**Review of the Debt Sustainability Framework for Low Income Countries (LIC DSF)**

**Discussion Note**

**August 1, 2016**

## I. BACKGROUND AND CONTEXT

1. **The LIC DSF, introduced in 2005, remains the cornerstone of assessing risks to debt sustainability in LICs and has key operational implications for setting IMF and World Bank debt conditionality and for the allocation of concessional financing by some multilateral development institutions.** The LIC DSF has been reviewed on three occasions: 2006, 2009, and 2012. The most recent review took steps to strengthen the analysis of total public debt, formally introduced remittances in determining debt thresholds, and introduced the option of using an alternative approach to complement the assessment of the risk of debt distress in countries classified as "borderline" cases. Annex I describes the main features of the LIC DSF following the 2012 reform.

2. **The financing landscape for LICs has evolved substantially since the early years of the DSF and changed the nature of the risks facing them.** A recent joint IMF-WB Board paper on *Public Debt Vulnerabilities in LICs: The Evolving Landscape* (IMF, WB 2015) shows that LICs are increasingly exposed to a wider set of vulnerabilities, including from market volatility and costlier debt, an environment many may not be familiar with. The challenging global environment suggests that debt vulnerabilities are likely to increase for many of these countries.

3. **Against this background, the ongoing review of the LIC DSF is timely, seeking to assess the performance of the DSF in areas that may warrant change and propose reforms to strengthen it.** The review will examine how well the framework has done in anticipating debt developments and highlighting risks to debt sustainability. It will then propose specific reforms to strengthen these aspects of the DSF. A broad objective of the review is to better align the DSF with the evolving nature of risks facing countries, in a context where many countries will be seeking funding to boost public investment. Selected country case studies will also be used to help illustrate the issues.

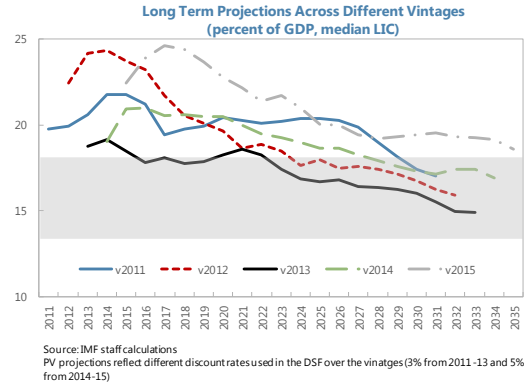
4. **This note, which is part of a wide consultation process with external stakeholders, presents staffs' preliminary findings and consideration of reforms.** It reflects work in progress and should not be seen as conclusive. It is primarily aimed to generate discussion both on the backward looking assessment and how to approach reforms to inform the final proposals in this regard. Feedback from external stakeholders will be reflected in the final review document. It is expected that the review will be presented to the Executive Boards of the IMF and World Bank in December 2016. Following Boards' approval of reforms, guidance will be prepared for the implementation of the revised DSF.

## II. Assessment of LIC DSF Performance

5. **Based on an analysis of the DSF performance to date, a number of issues have been identified.** These issues relate to both the ability of the DSF to anticipate debt developments and risks and the methodology used to derive the risk rating.

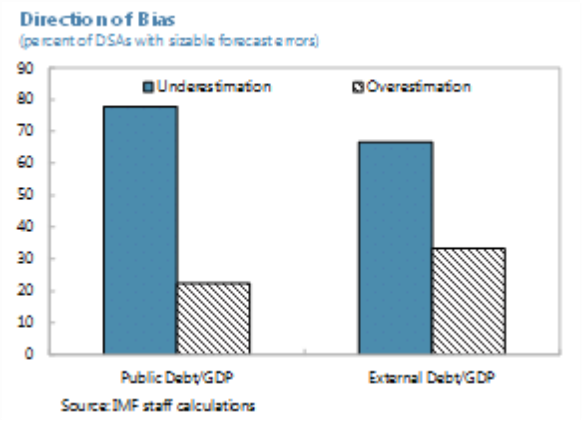
### A. Anticipating debt developments

6. **While debt projections over the near term have been broadly accurate, sizable forecast errors have been consistently detected over the medium term for both external and total public debt to GDP ratios.** Over 35 percent of DSAs produced during 2007-2010 contained medium term (5-year horizon) forecast errors greater than 15 percentage points in absolute value.[1] This result is consistent across country groups such as frontier LICs and commodity exporters. However, for small states, more than 50 percent of DSAs show sizable forecast errors, possibly reflecting their exposure to natural disasters. It is worth noting that the median present value of external public debt to exports projection across LICs for the last five DSA vintages broadly converges to a common long term level (Figure 1), suggesting a rigidity in forecasts which could be contributing to high forecast errors.



