimates of the posterior probability of stress at each level of the DFI can be derived empirically based on the share of countries within a given DFI range that went on to experience stress in sample. Figure AIV.3 depicts such estimates for each of 20 "bins" (intervals) of the DFI. While the limited number of "stress" observations mean that these probabilities can only be estimated imprecisely (particularly at higher values of the DFI where there

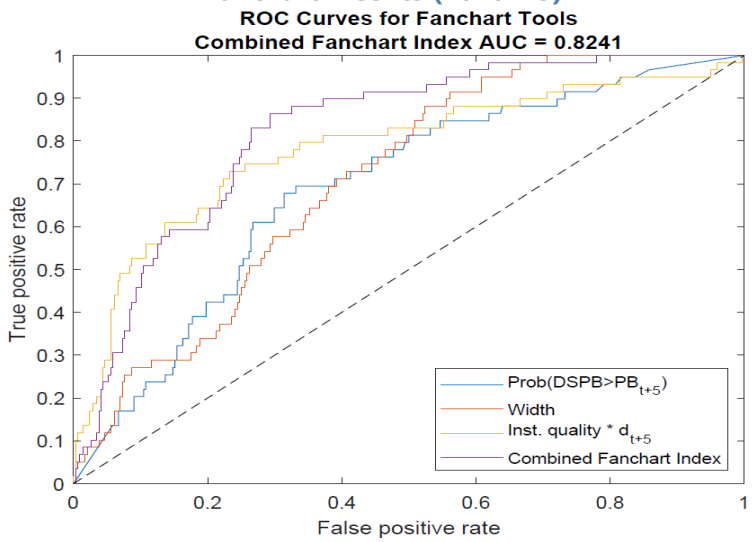
[7]Although fanchart metrics and signals are available for more recent period, the stress outcomes associated with these signals cannot be observed for the full medium term (5-year) prediction period. Hence, predictive performance is analyzed based on fanchart signals between 2010 (the earliest period available) and 2015.

[8]The ROC curve plots the share of correctly predicted crises (y-axis) against the share of false alarms (x-axis) for all possible thresholds. The further the ROC curve for a metric lies above the 45-degree line, the better the ability of that metric to distinguish crisis and non-crisis events.

[9]This metric also accounts for the link between sustainable debt levels and the outlook for the primary balance and interest-growth differentials. This probability can be inverted to give the likelihood that the baseline adjustment would be sufficient to put debt on a declining path. An alternative approach would be to focus on the probability that the primary surplus will be sufficient to achieve a given degree of debt reduction, but this would require taking a stance on whether debt is currently at a level from which it needs to be reduced, and the appropriate pace of debt reduction.

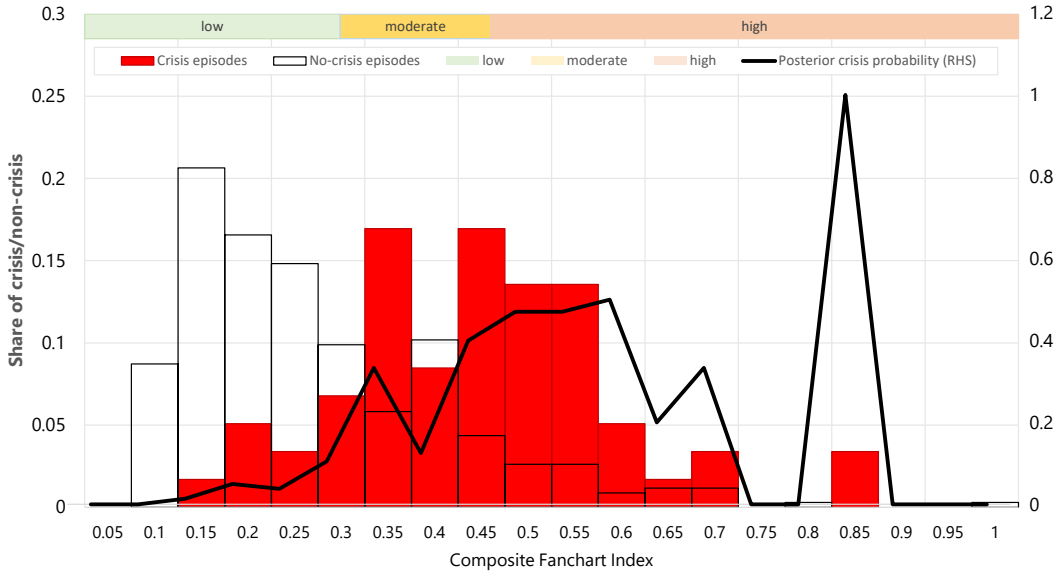
are fewer observations), there is a clear pickup in the posterior stress probabilities around the "low/moderate" and "moderate/high" thresholds. The figure suggests a posterior probability of stress conditional on a GFI "high risk" signal of at least 40 percent, and a posterior probability of stress conditional on a GFI "low risk" signal of at most 10 percent. The average posterior probability of stress for each of the three proposed risk zones is 44 percent for a "high risk" signal, 23 percent for a "moderate risk" signal, and 3 percent for a "low risk" signal.

Figure AVI.2. Receiver Operating Characteristic Curve for Fanchart Metrics (2010–15)



Source: Fund staff calculations.

