EXECUTIVE BOARD MEETING

SM/20/169
Supplement 1

December 1, 2020

To: Members of the Executive Board

From: The Secretary

Subject: **Review of the Debt Sustainability Framework for Market Access Countries—Annexes**

Board Action:	Executive Directors' **consideration** (Formal)
Tentative Board Date:	**Thursday, January 14, 2021**
Publication:	Proposed, with main paper
Questions:	Mr. Zettelmeyer, SPR (ext. 39482) Mr. Saenz, SPR (ext. 35408) Ms. Marchettini, SPR (ext. 38397)

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, Maximilien Kaffo, 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

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.

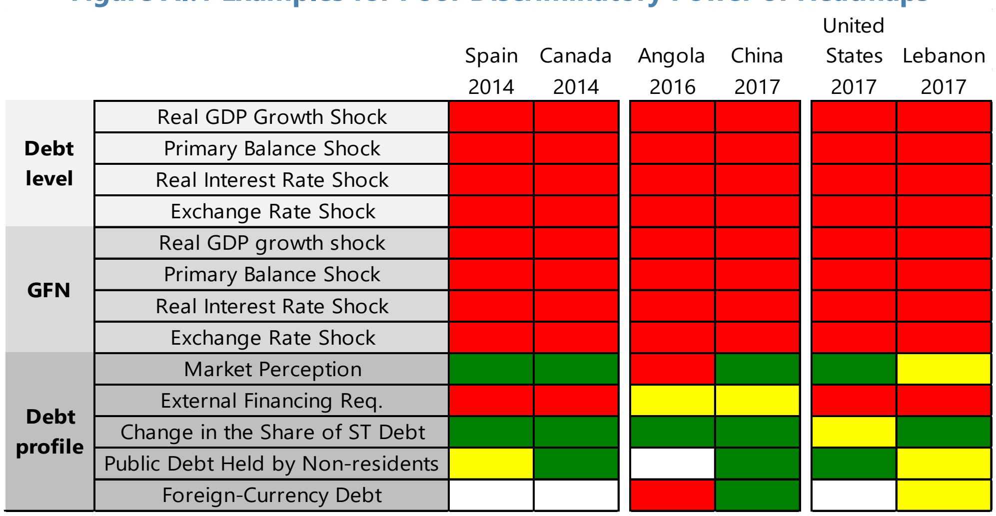
Discriminatory (Predictive) Capacity

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



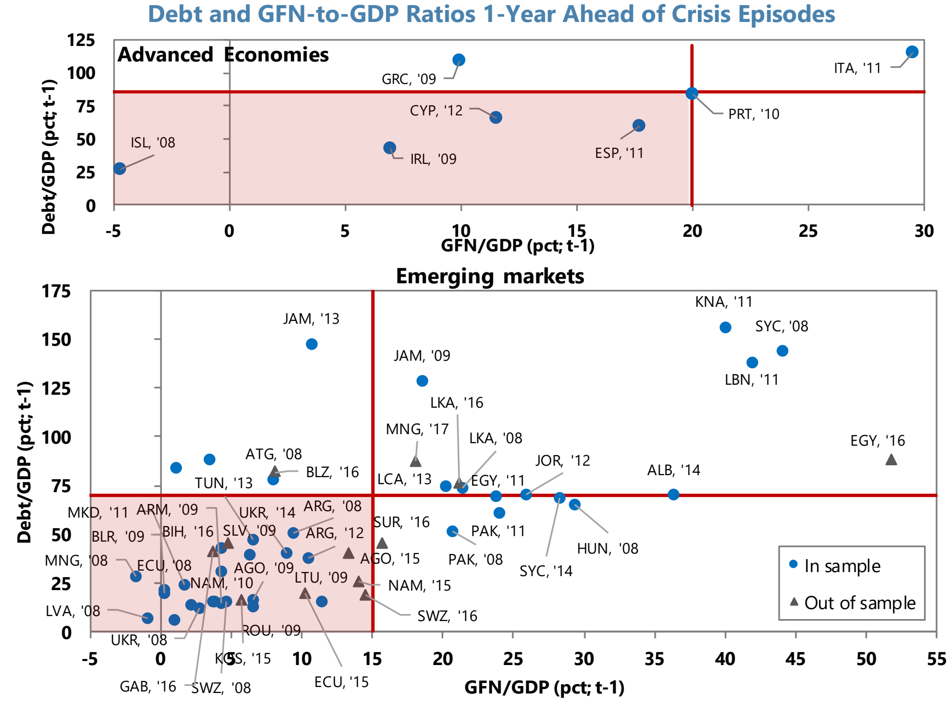
Source: MAC DSA Database.

[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	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%*

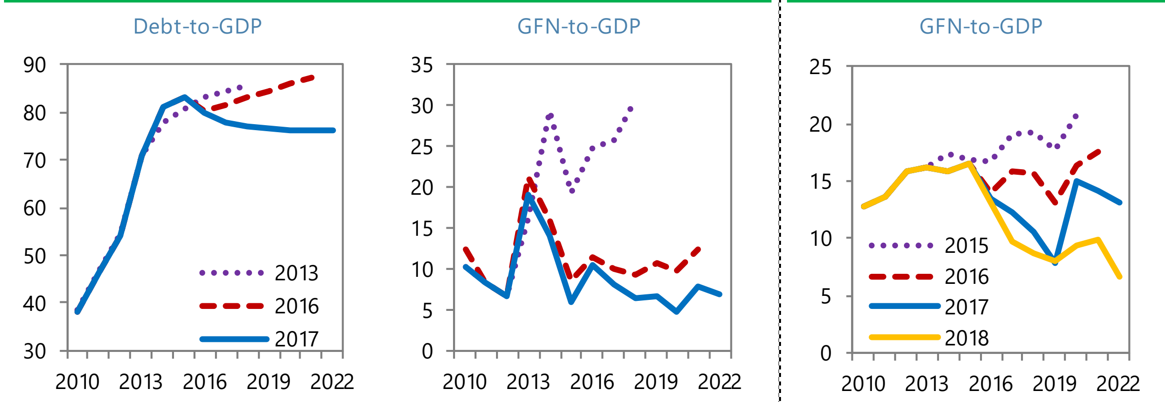
Sources: MAC DSA database and Fund staff calculations.

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)

Slovenia	Serbia
2013 assessment: ... debt has been growing rapidly ... Under the baseline scenario, Slovenia's public debt continues to grow ...	2015 assessment: .. Nevertheless, large gross financing needs... constitute significant risks.
2017 assessment: Fiscal consolidation has improved Slovenia's medium-term debt dynamics, but vulnerabilities remain... somewhat more benign public debt dynamics relative to the last Article IV projections.	2017 assessment: The sizable fiscal adjustment achieved ... has reversed the persistent increase in public debt ...