Source: IMF staff calculations
PV projections reflect different discount rates used in the DSF over the vinatges (3% from 2011-13 and 5% from 2014-15)

7. **The majority of DSAs with sizable medium-term forecast errors underestimated debt outcomes (Figure 2).** A breakdown of the medium-term forecast errors between those with debt outcomes greater and lower than projections shows that there is a stronger downward bias. Among external and public DSAs with sizeable forecast errors (larger than 15 percentage points in absolute value), about 70 and 80 percent, respectively, underestimated debt outcomes. This may reflect a late recognition of vulnerabilities and relatively more volatile macro environment for certain groups of countries.



Source: IMF staff calculations

[1] These estimates are broadly in line with those of the 2012 LIC DSF review. At that time, with limited data for a longer projection horizon analysis, it was found that 30 percent of DSAs produced during 2006-2007 had sizable forecast errors (larger than 15 percent of GDP in absolute value) over a 3-4-year horizon.

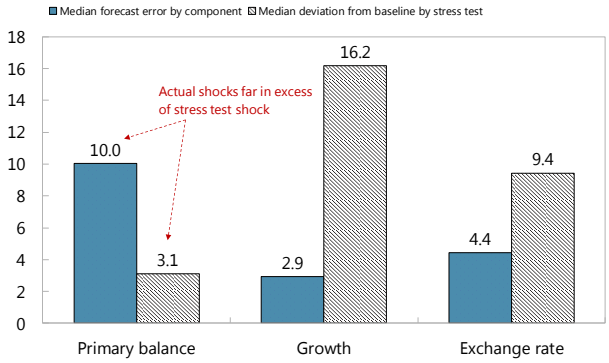
8. **Unexpected changes in debt ratios are primarily driven by fiscal deviations and balance of payment (BOP) shocks.** A decomposition of forecast errors of public and external debt over a 5-year horizon indicates that:[2] i) the 5-year forecast errors of total public debt were mainly driven by deviations of outcomes from projections of the primary deficit and unexplained factors as captured by an unexpected positive residual; and ii) trade balance changes and financial account flows are the major contributors to unexpected changes in external debt.

## B. Assessment of stress tests

9. **Recent evidence suggest that debt outturns have generally been within the debt ratios projected by the most extreme shocks.** In 14 and 15 cases out of 121 DSAs conducted in 2014-15 (about 12 percent),[3] the first-year projection for the PV of debt as a percentage of GDP and exports, respectively, exceeded the ratio projected for that year under the most extreme stress test in DSA vintages conducted three or four years ago. This suggests that, on average, stress tests may be too conservative in flagging risks facing LICs.

10. **At the same time, there seems to be a mismatch between the impact of stress tests in the DSF and actual behavior of the stressed variables.** For example, an assessment of the behavior of macroeconomic variables through time suggests that the magnitude of primary balance and BOP shocks analyzed in the DSF is relatively benign compared to the scale of such shocks in reality, while the magnitude of the exchange rate shock appears to be too extreme (Figure3).[4] This, along with the fact that the impact of stress tests has not adequately reflected the actual sources of forecast errors, suggest that stress tests need to be recalibrated (Figure 4).

**Sources of forecast error: modeled vs. actual**
(Percent, total public debt-to-GDP)



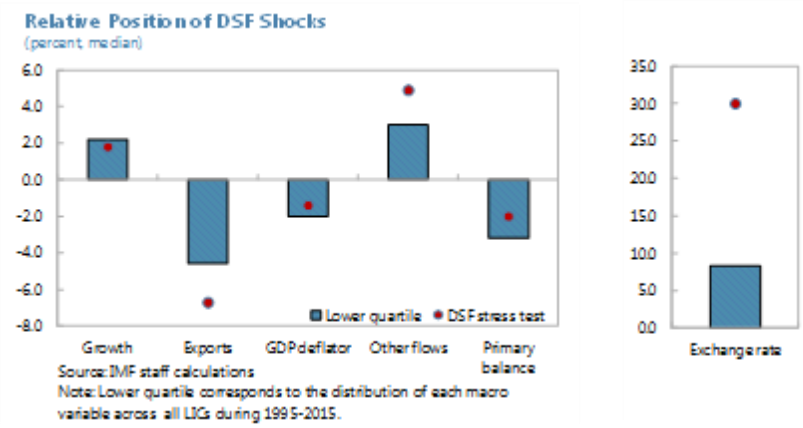

Sources: IMF Staff Calculations

[2] The assessment was done for all DSAs produced during 2007-2010 with sizable positive forecast errors.

