

Getting institutional assessments right

How qualitative scores shape
sovereign credit ratings

February 2026



Getting institutional assessments right

How qualitative scores shape
sovereign credit ratings

Acknowledgements

This paper was authored by Arend Kulenkampff. The author would like to thank Vera Songwe, Guillaume Thomassin, Hanan Amin Salem, Julie McCarthy, Justin Mundy, Isobel Cohen, Teal Emery, Chris Humphrey, and Erik Berglof for their contributions and valuable insights on this topic. NatureFinance (Amandine Ambregni, Lucy Martin, Natan Aquino, Roberta Zandonai), the Sustainable Sovereign Debt Hub (Barbara Oldani, Martina Tamvakou, Eva Sirp, Shereen Wiseman), and Justine Doody provided valuable support.

The views expressed in this paper are those of the authors alone and any potential errors/inaccuracies are their sole responsibility.

This work has been made possible through the support of the Children's Investment Fund Foundation (CIFF).

Comments and queries can be addressed to: arend.kulenkampff@naturefinance.net



Table of Contents

Executive summary	5
Section 1.	
Background and context	6
Section 2.	
Quality and transparency gaps in qualitative rating factors	12
Section 3.	
Solutions and recommendations	27
Conclusion	35
References	36

Abbreviations and acronyms

ADB	Asian Development Bank
AfDB	African Development Bank
APRM	African Peer Review Mechanism
AI	Artificial intelligence
BTI	Bertelsmann Transformation Index
CC	Control of Corruption
CCM	Country Ceiling Model
CIFF	Children's Investment Fund Foundation
Climate.VS	Climate Vulnerability Signals
CPI	Consumer Price Index
CPIA	Country Policy and Institutional Assessments
CRA	Credit rating agency
CRAM	Country Risk Assessment Model
CRE	Country Risk Expert group
DECVP	Development Economics Vice Presidency
DeMPA	Debt Management Performance Assessment
DI	Democracy Index
ECA	Export credit agency
EIU	Economist Intelligence Unit
EMDE	Emerging markets and developing economies
ESG	Environmental, Social, and Governance
GE	Government Effectiveness
GDP	Gross domestic product
GFS	Government Finance Statistics
HRMI	Human Rights Measurement Initiative
IIF	Institute of International Finance
IISD	International Institute for Sustainable Development
IMF	International Monetary Fund
IOSCO	International Organization of Securities Commissions
KPI	Key performance indicator
LLM	Large language model
LMIC	Low- and middle-income countries
LSEG	London Stock Exchange Group
MPS	Mapping Pathways for Security
NDB	National development bank
NOC	National oil company
OECD	Organisation for Economic Co-operation and Development
OPM	Free Papua Movement (Organisasi Papua Merdeka)
PEFA	Public Expenditure and Financial Accountability
PFM	Public financial management
PNG	Papua New Guinea
PTS	Political Terror Scale
PV	Political Stability and Absence of Violence
QO	Qualitative overlay
RL	Rule of Law
RQ	Regulatory Quality
RRI	RepRisk Index
SDDS	Special Data Dissemination Standard
SOE	State-owned enterprise
SRDSF-MAC	Sovereign Risk and Debt Sustainability Framework for Market Access Countries
SRM	Sovereign rating model
T&C	Transfer and convertibility
UNCTAD	United Nations Trade and Development
UNDP	United Nations Development Programme
VA	Voice and Accountability
WGI	Worldwide Governance Indicator

Executive Summary

As gatekeepers to international capital markets, credit rating agencies (CRAs) are critical nodes in the global financial architecture. Their assessments of governments' "willingness and ability" to honor sovereign debt obligations have a direct bearing on the cost of capital in emerging markets and developing economies (EMDEs). Transparent and unbiased rating models and high-quality inputs are essential to secure financing on fair terms, yet the "Big Three" CRAs—Fitch Ratings, Moody's, and S&P Global—fall short in this regard. Systematic gaps in their rating methodologies may be contributing to excessive risk premia and diverting urgently needed capital away from EMDEs. In the current context of historic public debt levels, heavy debt-servicing burdens, and escalating budgetary needs for security and climate resilience, any distortions in sovereign risk pricing increase the threat of debt distress for EMDE sovereigns.

Building on the growing body of research into CRAs as drivers of EMDE capital flows, this report focuses on the qualitative inputs into sovereign ratings. It complements ongoing efforts by market participants to improve the quality and transparency of quantitative data such as macroeconomic statistics and fiscal risk metrics. Shining a spotlight on how qualitative scores and third-party assessments—notably the World Bank's Worldwide Governance Indicators (WGIs)—influence the Big Three CRAs' ratings, the report reveals a fundamental asymmetry at the heart of sovereign risk analysis: those rating factors that are amenable to statistical measurement tend to be assessed in a relatively standardized and objective manner, whereas those that are based on subjective "expert judgments" are prone to inconsistent or even arbitrary treatment.