Figure AVI.3. Debt Fanchart Index Distributions in Stress and Non-Stress Cases (2010–15)



C. Customization of the Fanchart Tool in Special Cases

11. **Guidance will specify some adjustments to the standard methodology in special cases.**

- When the public sector holds large financial assets, for example in a stabilization or sovereign wealth fund (SWF), government solvency is typically stronger than would be suggested by the standard gross debt fancharts, since the sovereign can neutralize explosive debt paths by drawing down on the assets. Staff does not view incorporating these effects automatically in the construction of the fan chart as feasible due to data limitations. However, guidance could ensure that these factors are appropriately accounted for in the mechanical risk signals.[1]

- A second set of special cases are countries which have experienced obvious structural breaks. As in the ongoing restructuring cases discussed above, the fanchart's 'realism-adjustment' would not be appropriate, and the associated metrics would need to be based on the "standard" (symmetric) fan. Such situations would be expected to be rare but could include a major crisis in the past that is not expected to be repeated in the future; a major natural resource discovery or depletion in the projected horizon relative to the past; or regime changes like accession to a currency union. Guidance will flesh out how an escape clause to the "asymmetric fanchart" can be introduced for these cases.

- A third set of special cases are countries that are close to reaching a debt restructuring agreement. In these cases, it would be incorrect to apply the realism mechanism since the outcome of the debt restructuring is a debt reduction going forward. Therefore, staff proposal is to not apply the realism correction in those cases but to build the fanchart around the team's baseline by default. Moreover, to account for the fact that the volatility of the effective nominal interest rate is likely to be lower post-restructuring, this volatility could be scaled down by a factor corresponding to the ratio of new and past debt issuances.

12. **Staff also considered whether a fanchart adjustment was warranted in countries with past debt restructuring experiences, and concluded that is not the case.** In those cases, the question is whether past volatility of debt drivers remains a good guide for the future. Staff looked at a sample of recent restructuring cases to assess whether dropping the restructuring years from historical data would lead to a material change in the width of the 2020 fancharts. Except in the case of Greece—where the restructuring years were associated with a large recession—the width did not significantly decline, or even slightly increased, when the restructuring years were dropped from the historical sample (Table AVI.1).[2] These results support the idea that past volatility remains a good

[1]For instance, staff has identified seven MACs (Brunei, Kuwait, Norway, Qatar, Saudi Arabia, Singapore and the UAE) where SWF assets are in excess of both 100 percent of gross debt and 100 percent of GDP; it would seem reasonable to assign a low risk fanchart signal to these cases. For other countries, with assets that are significant but below one or both of these thresholds, team judgment appears better placed to assess the liquidity and availability of the assets (in other words, the mechanical fanchart signal could continue to be based on gross debt for these countries, but the overall medium-term risk *assessment* could be adjusted, as appropriate, by country teams).

[2]The fact that the width tends to increase when the restructuring years are dropped can be traced back to the fact that those years are generally associated with positive primary balance shocks, as the authorities undertook fiscal adjustment to show their commitment to restore debt sustainability. These tend to offset negative growth shocks and hence dampen the debt dynamics.

guide for future volatility for countries having experienced debt restructuring in the past and no correction to the fanchart methodology is needed in those cases.

Table AVI.1. Impact of Restructuring Years on Fanchart Widths

Countries	Restructuring years	Standard width	Width after dropping	Difference
Antigua and Barbuda	2010-11	67.2	68.2	**1.0**
Barbados	2018-19	48.8	47.7	**-1.2**
Belize	2012-13	31.4	32.4	**1.0**
Greece	2011-12	87.2	73.1	**-14.0**
Jamaica	2013-14	31.5	32.8	**1.3**
St. Kitts and Nevis	2011-12	47.0	48.8	**1.8**
Ukraine	2015-16	61.4	64.7	**3.4**
Average	**--**	**53.5**	**52.5**	**-1.0**
Median	**--**	**48.8**	**48.8**	**-0.1**

Source: Fund staff estimates.

Annex VII. Technical Notes on the GFN Module

This annex describes the new data requirements for the GFN module, explaining how centralized databases can limit resource implications. It also describes the generalized stress scenario including the implementation of macro-fiscal, financing, and debtholder shocks. Finally, it describes the composite index's construction and predictive performance.

A. Data Requirements for GFN Analysis

1. **The proposed GFN module creates several new data requirements.** In some cases, these new inputs will have a resource implication for Fund staff. Care has been taken to try to minimize this burden, including by relying on standardized cross-country databases (e.g., *Fiscal Monitor*, *International Financial Statistics*, and a centralized debt holder profile databases), which should limit the new effort required from individual country teams. However, estimates of amortization by debtholder is a key ingredient into the GFN analysis, and in many cases, it would be helpful to refine these further (Box AVII.1).

Box AVII.1. Debt Amortization by Debt Holder

Staff has used the holder profile of the *stock* of debt (*a la* the Arslanalp-Tsuda methodology) and the maturity profile information in country DSA files to produce working estimates of the holder profile of debt amortizations for almost all MACs. This is a key input into the proposed GFN module for analyzing rollover risks (see section under medium-term tools below). It would be useful, however, for country teams to go beyond these current working estimates and be able to enter more accurate information by the time the framework goes live in early 2021. Staff view this as feasible with the cooperation of from country authorities. Some pointers on how this data could be gathered follow.

- With respect to **external debt holders**, *total* external amortization on existing debt is already compiled by country authorities and widely available in IMF-World Bank databases (this is indeed an essential input into the financial account of the balance of payments). *Private* external amortization can, in principle, be calculated as the difference between total external amortization and amortization on non-marketable obligations (loans, swaps etc.) to *official* creditors, which should be available to country authorities (the COFER database could be enhanced to identify the maturity profile of marketable debt held by foreign central banks, which would otherwise show up in private external holdings).

- Turning to **domestic holders**, amortizations to the *domestic central bank* should be readily available. Amortization due to *domestic commercial banks* could be collated from banking surveys insofar as these contain information on the maturity profile of banks' government securities holdings. An alternative would be the country's securities registry which should be able to identify how much of each outstanding security is held by domestically commercial banks. With this in place, amortization due to the *domestic nonbank sector* obtains as the residual.