[3] This sample covers the most recent DSAs produced over 2014-15.

[4] The assessment compares the median value of each relevant macro variable (real GDP growth, export growth, GDP deflator, primary balance, and exchange rate) after a DSF shock is applied (the "DSF post-shock value") to the actual distribution of each macro variable. A DSF shock is considered relatively benign (extreme) if the DSF post-shock value is above (below) the 25$^{th}$ percentile of its distribution. The opposite applies to assessment of the exchange rate shock, since the lower values of its distribution capture the lowest depreciation rates, therefore DSF post-shock values for exchange rates falling below the 25$^{th}$ percentile of the distribution would suggest rather muted shock.

**Figure 4**




11. **In spite of the large number of stress tests included in the DSF, some important risks facing sub-groups of LICs have not been adequately covered.** LICs face debt pressures emanating from natural disasters, the materialization of contingent liabilities, and volatile commodity prices. While some of these shocks are indirectly covered by the existing stress tests (for example, export shocks may partially capture the effect of commodity price shocks), they are not explicitly evaluated by dedicated scenarios. As a result, assessments of these risks have been ad hoc and therefore lack cross-country comparability.

12. **Similarly, the framework does not provide explicit tools for assessing risks associated with market financing**. The DSF assessment of risks, in particular thresholds and benchmarks for debt service indicators as well as stress tests, was created based on the experience at a time when market access was rare among LICs. In recent years, a number of frontier LICs have been able to increase market and other forms of non-concessional financing. In the current environment, these countries face greater rollover risks because of a shorter debt maturity profile and exposure to market volatility. However, broader indicators of financing risk, like gross financing needs and market perception of risk, are not covered.

## C. Determining risk ratings

13. **Evidence from the past ten years suggests that risk ratings have sometimes not provided adequate signaling of impending debt distress.** To assess this, available risk ratings over the past ten years were compared against actual external debt distress events.[5] During this period, 7 countries experienced debt distress. Of those, in only one case the risk rating was high in the year of or preceding the debt distress event; two were deemed low risk, and four moderate risk.

[5] Debt distress events include those that involve liquidity stress.

14. **An assessment of features introduced into the DSF in 2012 –to allow use of more country-specific information –shows that they have helped clarify risk ratings in only a handful of cases**.

- **The probability approach** was used in only nine out of 68 possible cases. Of those, the final rating of only one country reflected the findings of the probability approach. In general, the use of the probability approach for all countries (not only borderline cases) would have resulted in higher risk ratings, particularly for countries in the moderate risk category.

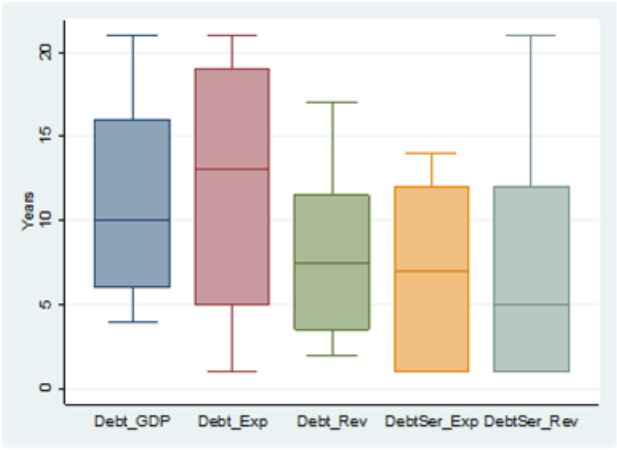
- **The remittances-augmented thresholds** were used in only 13 out of 68 countries. Of those, five saw a lower risk rating relative to the rating that would have resulted if the standard thresholds were used. It is worth mentioning that extending the application of the remittances-augmented debt thresholds for all countries (not those that qualified) since 2014 for which remittances data are available, would not lead to any improvement in the risk rating. In fact, a number of countries would see a worsening in their risk ratings.