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).**

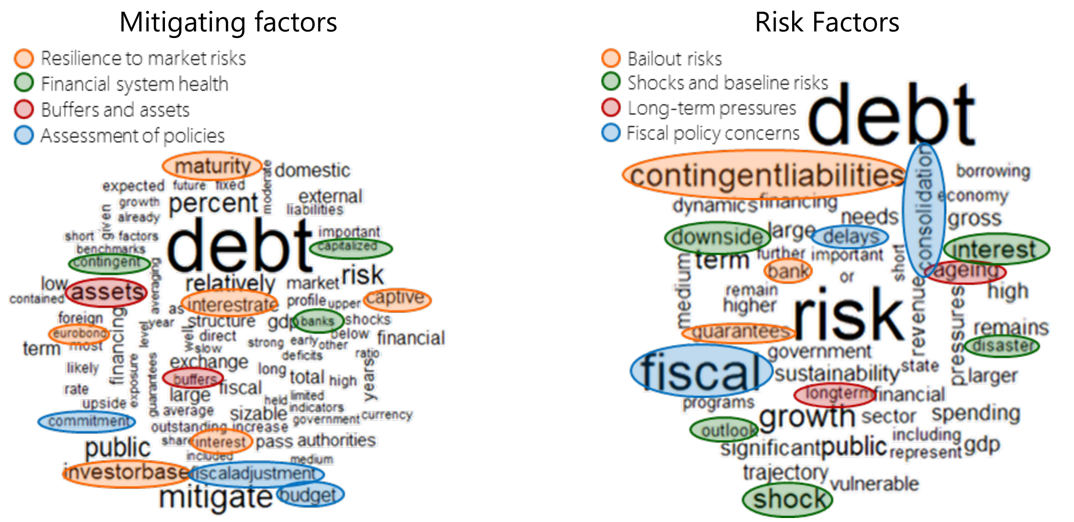
- *Resilience to market risks*. Teams often cited the existence of assets or buffers as a mitigating factor. The holder profile of debt was also referenced, at times as a mitigating factor (e.g., a stable or captive investor base) and at other times as a risk factor (e.g., vulnerability so sudden stops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



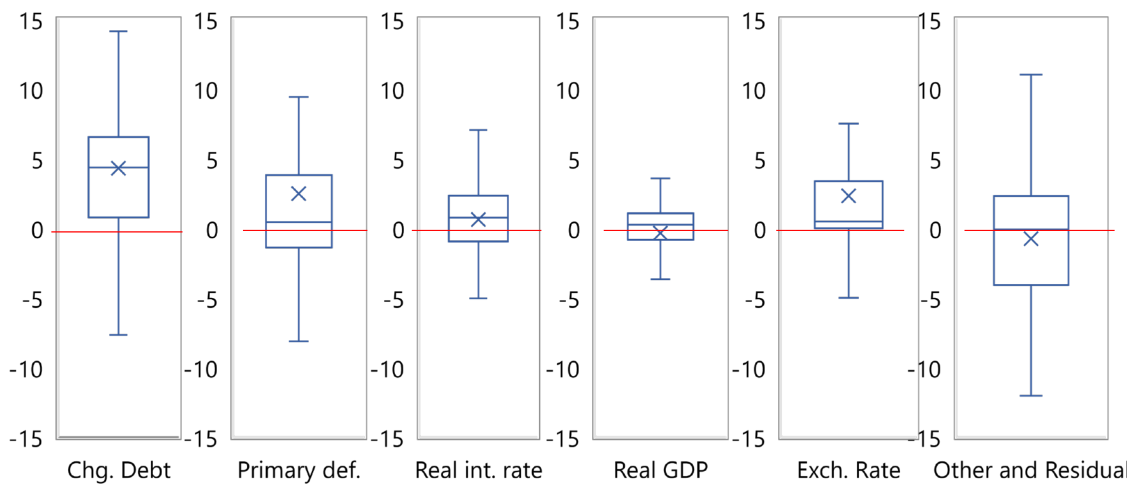
Source: Fund staff analysis of DSA writeups.

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.