It is important to recognize that even with the most accurate data, holder profile data can, by definition, never be pinned down exactly, as marketable debt can change hands over time. Thus, an accurate breakdown as of end of last month may not hold today. This said, even approximate holder profile data is critical for sovereign risk analysis, and hence warrants a serious data effort.

2. **To deepen the analysis of risks, financing assumptions, which form the centerpiece of GFN analysis, would be expanded beyond the current differentiation between domestic and**

external financing. Users will be asked to allocate domestic issuance among the central bank, commercial banks, and other resident sectors and divide external debt issuance among official and private creditors.[1] The implied data collection burden on teams should be limited because: (i) holder profiles for certain instruments (e.g., loans from bilateral/multilateral creditors) are quite obvious; (ii) BOP and monetary sector projections should easily link to, and inform, these assumptions; and (iii) where allocations are less obvious, teams can make a simplifying assumption that holdings remain equal to the share of existing debt held by that investor class.

3. **Information on government asset buffers and the non-bank financial sector could also be an important data input in key country cases.**

- *Assets*: Major commodity producers that have large sovereign wealth funds as well as several advanced economies whose public sectors hold significant financial assets would be the key countries where asset buffers would be expected to have a material impact on the analysis. As a default, DSA templates could be populated automatically from centralized databases like the *Fiscal Monitor*. The IE Foundation's *Sovereign Wealth Funds* annual report could be a supplementary source for key countries.

- *Non-bank financial institutions*: In countries where the sovereign relies on the non-bank financial sector as a source of stable financing, the option to bring this sector into the analysis would require information on the aggregate assets of the sector (to calculate the country's broader financial sector). This is likely to be applicable mainly to major advanced economies or large emerging markets. Here, information would likely need to be sourced from (already prepared) national balance sheet/flow of funds accounts data.

B. The Stress Scenario and Implementation of the Holder Shock

4. **The stress scenario combines macro-fiscal and financing shocks, which tend to raise GFNs:**

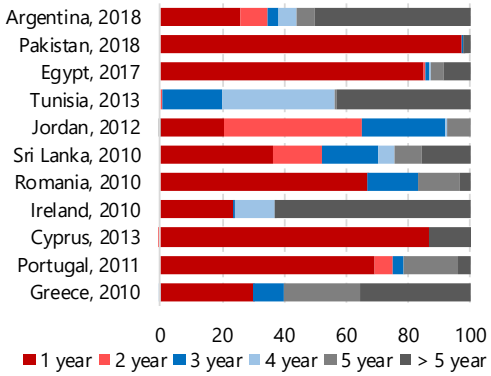
The macro-fiscal shocks are broadly similar to those of the existing MAC DSA's stress tests and include (i) a one standard deviation (computed over the last 10 years) reduction in the real GDP growth rate for two years; (ii) for countries outside currency unions and having their own legal tender, a one-year exchange rate shock equal to the largest annual depreciation observed in the last 10 years; and (iii) for currency union members and dollarized economies, a deflator shock equal half of the largest one-year change in inflation rates. These shocks are assumed to have additional knock-on effects on inflation. First, the exchange rate shock (where applicable) passes through to inflation (for a 1 percent depreciation, inflation rises by 25 basis points for EMs and 3 basis points for AEs). Second, the growth shock reduces inflation by 25 basis points for each 1 percentage point reduction in real GDP growth. Finally, the total of all these shocks affect the primary balance for two years, where the revenue/GDP ratio is fixed at the baseline level (e.g., an elasticity of 1) and

[1]The module introduces a minor new requirement to indicate whether an instrument was marketable or not.

expenditures are fixed at baseline nominal levels (e.g., an elasticity of zero).[2] Backtesting of these assumptions indicates that they are severe but not extreme (see paragraph 12 below).

Financing shocks involve shortening of maturities, which also adversely impacts GFNs through higher amortization payments. The scenario has its own financing assumptions, in which debt issuance to meet the stressed GFNs is mainly composed of shorter-term instruments. The shares allocated to each instrument follow the average maturities of bond issuances in recent crisis events (Figure AVII.1), with about half of all issuance concentrated in T-bills.

Figure AVII.1 Shares of Net New Market-Based Debt Maturing, By Year



Note: A year's net new debt issuance is measured as the change in bonds outstanding by year in percent of the change in bonds outstanding in all years.
Source: Bloomberg and Fund staff calculations.

5. **The next step is to impose assumptions on how the GFNs are financed.** As a preliminary step, prior to imposing the "holder shock" described below, the debt issuance required to meet the stressed GFNs needs to be broken down according to the 5 creditor groups: central bank, domestic commercial banks, other domestic creditors, foreign official and foreign private creditors. Allocating debt issuance among these creditors could be done in a standardized (according holdings of existing debt) or a customized manner, as decided by the team.

6. **Based on the debt issuance projections generated by these assumptions, the final step establishes the domestic financing requirements created by an external debtholder (rollover) shock.** The holder shock simulates a loss in foreign appetite for a country's sovereign debt. In the shock, which is built on top of the higher stressed GFN, foreign private rollovers drop to a 67 percent and investors are unwilling to finance any new borrowing requirement (for example, primary deficits), over a two-year period.[3] The first line of defense from this shock is any government asset buffers, but if these buffers are fully depleted, then the domestic banking sector is assumed to absorb any residual financing needs (Box AVII.2).

7. **Importantly, the risk signals derived from the test are not sensitive to how exactly either the stress scenario or the holder shock are defined.** The parameters of the test determine the size of financing that the domestic banks need to absorb when the holder shock is imposed on the stress scenario. As described in paragraph 9 below, this constitutes one of the metrics of that enters the GFN module's risk index. However, the risk signals derived from this index are calibrated based on the probabilities of missed crises and false alarms associated with each index value, *for a given test definition* (see main paper, paragraph 31). Hence, if the shock were defined to be more

[2]These calibrations are consistent with the current MAC DSA. To rule out counterintuitive results, caps were put on inflation and the fiscal balance to prevent situations where very high inflation caused improvements in GFNs relative to the baseline.