- **An overall risk rating was reported in only 10 countries since 2014.** Only 6 of those provided an in-depth discussion of domestic public debt vulnerabilities. Most DSAs did not provide a discussion of the extent of vulnerabilities stemming from rollover risks, the increasing participation of non-residents in domestic local-currency bond markets, as well as from the structure of domestic public debt.

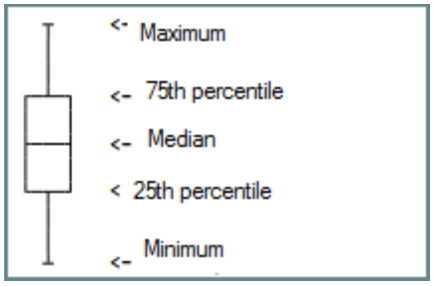
- **There appears to have been a bunching of countries within the "moderate risk" category.** There is great diversity in the number of years and magnitude of breaches of thresholds within this risk category. Figure 4 shows, for the most recent DSA vintages produced from 2014-15, a significant diversity in the number of years for which each debt burden indicator breached thresholds under stress tests for countries assessed as moderate risk; with some countries having only one-off breaches while others showing protracted ones. A similar assessment done for size of breaches of debt thresholds also revealed a significant diversity among this group, with some countries having sizeable breaches of thresholds (suggesting the rating could be high risk), while others showing much more minor breaches (suggesting a low risk could be more appropriate).[6]

[6] For each country, diversity in size is calculated as the difference between the debt burden indicators (i.e. PV of debt to GDP, exports, revenue ratio and the debt service to exports and revenue) under the most extreme shock and the corresponding DSF threshold for each country.

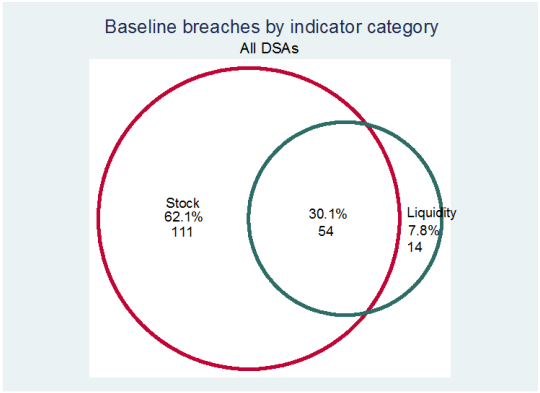
**Figure 4 Dispersion within moderate risk countries: Number of years breaching thresholds**




*Note: The box plots show the medians, the min-max values and the interquartile ranges of the sample as described below.*




15. **Some debt burden indicators have played a bigger role than others in determining the risk rating.** An assessment of all DSAs done since the inception of the framework indicates that the PV of external debt to fiscal revenues has rarely played a determining role in the risk rating. In cases where this indicator breached a threshold, at least one other debt indicator thresholds was also breached. Breaches of debt service thresholds under the baseline scenario have been much less frequent than breaches of stock thresholds. Only 14 out of 179 DSAs with breaches of thresholds under the baseline had breaches of the debt service thresholds without any breach of the stock thresholds. The high debt service thresholds are an outcome of the design of the LIC DSF as the current definition of debt distress episodes used to estimate debt thresholds captured mainly severe debt distress events.



## III. Potential Reforms to the DSF

16. **Based on the initial findings of the DSF performance, a number of reforms are under consideration.** To ensure that the DSF is aligned with the evolving nature of risks facing countries, changes to improve the accuracy of debt projections and stress testing, and changes in and simplification to the over-arching framework for determining risk ratings will be considered.

### A. Enhancing the accuracy of debt projections and identification of risks

17. **The review will aim to develop tools to highlight potential optimism/pessimism in the projections underlying the DSA.** The focus will be on key relationships (e.g. fiscal adjustment and growth projections, investment-growth nexus) and on variables shown to be most responsible for projection errors. While projections of macroeconomic variables take place outside the DSF, they are an input to the framework, and thus their accuracy is critical for the quality of DSAs.

18. **The track record of stress tests in the DSF suggests a need to recalibrate, refocus, and selectively add stress scenarios.** Staffs proposes to (i) focus the stress tests on those shocks that have been shown to play a lead role in explaining projection errors; (ii) improve the calibration of these shocks, including to better reflect the dynamic interaction of the macroeconomic variables after a shock is applied where relevant; and (iii) introduce a menu of shocks to be applied where relevant to capture important risks currently not in the DSF, while accounting for the large diversity of LICs. Key shocks to consider as part of this menu include natural disaster shocks for countries prone to such events; contingent liabilities that may materialize from state owned enterprises and public-private partnerships, where these present risks; terms of trade shocks for commodity exporters; and tightening of market conditions for market access countries.