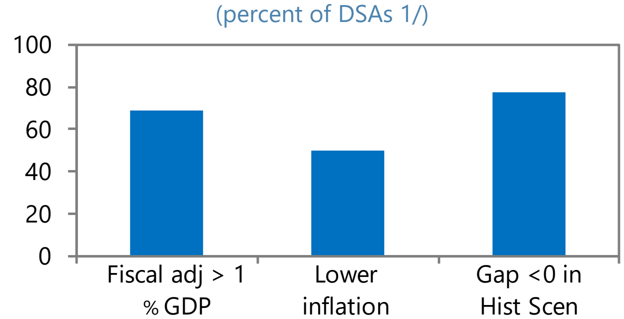
Table AI.4 Tests for Projection Biases

	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.*

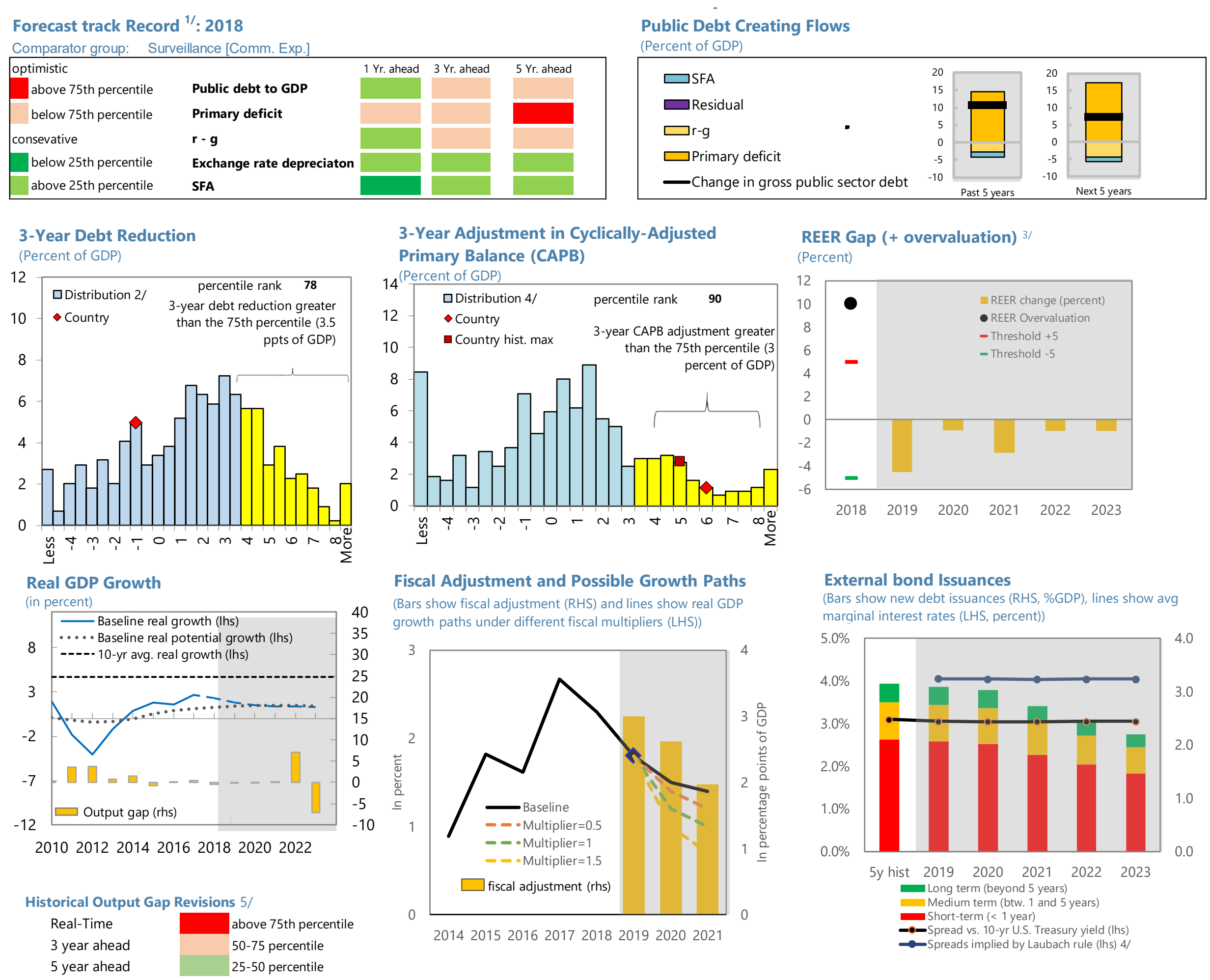
[1]This tool is based on Kangur et. al. (2019) and staff analysis showing the existence of real-time output gap biases for a majority of market access countries.

- 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 observations 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.

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 <u>large IMF programs</u> (data from the IMF Finance Department and the MONA database) and <u>exceptional financing from other IFIs and donors</u>. 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 <u>default</u>. 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 <u>restructuring</u> episodes. Conditions for stress event:
   - List from Das et al. (2012), complemented with Guscina et al. (2017).

iv. Episodes associated with <u>hyperinflation</u>. 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. <u>Episodes flagged by market-related indicators</u>.
   - 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:

(blue)	Stress Country-Year identified by mechanical criteria
(green)	Country-Year separating two stress country-years identified by mechanical criteria
(yellow)	Stress Country-Year identified by judgment
(red)	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:

(blue)	Stress Country-Year identified by mechanical criteria
(green)	Country-Year separating two stress country-years identified by mechanical criteria
(orange)	Stress Country-Year identified by judgment
(red)	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.

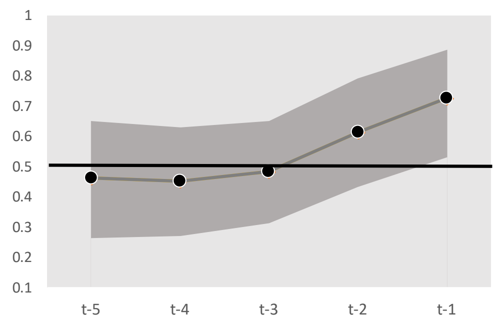
While the option of using fixed effects was dismissed for the reasons above, the analysis provided an additional robustness test for the estimates. The significance and sign of the coefficients remains broadly stable when fixed effects are estimated, except for the coefficient of “stress history”, which 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.

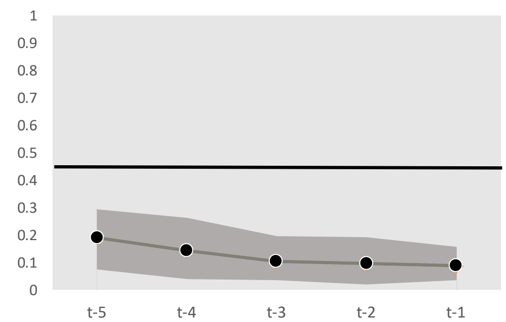
Figure AV.1 Predictive Performance (in terms of AUC) of Individual EWI at *t*+1 to *t*+5

ADVANCED ECONOMIES

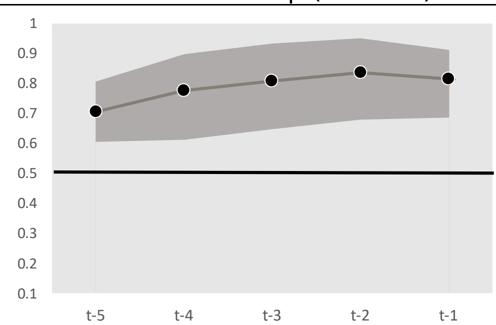
Government Debt Gap (HP Filter)



Current Account Balance



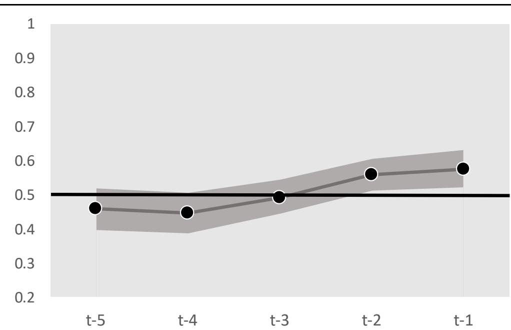
Credit to GDP Gap (HP Filter)



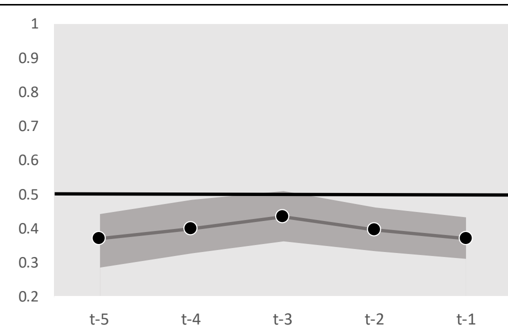
The Figure shows that, for AE, debt increases (in the picture measured as gap from a trend) are significant predictors of sovereign stress 1-2 years before the stress materializes. Cyclical indicators of external of financial imbalances are significant predictors over the full 5-year projection horizon.