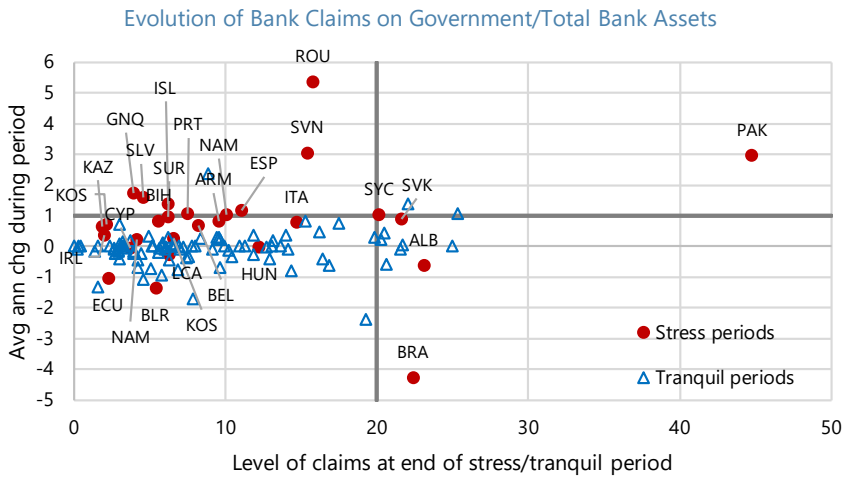
[3]These assumptions, made in consultation with area departments, are meant to capture a typical rollover shock.

severe, the thresholds that determine the mechanical signal associated with the test would be set higher than if the shock were defined to be less severe.

Box AVII.2. Behavior of the Domestic Banking Sector in Sovereign Stress Episodes

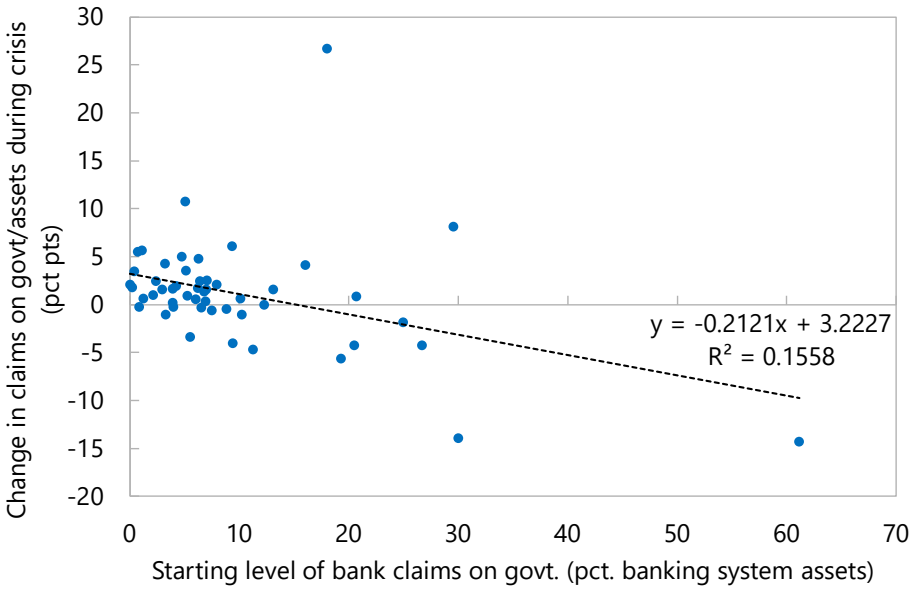
When there is an outbreak of sovereign stress, the domestic banking system tends to increase its exposure to the government, and thus serve as a residual financing source. However, the ability of banks to increase their government debt holdings will be constrained by existing exposures. Empirically, bank claims on the government seldom rise above 20 percent of banking system assets (Figure 8) and tend to rise less in stress events when starting exposures are high (Figure 9).[1]

Bank Claims on Government/Total Bank Assets—Stress Vs. Tranquil Periods



Sources: International Financial Statistics and Fund staff calculations.

Bank Claims on the Government in Stress Periods



Sources: International Financial Statistics and Fund staff calculations.

[1]These findings are also consistent with Arslanalp and Tsuda (2014), who used 15 percent of assets as a risk threshold.

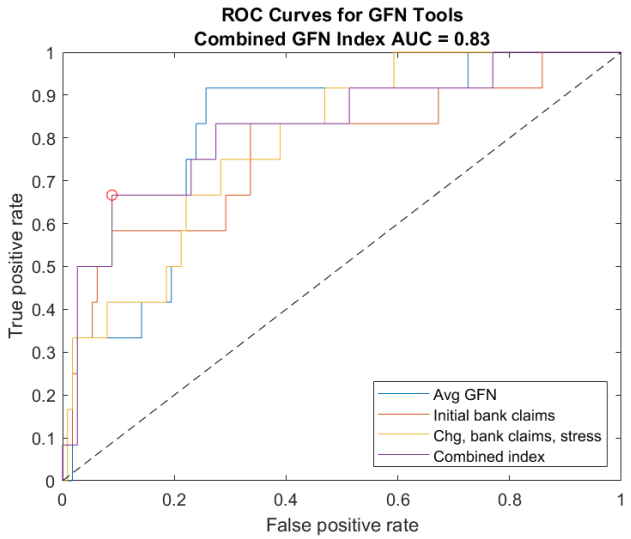
C. Derivation of the GFN Module's Risk Index and Backtesting

8. **Staff simulated the GFN module using past DSA templates.** This involved running the GFN module using the macroeconomic, fiscal, and financing assumptions that could be obtained from MAC DSA templates prepared over 2014-15 and submitted to the MAC DSA archive.[4] These were augmented with debt holder profile data from the last observed year in the corresponding MAC DSA template and information on banking system assets obtained from the IFS. Altogether, this process provided 125 observations (corresponding to about 60-70 country DSAs per year).[5] These were used to derive key potential risk metrics that might be produced by the module, as described below.