This "soft" side of sovereign risk analysis has received comparatively less attention from researchers and advocates of CRA reform. However, assessments of institutional quality, political risk, and other determinants of "willingness to pay" have disproportionate weight in rating decisions. Because they often cannot appeal to standard metrics or objective benchmarks when assigning qualitative scores to these factors, their judgment calls are susceptible to errors and bias that are difficult to spot and contest.

The report documents these critical issues and presents recommendations for how to address them, both via technical fixes to the existing methodologies as well as through more fundamental changes in approach. The latter include alternative scoring models that leverage latest advances in artificial intelligence (AI) such as super-forecasting and large language models (LLMs). The report also considers alternative ways to conceptualize and track willingness to pay using performance-based frameworks as markers of prudent fiscal governance.

Innovative approaches are needed to cope with the profound changes in the information landscape confronting CRAs. Emerging risks such as climate change, geoeconomic fragmentation, and technological disruption broadens the scope and complexity of sovereign analysts' job. The rise of AI, while offering new analytical tools, may also make it harder to parse signals from noise. Failure to tackle these challenges by updating their qualitative assessment methodologies threatens to erode the credibility of CRAs' ratings and undermine the legitimacy of their pivotal role in channeling capital flows to EMDEs.



Section 1

Background and context

Section 1

Background and context

As gatekeepers to international capital markets, sovereign credit rating agencies (CRAs) have a major bearing on capital flows to emerging markets and developing economies (EMDEs). The sovereign credit ratings of the Big Three agencies (Fitch Ratings, Moody's Ratings, and S&P Global) send powerful signals to international creditors about the relative risk of investing in the debt instruments of sovereign issuers across the globe. These signals are in turn reflected in the interest rates and borrowing terms that sovereigns face in international markets.

Sovereign issuers that depend on external financing to cover budgetary needs rely on the due diligence of CRAs to accurately assess their "ability and willingness to pay," and on that basis assign ratings that best match the credit fundamentals. This is especially critical for the EMDEs that currently face mounting fiscal pressures from heavy debt servicing bills and contingent liabilities from climate-related disaster spending. Ratings are a key driver of public debt dynamics. For the 54% of low- and middle-income countries (LMICs) currently facing debt distress (World Bank 2025a), the prospect of downgrades threatens to push some of them into default.

As the interlinkages between debt distress and development traps come into sharper relief, the rating practices of the Big Three CRAs are drawing public scrutiny and accusations of malpractice (for example, see Ryder et al. 2024). Fairly or not, the view that sovereign ratings are biased and distorting sovereign risk pricing has taken hold amongst certain market commentators and development practitioners. The high sovereign borrowing costs of sub-Saharan African countries relative to peers in other regions is cited as evidence of faulty methodologies and questionable judgment calls (Griffith-Jones and Kraemer 2021; Cash and Khan 2024; Tinubu 2026; UNDP 2025; UNCTAD 2025).

Several studies have investigated these claims (Barta and Johnston 2023; Gbohouni et al. 2023; Kulenkampff 2025), but have not found conclusive evidence of a systematic cultural, political, or ideological bias on the part of CRAs. Others have pointed to gaps on the issuer side with debt reporting and data quality that may be skewing investors' risk perceptions (UNDP 2025; Tshabalala 2025; World Bank 2025c). These concerns are valid, and efforts to improve reporting are welcome, yet they tend to focus on only a subset of data: quantitative macro-fiscal statistics.

Data quality and transparency considerations differ for quantitative and qualitative rating factors. For the former, the issues pertain mainly to how the metrics (e.g., inflation, economic growth, public debts and deficits, current account balances, etc.) are compiled. Quality varies widely from country to country due to the institutional capacity constraints of statistical bureaus and the lack of a common global authority to enforce uniform reporting requirements on sovereigns. Even so, the definition and measurement of key quantitative data for sovereign risk analysis are largely standardized and can be evaluated using relatively objective statistical benchmarks.¹

Qualitative factors in sovereign risk analysis tend to be more ambiguous and harder to measure. This is either because the phenomena are intrinsically non-numerical and descriptive (e.g., political stability), or because the tools to measure them are not available or not sharp enough (e.g., corruption). They cover a wide range of subjects (e.g., institutional quality, political event risk, and regulatory quality), of which many fall under the umbrella of "governance". Qualitative factors are typically assessed along ordinal scales, with values assigned by in-house analysts or outside experts at development banks, think tanks, and commercial research firms, or coded according to opinion surveys conducted by independent pollsters.

¹ For instance, many sovereigns are aligning with the International Monetary Fund's (IMF) Government Finance Statistics (GFS) Manual and Special Data Dissemination Standard (SDDS), which "guide members that have, or might seek, access to international capital markets in the provision of their economic and financial data to the public." See IMF (n.d.).

Scores and surveys cover both *de jure* features of the legal, regulatory, and institutional environment— e.g., the enactment of fiscal responsibility laws— as well as their *de facto* implementation—e.g., the incidence of corruption. Whereas the former are relatively easy to track, the latter reflect the subjective perceptions of the individuals being consulted. This makes them susceptible to errors and bias that can negatively skew the ratings of certain countries.