EMERGING MARKETS

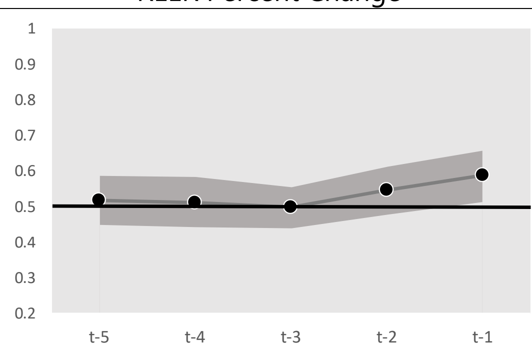
Government Debt Growth



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

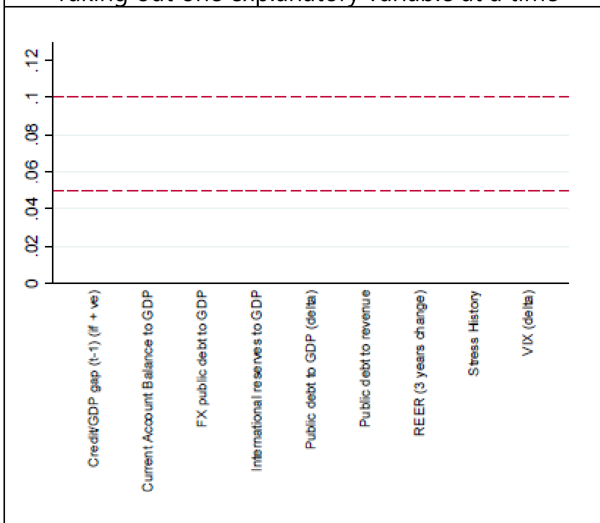
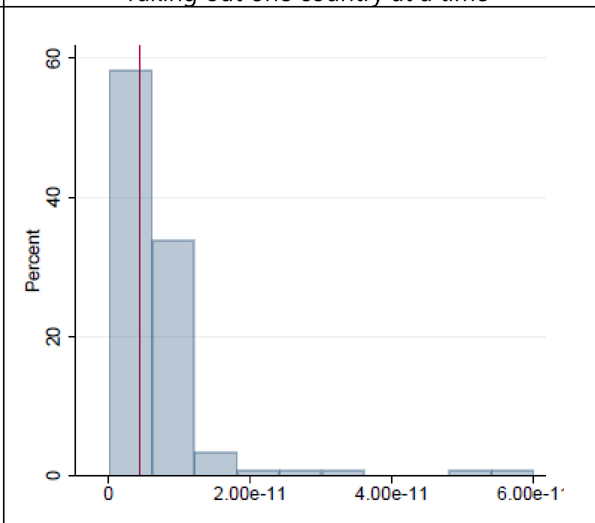
To capture country heterogeneity in a granular continuous way, Staff examined several slow-moving variables which, reflecting structural characteristics, could inform on countries' inner "debt carrying capacity". 1/

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).**

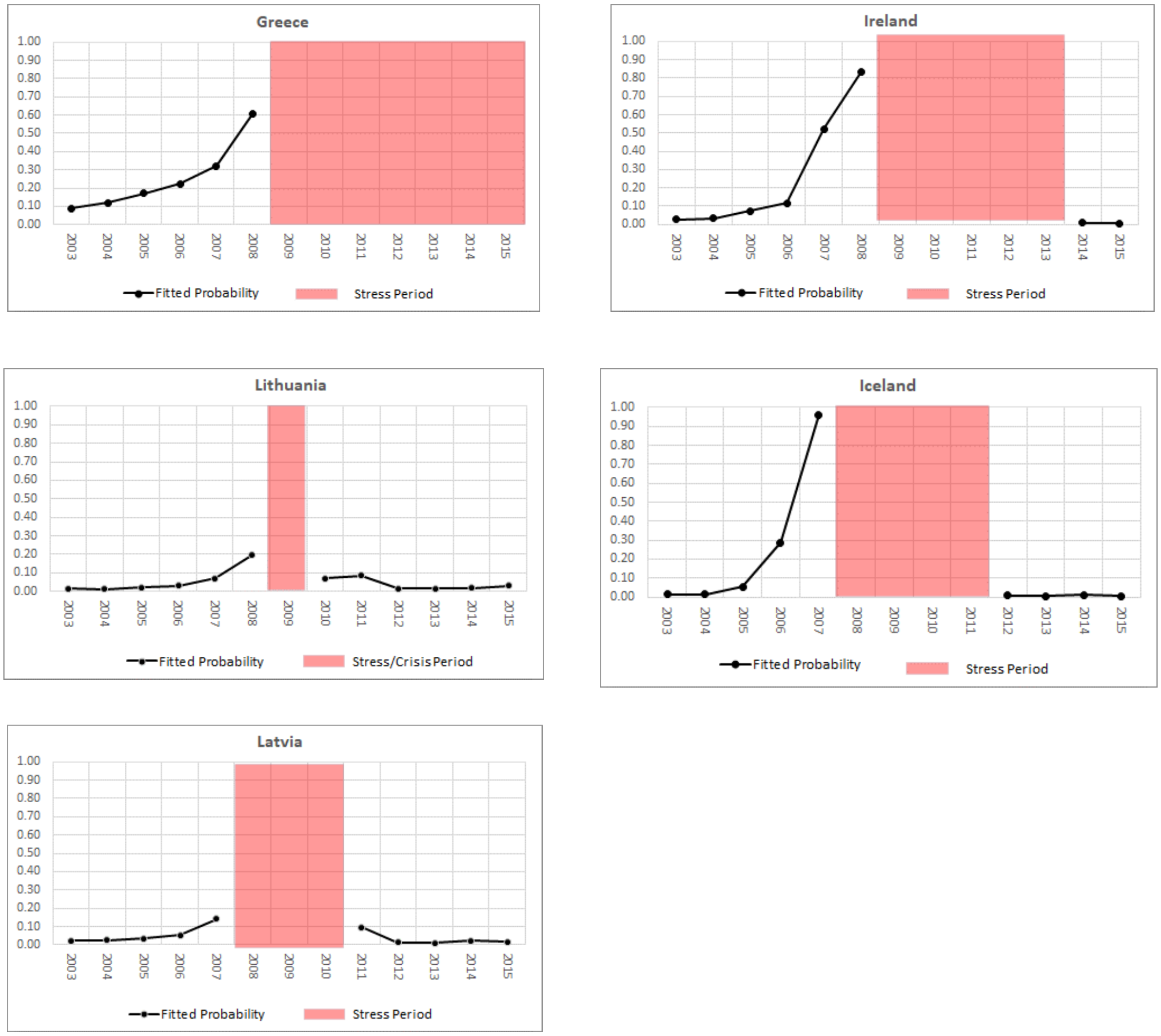
Predictive performance in countries that experienced stress due to regional spillovers is higher for the logit specification which includes the share of CU MACs in stress (Figure AV.3).