9. **Staff examined several potential risk metrics and concluded that an index composed of the following three indicators showed the best performance:**

- *GFN levels*: GFN levels have significant explanatory power in predicting crises (although not in the non-linear fashion implicitly assumed by threshold-based signals). ROC curve analysis on the average GFN projections in past DSAs submitted to the MAC DSA database suggests an in-sample AUC of 0.81.

Figure AVII.2 ROC Curves for GFN Module Risk Indicators



Source: Fund staff calculations.

- *The volume of financing needed from domestic banks*: Intuitively, the banking system would not be able to purchase outsized amounts of government debt. Staff tested the change in the ratio of bank claims on the government to banking system assets under both the baseline and stress scenarios. The baseline did not have any explanatory power. However, the change in bank claims on the government in the stress scenario showed an AUC of 0.79, also indicating potential as a stress indicator.

- *The level of initial bank claims on the government*: If bank exposures to the government are already high, then they should be less able to further increase holdings if needed. This

[4]Although GFN metrics and signals are available for more recent period, the stress outcomes associated with these signals cannot be observed for the full medium term (5-year) prediction period. Hence, predictive performance is analyzed based on GFN signals between 2014 (the earliest period available) and 2015.

[5]Where more than one DSA was undertaken in a year, the results were averaged.

conjecture was confirmed by ROC curve analysis of the predictive power of the most recent level of bank claims on the government (in percent of assets).

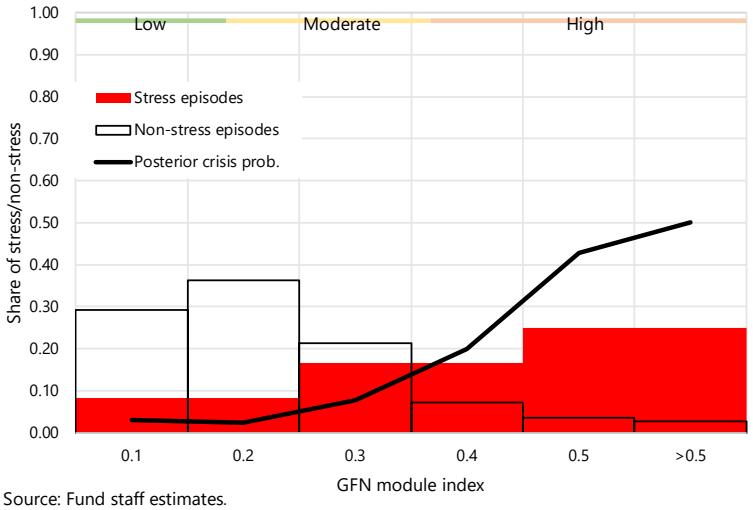
10. **To aggregate the information from these three indicators, staff combined them into an aggregate GFN Financeability Index (GFI), weighted by their explanatory power (i.e., their AUC).** This overall index has an AUC of 0.83, implying that it is an improvement over each of these indicators in isolation (Figure AVII.2).

11. **Back-testing based on archived DSAs over 2014-15 reveals a substantially improved performance for these metrics relative to the existing GFN/GDP thresholds.** The composite index has a high AUC of 0.83 and a low TME of 42 percent. By comparison, the existing MAC DSA GFN thresholds that are associated with average missed crisis and false alarm rates of 68 and 15 percent, respectively. Moreover, an additional illustration of the GFI's explanatory power is given by the limited overlap between the stress and non-stress distributions displayed in Figure AVII.3. As with the fanchart and logit model, the GFI is conducive to a three-risk zone (high, moderate, low) classification.

12. **The backtesting exercise was also used to assess the plausibility of the stress scenario, by examining the location of the implied debt path within each debt fanchart.** The implied path lies above the median but below the 95th percentile, five years out, in a large majority of cases (89 percent). This suggests that the scenario is severe but not an extreme tail risk.

13. **As with the other DSA tools, the relationship between the risk ratings obtained from the model and the likelihood of stress can be examined by estimating posterior stress probabilities**. While the level of the GFI is not a direct estimate of the probability of a "stress" event, estimates of the posterior probability of stress at each level of the GFI can be derived based on the share of countries within a given GFI range that went on to experience stress in sample. Figure A.VII.3 depicts such estimates for each of 6 "bins" for the GFI (5 bins for GFI values below 0.5 and one bin for values above 0.5). While the limited number of "stress" observations mean that these probabilities can only be estimated imprecisely (particularly at higher values of the GFI where there are fewer observations), there is a clear pickup in the posterior stress probabilities around the "low/moderate" and "moderate/high" thresholds. The figure suggests a posterior probability of stress conditional on a GFI "high risk" signal of at least 24 percent. The average posterior probability of stress for a country conditional on the GFI falling in each of the three risk zones is 42.1 percent for the "high risk" zone, compared to 4.1 percent and 3.5 percent for "moderate" and "low" risk zones, respectively.