19. **In addition to the above, the merit of further analysis of market-related risks for countries that have or are likely to have market access will be considered.** In this regard, staffs are considering developing benchmarks for a number of indicators that are key to assessing liquidity pressures. This includes gross financing needs, the EMBI spreads, and the relative importance of domestic public debt.

### B. Improving the risk assessment in the DSF

20. **A core element of the review will be to re-estimate and improve the underlying model used to determine debt thresholds.** With a view to reflecting developments in the past ten years, staffs are considering a number of improvements to the econometric model underlying the DSF along several dimensions. First, in light of the findings regarding the limited ability of the framework to capture both solvency and debt service related events, staffs will look more closely at the definitions of external debt distress episodes and of debt distress signals to establish whether they adequately capture all relevant debt distress events

and, if not, how they should be modified. In addition, with a view to improving the fit of the model, staffs are exploring the merit of reflecting the impact of additional macroeconomic variables that play an important role in debt sustainability. Staffs are considering the robustness of the CPIA measure of governance and institutional capacity relative to other measures and CPIA sub-components

21. **The review will examine how debt distress thresholds—and thus ratings—are derived.** Staffs aim to improve the process of deriving risk ratings by assessing the merit of: i) using more information (beyond just the CPIA); ii) providing more granularity to the risk rating categories, particularly for the moderate category, and iii) streamlining the number of debt indicators used to derive risk ratings. This will be informed by the analytical findings regarding the usefulness of the different indicators in the DSF.

22. **The review will also examine how to improve the assessment of the risk of debt distress, given updated thresholds for indicators.** Consideration will be given to refining the guidance on how to account for the timing and extent of breaches of thresholds when assigning risk ratings. It will also consider alternatives to reflect the effect of remittances in countries where these are significant in order to balance the need for streamlining the framework with the need to account for the importance of remittances in those countries. In addition, to better guide policy responses by debtors and creditors, consideration will be given to providing more guidance to detect instances of unsustainable debt (as distinct from high risk of debt distress) before arrears begin to accumulate (i.e., before debt distress happens).

23. **The review will also explore the merit of introducing formal thresholds on total public debt, a decision that will be guided by the empirical findings**. To the extent that the findings are robust, staffs will consider the merit of introducing a formal risk rating on overall public debt to be used alongside the external risk rating.

## Annex I. Elements of the LIC DSF

*This annex describes the main features of the IMF-WB debt sustainability framework for low-income countries (LIC DSF). The framework produces a formal risk rating for external debt distress, determined by comparing projected levels for five external public debt burden indicators to thresholds derived from econometric analysis. Risks associated with public domestic debt or external private debt may qualify the external risk rating.*

### I. Model underlying the LIC DSF

**The probability of debt distress is estimated using a probit model.** The dataset used in the last review of the framework included 130 countries—61 LICs and 69 MICs—for the period 1970-2007.[7] During this period, 105 debt distress episodes and 654 non-distress episodes were identified. The model is as follows:

$$P(\text{debt distress}_t) = \Phi(\beta_1 * \text{debt burden}_{t-1} + \beta_2 * \text{governance}_{t-1} + \beta_3 * \text{shock}_{t-1} + \beta_4 * \text{other}_{t-1}) \ldots (1)$$

where debt distress equals 1 when the country is in debt distress (0 otherwise), and other variables as defined in Table 1.

**Table 1. Probit Variables**

| | For external debt thresholds | For public debt benchmarks |
|---|---|---|
| **Debt burden 1/** | PV of PPG external debt or debt service scaled by GDP, exports, or government revenue | PV of total public debt scaled by GDP |
| **Governance** | CPIA score | |
| **Shock** | Real GDP growth | |
| **Other** | Interaction dummy (for MIC debt burden) | |

1/ For countries where remittances contribute a significant share of GDP and export receipts, the debt burden indicators, and accordingly the thresholds, are adjusted to allow for their role.

Source: Annex I, "Revisiting the Debt Sustainability Framework for Low-Income Countries," IMF and World Bank 2012.