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



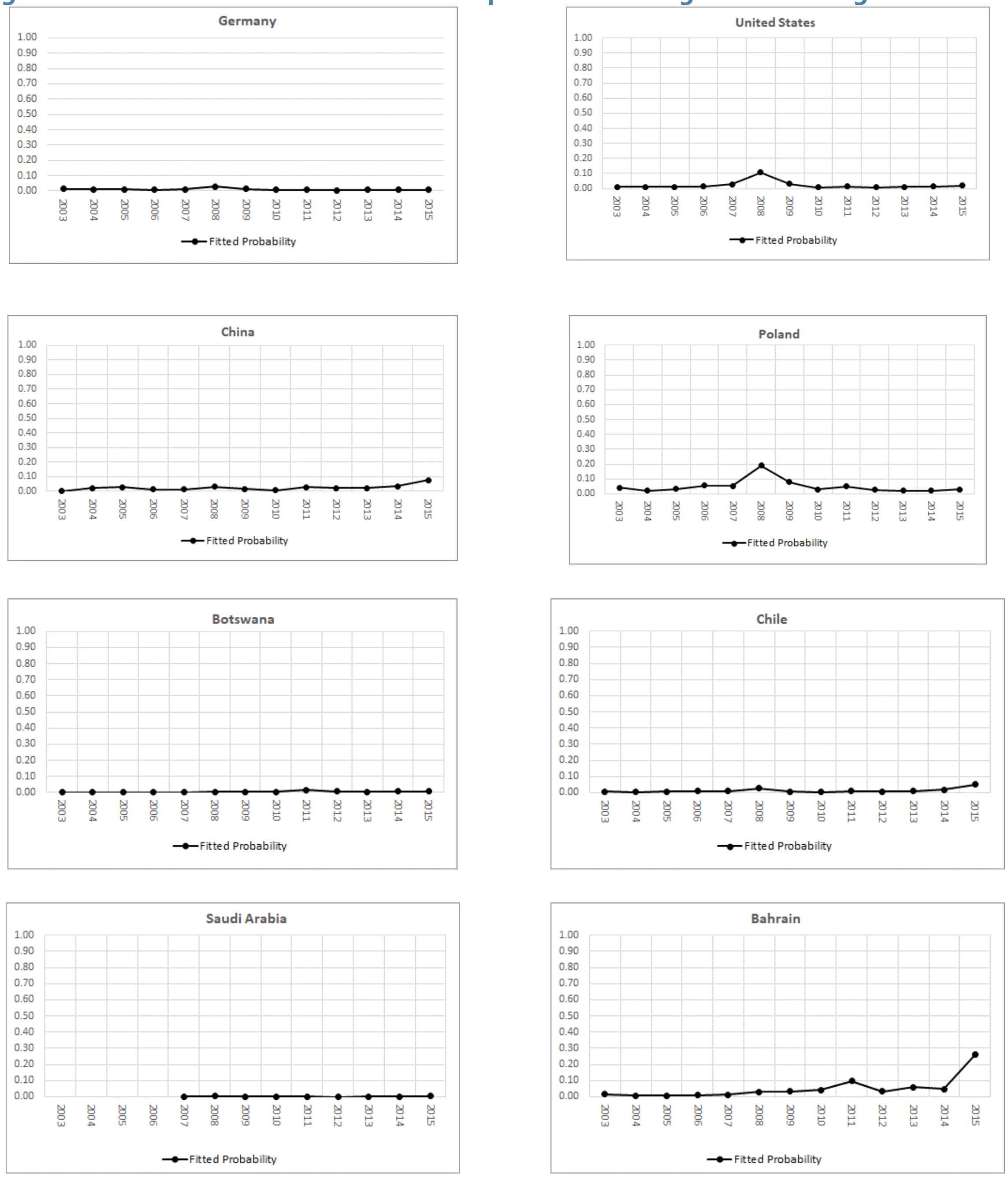
Source: Fund staff estimates.

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

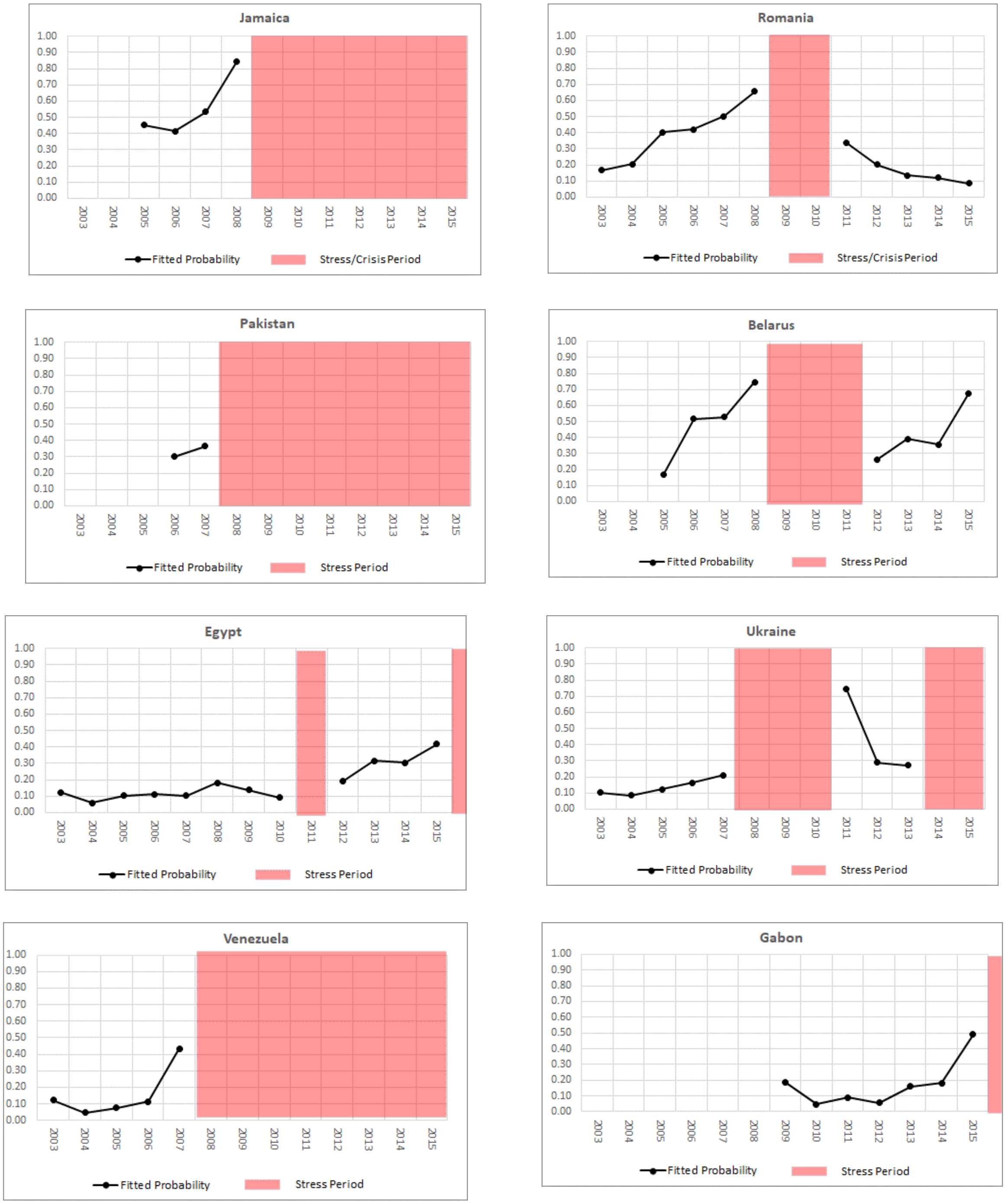
Source: Fund staff estimates.