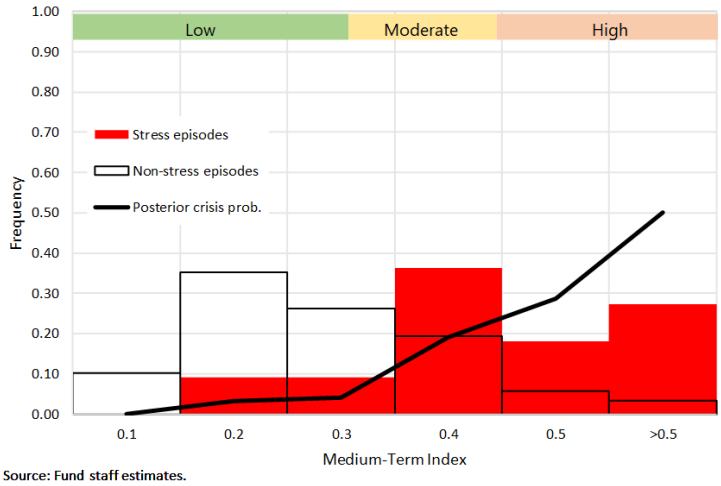
Figure AVII.3. GFN Financeability Index Distributions in Stress and Non-Stress Cases



Source: Fund staff estimates.

14. **Similarly, posterior probabilities of stress conditional on a risk signal can be derived for the medium-term risk index (MTI) that is derived averaging the fanchart and GFN indices of risk.** Figure A.VII.4 depicts such estimates for each of 5 "bins" for the MTI (5 bins for MTI values below 0.5 and one bin for values above 0.5). The figure suggests a posterior probability of stress conditional on a GFI "high risk" signal of at least 20 percent. The average posterior probability of stress for a country conditional on the MTI falling in each of the three risk zones is 43 percent for the "high risk", 9 percent for the "moderate risk", and 4 percent for the "low" risk zones, respectively.

Figure AVII.4. Medium-Term Risk Index Distributions in Stress and Non-Stress Cases



Source: Fund staff estimates.

D. Customization of the GFN Tool in Special Cases

15. **The GFN tool will incorporate some standardized customizations and guidance will be provided to better account for the following factors that may contribute to, or detract from, the availability of liquidity to the government:**

(i) *Use of government assets to offset funding pressures*: Teams will be able to customize the size of any available liquid asset buffers that can be used to meet financing needs generated by the holder shock before allocating new claims to the banks.[6]

(ii) *The role of the domestic non-bank financial sector as a residual creditor*: For some countries, the domestic non-bank sector is larger than the domestic banking system and a major source of stable government funding.[7] In these cases, funding needs arising in the stress scenario can be absorbed by the broader financial sector (rather than just domestic commercial banks), resulting in a smaller overall demand to increase exposures to the sovereign (relative to the combined assets of this broader financial sector).[8] This adjustment is particularly relevant for reserve currency issuers, whose Treasury securities are often held disproportionately by non-residents but whose large domestic nonbank sectors should be available to meet any financing gap generated by reduced non-resident participation.

(iii) *Non-bank financial intermediaries as a source of government funding risk*. In contrast to (ii), in these cases, teams would be expected to identify the portion of this sector's financing that is subject to a sudden stop, which would then be treated in a parallel manner as financing from private foreign creditors under the holder shock.

(iv) *The timing of the onset of stress*: The start of the rollover shock, macro-fiscal, and maturity shortening shocks would be adjustable to allow teams to align the possible onset of market stress with a specific event (for example, a political event such as elections) and to accommodate cases where a country has no significant external private debt maturing during the first two projection years.

(v) *More granular information available to teams*: There may be certain circumstances when staff has more detailed information on the banking system's capacity to absorb additional government debt in a stress situation (e.g. analysis performed in the context of an FSAP). Other information, including a country's banking regulations, arrangements between public sector entities, capacity to conduct liability management operations and/or capital flow measures may also impact the analysis of government financing risks. Guidance would spell out how to use this information, including whether it could be integrated with the standard approach.

[6]Liquid asset buffers are likely to produce important effects in the GFN module in major EM commodity producers with large sovereign wealth funds (e.g., Kuwait, Qatar, Saudi Arabia, United Arab Emirates), as well as in several advanced economies with sizable financial assets (e.g. Canada, France, Germany, Japan, Singapore, United Kingdom and United States).

[7]For instance, in cases where mutual funds or other investors might be the primary financiers of government debt.

[8]It would also be appropriate to incorporate the non-bank financial sector in the initial government exposures/asset ratio included in the GFI. When the non-bank sector is large, its inclusion would likely cause this ratio to fall, capturing the benefit of a deep financial system.

Annex VIII. Resource Requirements for the New MAC DSA

1. **While transitioning to the new framework will involve a time and resource investment, it should not be costlier to maintain than the current one once it is up-and-running.**

- *Data requirements*: Most data requirements, including those that seem new, carry over from the existing MAC DSA framework (see table A.VIII.1). Fresh data requirements arise in four areas—debt holders, 10-year projections, inputs for stress tests and long-term modules, and debt disclosures. Some of these data needs, such as those for customizing triggered stress tests and long-run analysis apply only in special cases. The 10-year projection horizon, which is a new requirement, does not imply a need for a full financial program and instead can be satisfied through a careful extrapolation of key variables after the normal 5-year horizon. However, certain debt disclosures may constitute a new data requirement for teams with the support of country authorities, although this information provides critical information on debt risks.

- *Centralization and automation*: Many variables required by the new framework can be sourced from existing cross-country databases. The new template will be pre-populated with default parameter settings and centrally warehoused data. Staff has already made significant use of these sources and default settings to design the tools. In testing the new tools, the review team has already run them on nearly every MAC, proving the feasibility of implementing the new framework.