### II. Debt burden thresholds/benchmarks

**Debt burden thresholds are derived using the following equation:**

$$Threshold = \frac{\Phi^{-1}(P^0(debt\ distress)) - (\widehat{\beta_2} * \overline{governance} + \widehat{\beta_3} * \overline{shock} + \widehat{\beta_4} * \overline{other})}{\widehat{\beta_1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Thresholds are calibrated for three values of the CPIA associated with weak, medium, and strong governance (3.25, 3.5, and 3.75, respectively). The shock variable is set to

[7] See "Revisiting the Debt Sustainability Framework for Low-Income Countries," IMF/World Bank, January 2012 and references cited there.

the average real GDP growth for all LICs in the sample. Fixing the values of governance, shock, and "other", the model varies the value of the probability of debt distress until it yields the debt burden threshold that minimizes the sum of Type I and Type II errors in the debt distress sample.[8] The derived thresholds are then rounded for operational purposes. Table 2 shows the calibrated debt burden thresholds associated with the probability levels that minimized the sum of Type I and Type II errors.

**Table 2. LIC DSF Debt Burden Thresholds and Implied Probability Thresholds**

| | PV of PPG external debt as percent of | | | PPG external debt service as percent of | |
|---|---|---|---|---|---|
| | GDP | Exports | Revenue | Exports | Revenue |
| Weak | 30 | 100 | 200 | 15 | 18 |
| Medium | 40 | 150 | 250 | 20 | 20 |
| Strong | 50 | 200 | 300 | 25 | 22 |
| Probability | 14% | 13% | 15% | 14% | 15% |

Source: Table A3, Annex I, "Revisiting the Debt Sustainability Framework for Low-Income Countries," IMF and World Bank 2012.

## III. Stress tests

**The external (public) DSA considers two (three) alternative scenarios and six (five) bound tests.** Alternative scenarios represent permanent shocks over the projection period while bound tests are temporary shocks that last one to two years, after which the modified variables return to baseline values. The stress tests are primarily standalone shocks, except for the combination shock scenarios in which multiple shocks are introduced. No explicit correlations among the macro variables are assumed.

## IV. Risk ratings

**The LIC DSF assigns an explicit rating for a country's risk of external debt distress.** In the external DSA, debt burden indicators are compared to their respective thresholds. If all indicators are below their thresholds, the country is assigned a **low risk rating**. If any one indicator breaches the relevant threshold in at least one of the stress tests, the assigned risk rating is **moderate**. If one or more thresholds are breached in the baseline scenario but the country is not facing payment difficulties, the **high risk rating** is assigned. Lastly, a situation where either (i) current debt and debt service ratios are in significant or sustained breach of the

[8] A Type I error denotes the failure to predict a distress episode (missed crisis) while a Type II error denotes the failure to predict a non-distress episode (false alarm).

thresholds, or (ii) there is an actual or impending debt restructuring, or (iii) the country is in arrears on external debt service is taken as indicating that the country is **in debt distress**. While the indicative thresholds play a central role in the determination of the risk rating, staff is expected to apply judgment in the final risk rating by considering, for example, the magnitude, duration, and number of breaches, or the relevance of a given stress test.[9] When a country is on the border between two categories, the probability approach (described below) may be used to help determine the risk rating.

**When relevant, the LIC DSF also includes an assessment of the country's overall risk of debt distress.** The evolution of overall public debt ratios is compared to the relevant public debt benchmarks, derived as described above. Overall public debt includes both external and domestic debt.] If public domestic debt or private external debt levels are a significant cause of concern, the LIC DSF assessment will also report the overall risk of debt distress to flag potential vulnerabilities not captured by the external risk rating.

## V. Probability thresholds for the probability approach

**At times comparing the baseline and alternative scenarios to the relevant thresholds produces inconclusive assessments of debt sustainability.** Countries where the highest projected values of a debt burden under the baseline and standardized stress tests fall within 10 percent of the relevant thresholds (i.e., a breach or near breach) are considered to be "borderline cases." When the risk rating straddles two risk categories, the "probability approach" may be used to help make the final assessment.

**Under the probability approach, the evolution of a country's probability of debt distress over time is compared to probability thresholds.** In this approach, the country's probability of debt distress is estimated for each debt burden indicator (as in equation (1)) except that the model uses the country specific (continuous) 3-year moving average CPIA score (rather than the weak, medium, and strong categorization) and the 25-year average real growth rate (5 years of history and 20 years of projection). The estimated probability paths associated with the 5 debt burden indicators are then plotted against their respective thresholds; these threshold levels (shown in the last row of Table 2), are the probability values implied by the current LIC DSF debt burden thresholds estimated in Equation (2).

[9] There is no specific guidance on how these various factors should be weighed.