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: Minimum Total Misspecifcation Error (TME)	Loss Function Minimization: *of which: Missed Crisis Rate*	Loss Function Minimization: *of which: False Alarm Rate*	# of Stress Episodes
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

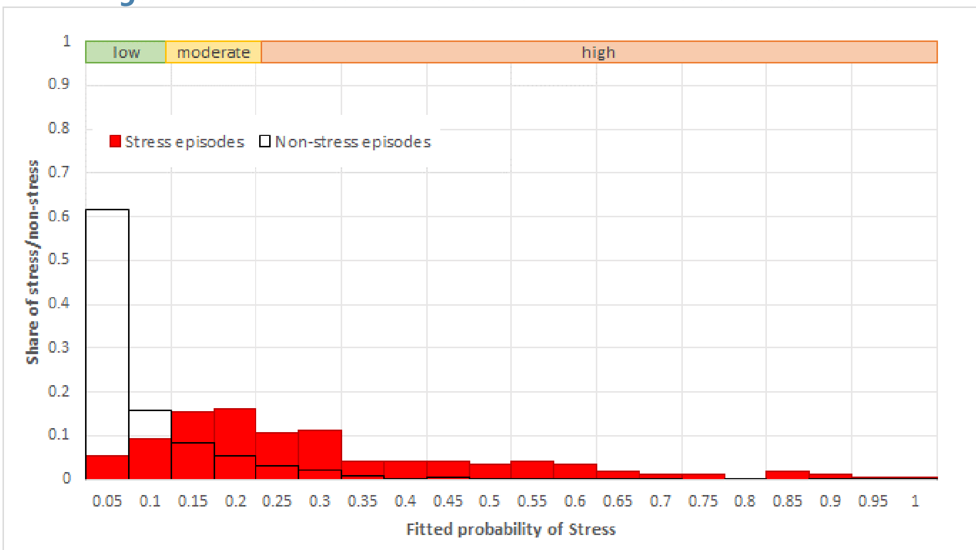


Table AV.4. Signal Derived with the Proposed Decision Rule in Selected Stress Episodes

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.

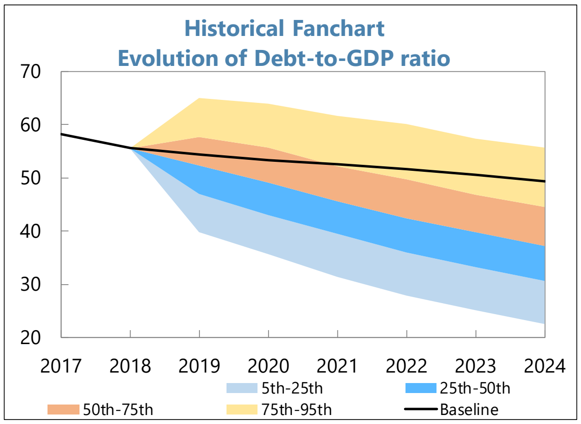
[2]Uncertainty about the *initial* level of the debt-to-GDP ratio is also incorporated by appealing to the historical WEO debt data revisions for the country. This adjustment for base effect risk was proposed in place of the initial proposal of using stock-flow-adjustment (SFA) shocks, which was dropped 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.

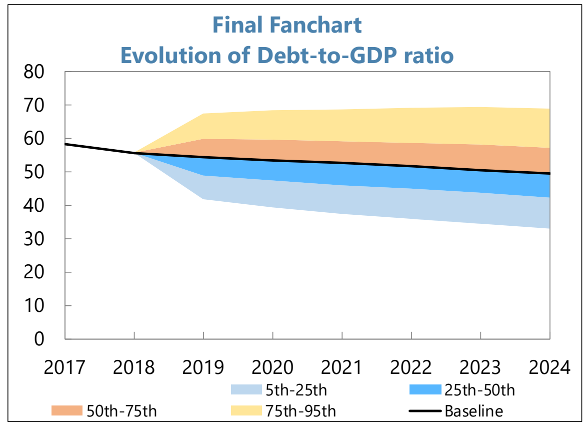
In this case, the team's baseline would be assessed as sufficiently realistic and representing an adequate balance of risks. The forward-looking information included in the baseline fully determines the (upward/horizontal/downward) "direction" of the fanchart; while its width and skew is determined by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

Figure AVI.2. Application of the New Fanchart Methodology

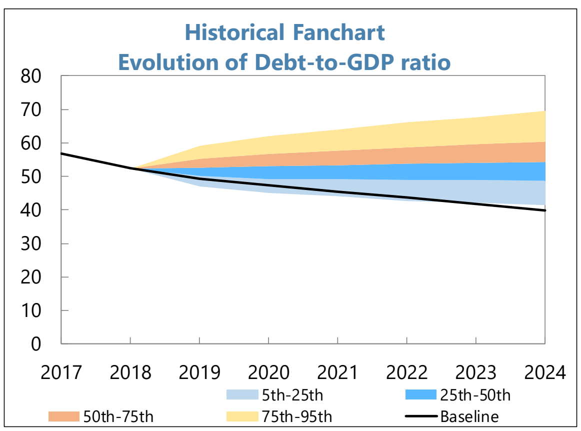
County 1 (2019): Baseline above 20th percentile of historical fanchart



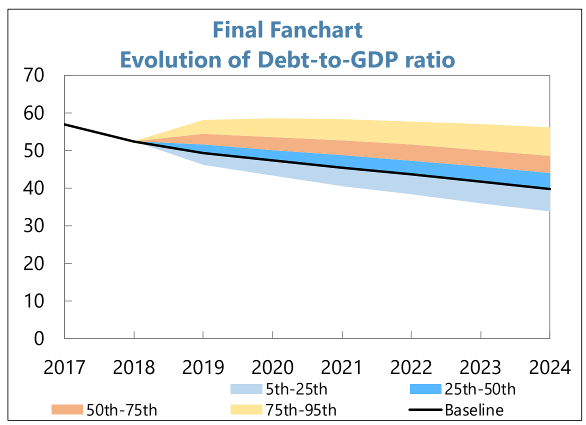
Country 1 (2019): Final fanchart obtained by applying de-meaned shocks to baseline (standard fanchart)