- *Transitioning to the new template*: While the template will be designed to be as user friendly as possible, country teams will require some assistance in transferring their databases and projections to the new template and potentially customizing it to reflect country-specific factors. For this purpose, SPR would provide intensive support to area departments through an implementation team, drawing on the experience of the smooth LIC DSF rollout. After this transition, updating and running the new framework is not expected to be more demanding than in the present framework, given the centralization and automation of data sourcing and the fact that the new tools should enable shorter and more focused writeups.

2. **After implementation, staff will carefully facilitate transitional arrangements for PRGT-graduating members and other frontier countries that use the MAC DSA**. Transitioning between frameworks does entail an effort from both the country team and the authorities. While there are similarities between the MAC and LIC debt sustainability frameworks, the requirements are not fully overlapping and may require country authorities to collect new data. It will also involve a training effort to become fully abreast of the new framework and its interpretation. However, early identification of potential graduates/new users of this template should help provide an ample transition period to provide training/technical support where needed and help deliver a smooth changeover. Further considerations will be handled in the Guidance Note.

Table A.VIII.1 Data Requirements for the New MAC DSA Framework

Variable	Module: Realism tools, debt profile	Module: Near-term (logit)	Module: Medium-term: Debt fan-chart	Module: Medium-term: GFN module	Module: Medium-term: Tail risks	Module: Long-term: Aging	Module: Long-term: Nat. res.	Module: Long-term: Large amort-ization	Always needed or subj. to trigger	Existing or new data require-ment	Scope for central-ization	Source
Fiscal data/projections (up to t+5)												
Primary revenues, expenditures, balance	●	●	●	●	●				Yes	Existing	No*	Teams/WEO
Interest bill and receipts	●		●	●	●				Yes	Existing	No	Teams
Debt									Yes	Existing	No*	Teams/WEO
By residency (incl. external debt)		●		●					Yes	Existing	No	Teams
By currency			●		●				Yes	Existing	No	Teams/WEO
By maturity	●			●					Yes	Existing	No	Teams
By holder	●			●					Yes	New	Yes	Arslanalp-Tsuda**
By legal basis	●								Yes	New	No	Authorities
Amortization of existing debt				●					Yes	Existing	No	Teams
Assumptions on new debt issuance	●			●					Yes	Existing	No	Teams
Gross financing need (calculated)		●		●	●				Yes	Existing	No*	DSA calculation
Historical stock-flow adj. (as validated)	●		●						Yes	Existing	Yes	SPR**
Government liquid assets				●				●	Yes	New	Yes***	Fiscal Monitor
Cyclically adjusted primary balance	●								Yes	Existing	No	Teams
Forecast track record (PB & debt drivers)	●								Yes	Existing†	Yes	SPR
Average maturity of public debt	●								Yes	New	No	Authorities/teams
Debt coverage disclosures	●								Yes	New	No	Authorities/teams
Intra-governmental holdings	●								Yes	New	No	Authorities/teams
Major macro variables/proj. (up to t+5)												
Real and nominal GDP and deflator	●		●	●	●	●			Yes	Existing	No	Teams/WEO
Current account balance		●							Yes	Existing	Yes**	Teams/WEO
Nominal bilateral ER		●	●	●	●				Yes	Existing	Some*	Teams/WEO
Real effective ER	●	●	●						Yes	Existing	Some*	Teams/IMD
International reserves		●							Yes	New	Yes	IFS
Potential GDP and output gap	●								Yes	Existing	No	Teams/WEO
Forecast track record for key variables	●									New	Yes	SPR
Financial sector & structural indicators												
Bond spreads	●								Yes	New	Yes	Teams/Bberg
VIX		●							Yes	New	Yes	CBOE
U.S. long-term interest rate	●	●							Yes	New	Yes	Haver
Governance composite indicator		●							Yes	New	Yes	WEF
Stress history		●							Yes	New	Yes	SPR**
Share of CU MACs in stress		●							Yes	New	Yes	SPR
Financial sector credit and gap		●			●				Yes	Existing	Yes	BIS/IFS
Financial sector deposits					●				Trigger	Existing	Yes	IFS
Banking system assets				●					Yes	New	Yes††	IFS/IMD/Haver
Estimated exchange rate overvaluation	●				●				Trigger	Existing	No	EBA/EBA lite
Frequency/cost of natural disasters					●				Trigger	New	Yes	EMDAT
Adverse commodity path					●				Trigger	New	Yes	RES
Financial soundness indicators					●				Trigger	New	Yes	MCM
Specialized long-term analyses												
Pension program information												
Demographic and labor indicators						●			Trigger	New	Yes	UN Pop/ILO
Current beneficiaries						●			Trigger	New	No	Authorities
Current revenues/GDP						●			Trigger	New	No	Authorities
Current benefit payments/GDP						●			Trigger	New	No	Authorities
System reserves						●			Trigger	New	No	Authorities
Natural resource sector data/projections												
Proven reserves							●		Trigger	New	Yes	BP
Investment and production plans							●		Trigger	New	No	Various
Long-term data/projections (t+6 to t+10)												
Amorization of existing and new debt								●	Trigger	New	No	Authorities
Real and nominal GDP and deflator								●	Trigger	New	No	Team
Primary revenues, expenditures, balance								●	Trigger	New	No	Team
Interest bill and receipts								●	Trigger	New	No	Team

* For near-term assessment/logit model data can be updated and run centrally for periodic updates. ** Based on existing estimates, which some teams may be periodically requested to validate/update. *** Where data are unavailable in the Fiscal Monitor a default option of zero would exist, though teams may wish to adjust. †SPR plans to expand the dataset of forecast errors for several additional variables to also include exchange rate, SFAs, and r-g. ††A limited number of teams may need to provide a source for bank assets, when countries do not report to STA and there is no data coverage in Haver.

References

Arslanalp, S., and Tsuda, T., 2014, "Tracking Global Demand for Emerging Market Sovereign Debt," IMF Working Papers No. 14/39. (Washington: International Monetary Fund).

Baldacci, E., Petrova, I.K., Belhocine, N., Dobrescu, G. and Mazraani, S., 2011, "Assessing fiscal stress," IMF Working Papers, pp.1-41. (Washington: International Monetary Fund).

Bassanetti, A., Cottarelli, C., and Presbitero A., 2019, "Lost and Found: Market Access and Public Debt Dynamics," *Oxford Economic Papers*, Volume 71, Issue 2, April 2019, Pages 445–71.

Carvalho, C.M., Polson, N.G. and Scott, J.G., 2010, "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), pp.465-480.

Catão, L.A. and Milesi-Ferretti, G.M., 2014, "External liabilities and crises," *Journal of International Economics*, 94(1), pp.18-32.

Cecchetti, S.G., Mohanty, M.S. and Zampolli, F., 2011, "The real effects of debt"

Cerovic, S., Gerling, K., Hodge, A., Medas, P., 2018, "Predicting fiscal crises," IMF Working Paper 18/181. (Washington: International Monetary Fund).

Das, U.S., Papaioannou, M.G. and Trebesch, C., 2012, "Sovereign debt restructurings 1950-2010: Literature survey, data, and stylized facts (p. 12)," (Washington: International Monetary Fund).

Drehmann, M. and Juselius, M., 2014, "Evaluating early warning indicators of banking crises: Satisfying policy requirements," *International Journal of Forecasting*, 30(3), pp.759-780.

Erce, A. and Mallucci, E., 2018, "Selective sovereign defaults," FRB International Finance Discussion Paper, (1239).

Faria, A. and Mauro, P., 2009, "Institutions and the external capital structure of countries," *Journal of International Money and Finance,* 28(3), pp.367-391.

Fournier, J.M. and Bétin, M., 2018, "Sovereign defaults: Evidence on the importance of government effectiveness," OECD Economic Department Working Papers, (1494).

Gardner, E.H. and Schimmelpfennig, A., 2008, "Lebanon-Weathering the Perfect Storms," IMF Working Papers 08/17. (Washington: International Monetary Fund).

Guscina, A. and Malik, S. and Papaioannou, M. G., 2017, "Assessing Loss of Market Access: Conceptual and Operational Issues," IMF Working Paper No. 17/246. (Washington: International Monetary Fund).

International Monetary Fund, 2013, "Staff Guidance Note for Public Debt Sustainability Analysis in Market Access Countries," (Washington: International Monetary Fund).

International Monetary Fund, 2014, "Republic of Slovenia: 2013 Article IV Consultation," IMF Country Report No. 14/11. (Washington: International Monetary Fund).

International Monetary Fund, 2016, "Argentina: Economic Developments," Country Report No. 16/69 (Washington: International Monetary Fund).

International Monetary Fund, 2016, "Republic of Serbia: Fourth And Fifth Reviews Under The Stand-By Arrangement And Rephasing Of The Arrangement, Seventh Review Under the Stand-By Arrangement and Modification of Performance Criteria—Press Release; Staff Report; And Statement by The Executive Director for The Republic of Serbia," IMF Country Report No. 16/287. (Washington: International Monetary Fund).

International Monetary Fund, 2017, " Collaboration between Regional Financing Arrangements and the IMF," (Washington: International Monetary Fund).

International Monetary Fund, 2017, "Republic of Slovenia: 2017 Article IV Consultation-Press Release; Staff Report; And Statement by The Executive Director for The Republic of Slovenia," IMF Country Report No. 17/125. (Washington: International Monetary Fund).

International Monetary Fund, 2017, "Republic of Serbia: 2017 Article IV Consultation, Seventh Review Under the Stand-By Arrangement and Modification of Performance Criteria—Press Release; Staff Report; And Statement by The Executive Director for The Republic of Serbia," IMF Country Report No. 17/263. (Washington: International Monetary Fund).

International Monetary Fund, 2020, "How to Assess Country Risk: Vulnerability Exercise Approach Using Machine Learning," (Washington: International Monetary Fund).

Kangur, M.A., Kirabaeva, K., Natal, J.M. and Voigts, S., 2019, "How Informative Are Real Time Output Gap Estimates in Europe?" IMF Working Papers No 19/200. (Washington: International Monetary Fund).

Kraay, A. and Nehru, V., 2006, "When is external debt sustainable?", The World Bank Economic Review, 20(3), pp.341-365. (Washington: World Bank).

Kumar, M. and Woo, J., 2010, "Public debt and growth," IMF working papers, pp.1-47. (Washington: International Monetary Fund).

Laubach, T., 2009, "New Evidence on the Interest Rate Effects of Budget Deficits and Debt," *Journal of the European Economic Association*, MIT Press, vol. 7(4), pages 858-85, June.

Makalic E. & Schmidt, D. F., 2016, "High-Dimensional Bayesian Regularised Regression with the BayesReg Package," arXiv:1611.06649 [stat.CO].

Manasse, P. and Roubini, N., 2009. "Rules of thumb for sovereign debt crises," *Journal of International Economics*, 78(2), pp.192-205.

Medas, P., Poghosyan, T., Xu, Y., Farah-Yacoub, J. and Gerling, K., 2018, "Fiscal Crises," *Journal of International Money and Finance*, 88, pp.191-207.

Schimmelpfennig, M.A., Roubini, N. and Manasse, P., 2003, "Predicting sovereign debt crises," (Washington: International Monetary Fund).

Sumner, S.P. and Berti, K., 2017, "A complementary tool to monitor fiscal stress in European economies," Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Reinhart, C.M., Rogoff, K.S. and Savastano, M.A., 2003, "Debt intolerance," (No. w9908), National Bureau of Economic Research.