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:

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

where $debt_{x,t+j}^{p}$ 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_x^{max}=max($d_{x,0}^{p}$, $d_{x,1}^{p}$,…., $d_{x,5}^{p}$) 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_x^{max}.
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:*

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

where $debt_{c,t+j}^{a}$ 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 $d_{x,j}^{p}$ are computed using the team's debt projections, the distribution of departures from the historical trend is derived using the distances $d_{c,j}^{a}$ 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

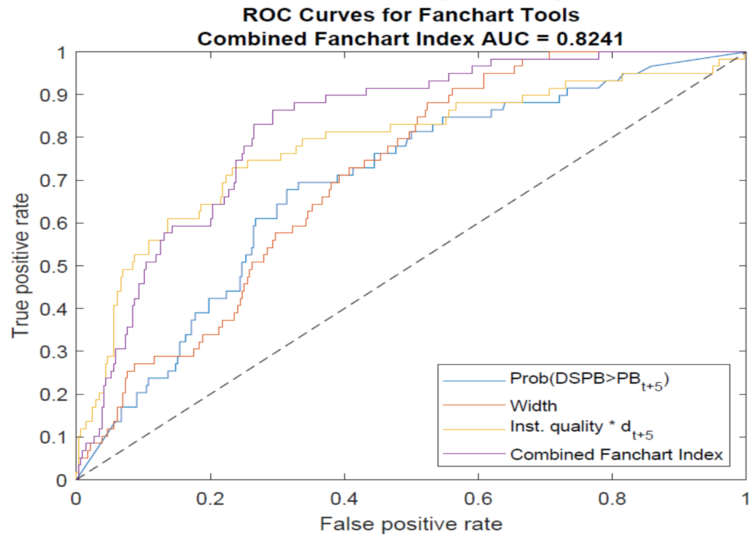
[7] Although fanchart metrics and signals are available for more recent period, the stress outcomes associated with these signals cannot be observed for the full medium term (5-year) prediction period. Hence, predictive performance is analyzed based on fanchart signals between 2010 (the earliest period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

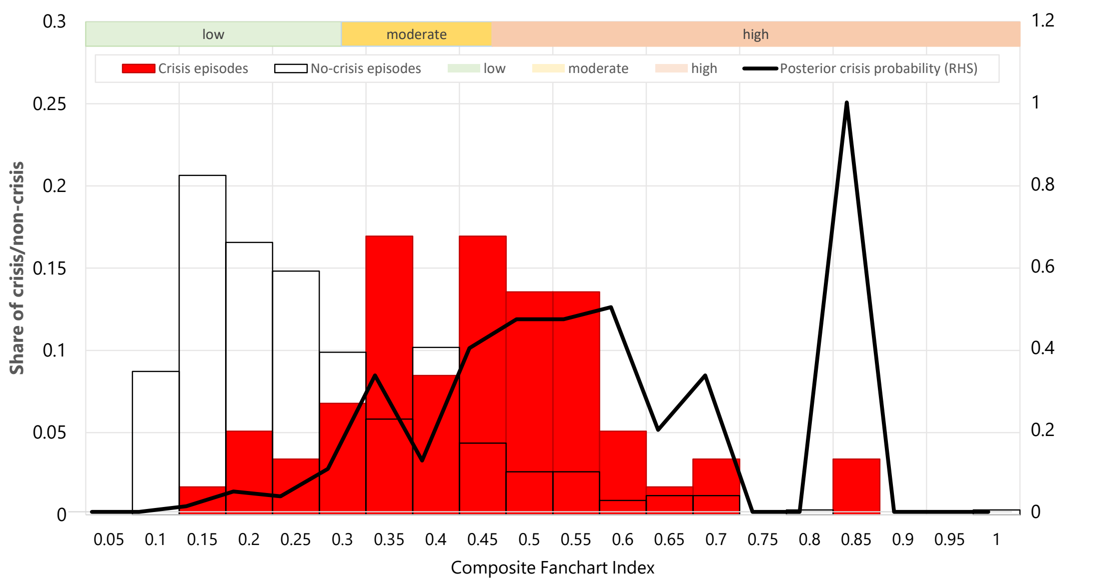
are fewer observations), there is a clear pickup in the posterior stress probabilities around the "low/moderate" and "moderate/high" thresholds. The figure suggests a posterior probability of stress conditional on a GFI "high risk" signal of at least 40 percent, and a posterior probability of stress 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:**

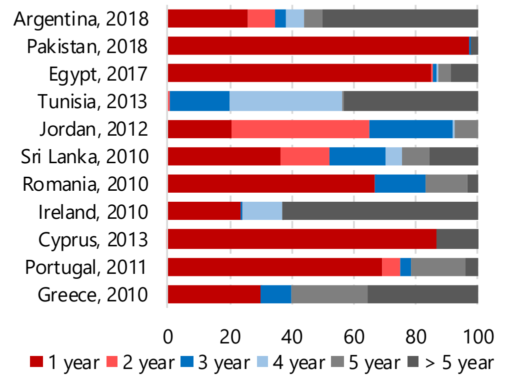
The macro-fiscal shocks are broadly similar to those of the existing MAC DSA's stress tests and include (i) a one standard deviation (computed over the last 10 years) reduction in the real GDP growth rate for two years; (ii) for countries outside currency unions and having their own legal tender, a one-year 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.

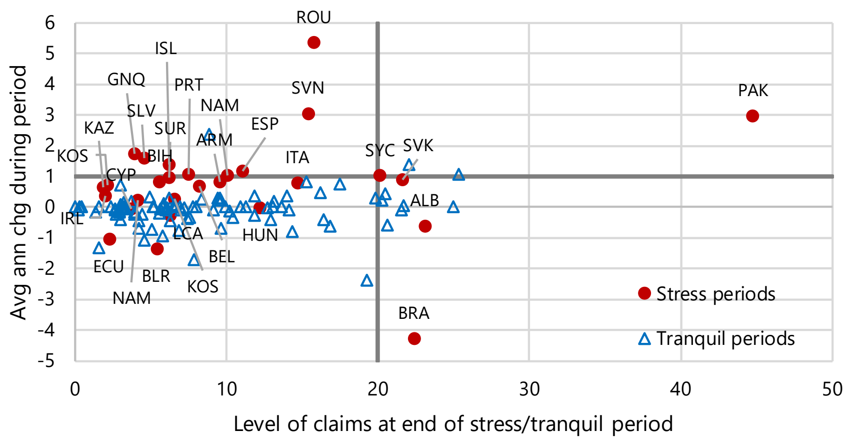
severe, the thresholds that determine the mechanical signal associated with the test would be set higher than if the shock were defined to be less severe.

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]

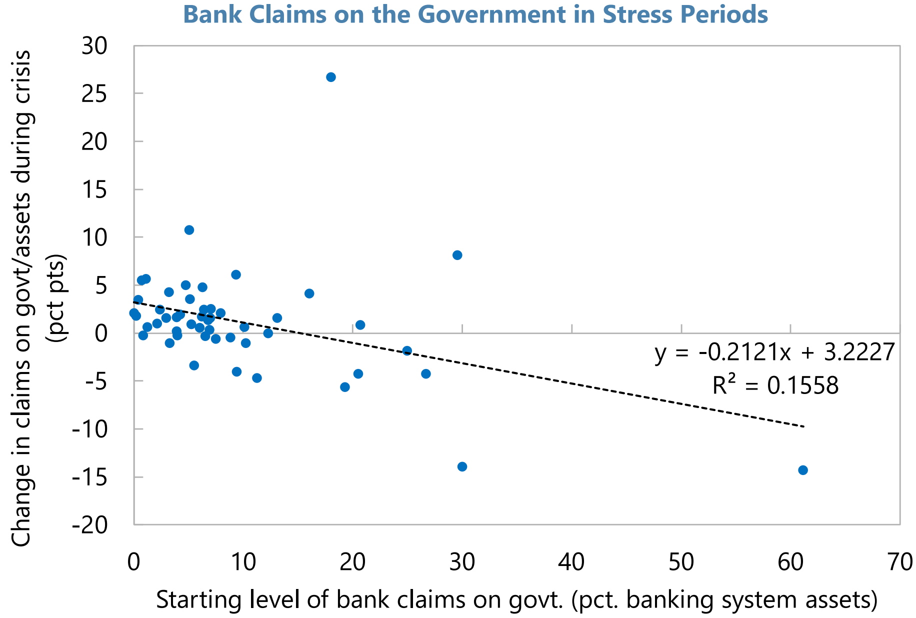
Bank Claims on Government/Total Bank Assets—Stress Vs. Tranquil Periods

Evolution of Bank Claims on Government/Total Bank Assets



Sources: International Financial Statistics and Fund staff calculations.

Bank Claims on the Government in Stress Periods



Sources: International Financial Statistics and Fund staff calculations.

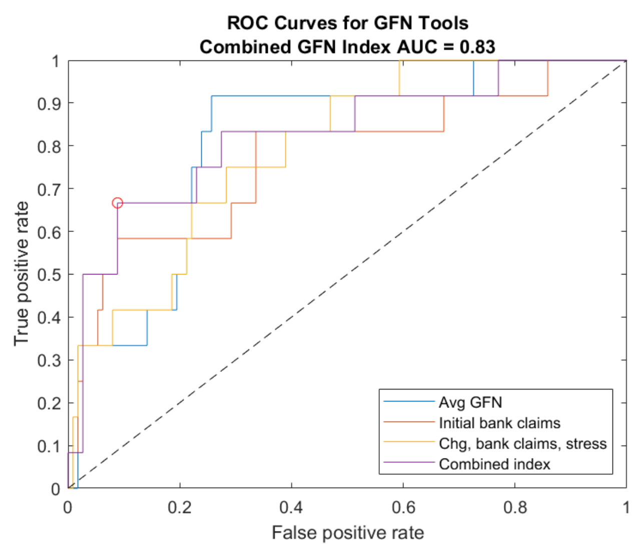
[1]These findings are also consistent with Arslanalp and Tsuda (2014), who used 15 percent of assets as a risk threshold.

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:**

Figure AVII.2 ROC Curves for GFN Module Risk Indicators



Source: Fund staff calculations.

- *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.

- *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).

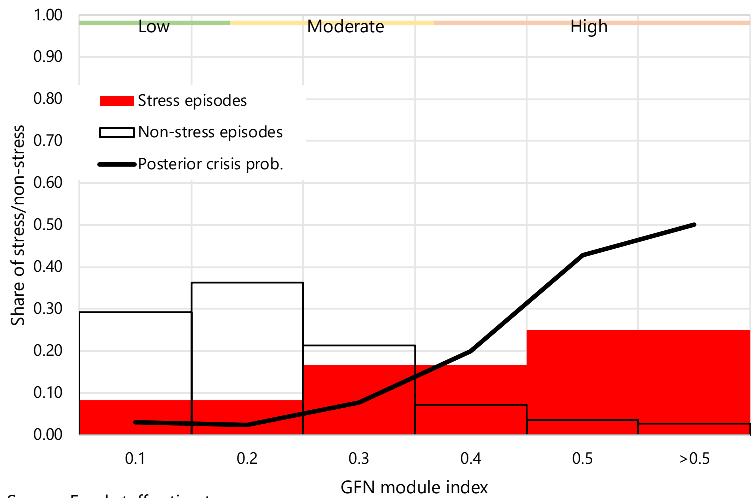
10. **To aggregate the information from these three indicators, staff combined them into an aggregate GFN Financeability Index (GFI), weighted by their explanatory power (i.e., their AUC).** This overall index has an AUC of 0.83, implying that it is an improvement over each of these indicators in isolation (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 95^{th} 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.

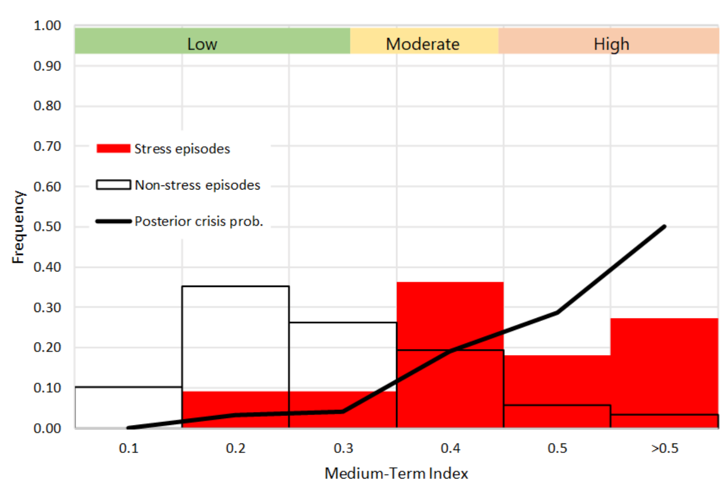
Figure AVII.3. GFN Financeability Index Distributions in Stress and Non-Stress Cases



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: Medium-term: 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.