Reference criteria, quantitative “guideposts” such as third-party indicators, and review by the internal credit committees are meant to mitigate this risk. Yet these checks and balances often fall short for several reasons discussed below. The next section will examine these issues in closer detail through a quality and transparency lens, while Box 1 summarizes how the Big Three CRAs incorporate qualitative factors into their methodologies.

Box 1. Qualitative factors in the Big Three CRA

The Big Three CRAs have distinct approaches to combining hard and soft factors in their respective methodologies. S&P and Moody’s both use scorecards to determine their ratings, whereby the sovereign analysts assign weighted scores to a defined set of factors that roll up into an indicative rating. S&P grades along a six-point numerical scale (1 for strongest to 6 for weakest) across five thematic pillars (“institutional,” “economic,” “external,” “fiscal,” and “monetary”), whereas Moody’s allows for finer gradations along a 17-grade scale (from *aaa*, at best, to *caa1*, at worst) across four factors (“economic strength,” “institutions and governance strength,” “fiscal strength,” and “susceptibility to event risk”). In both cases, adjustments can be made to each of the sub-factors as well as to the final rating.

Fitch employs a proprietary econometric sovereign rating model (SRM) based on 18 variables to generate an indicative quantitative score, which is adjusted via “qualitative overlay” (QO) to arrive at the final rating. The agency considers four analytical pillars (“structural features,” “macroeconomic performance, policies, and prospects,” “public finances,” and “external finances”). For illustration, roughly one-third of the 120 sovereigns covered by Fitch were rated at the SRM output, meaning that qualitative factors outweigh quantitative scores in most cases. Furthermore, the SRM includes the WGI as one of its coefficients—indeed, as the variable with the highest weight—by taking a simple average of a sovereign’s percentile rank across the six indicators (Rule of Law, Control of Corruption, Government Effectiveness, Voice and Accountability, Regulatory Quality, and Political Stability and Absence of Violence).

Qualitative factors also play a major role in determining the “country ceiling,” which is the highest rating that any non-sovereign issuer can obtain on their debt. The ceiling is set at or “notched up” from the sovereign rating based on a mix of quantitative and qualitative factors. The difference mainly reflects transfer and convertibility (T&C) risk—the risk that private sector actors face capital and/or exchange controls that impede their ability to service foreign currency obligations. Fitch evaluates the constraints and/or incentives faced by the authorities for and against imposing such controls using a separate Country Ceiling Model (CCM), which contains multiple qualitative scores and adjustment factors, including the composite WGI (Fitch 2023). Moody’s Country Ceilings Methodology also uses the sovereign rating as a starting point, although it first determines a “local currency country ceiling” to reflect jurisdictional risk, and then applies T&C notching to arrive at the Foreign Currency ceiling (Moody’s 2020). Qualitative factors have a disproportionate weight in both components, with WGI again featuring prominently as inputs.² S&P, by comparison, employs very broad and vague criteria related to the openness of the foreign exchange regime and degree of interventionism (S&P 2009).

Besides their role in determining the rating, qualitative considerations are also instrumental in formulating the rating *outlook*, which captures the CRAs’ view on whether the near-term rating trend is positive, negative, or stable. For this critical market signal, there are no specific reference criteria or guidelines—only a judgment on whether the balance of risks is tilted to the upside, downside, or neutral. Qualitative considerations play an important role in shaping the analysts’ views about the evolving risk environment, which also feeds into forecasts of quantitative metrics such as gross domestic product (GDP) growth and fiscal balance. After all, the macro-fiscal forecasts depend heavily on anticipated policy settings and political developments, making the exercise as much a qualitative scenario framing as a quantitative modeling exercise.

² For instance, Moody’s uses the WGI Rule of Law and Regulation Quality indicators in its Predictability and Reliability of Institutions and Government Actions sub-factor (Moody’s 2020).

Despite the disproportionate weight of qualitative factors in sovereign risk analysis, there has been limited research into either the validity of the scoring criteria or the rigor with which they are applied. Fitch and to a lesser extent Moody's reliance on the World Bank's Worldwide Governance Indicators (WGIs) is a notable gap in the literature, given their widespread

use (see Box 2). For instance, recent reports (Lysenko 2025; UNDP 2025) highlight the importance of qualitative factors in rating assessments, yet they do not probe the quality or the statistical significance of their underlying surveys and expert assessments.

Box 2. The ubiquity and power of the World Bank Worldwide Governance Indicators (WGI)

The WGIs are a long-standing data product compiled by the World Bank using 35 data sources covering different dimensions of governance grouped into six headline indicators:

- 1. Rule of Law (RL):** The extent to which agents have confidence in and abide by the rules of society, in particular the quality of contract enforcement, property rights, the police, and the courts.
- 2. Control of Corruption (CC):** The extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.
- 3. Government Effectiveness (GE):** The quality of public services and civil service, and the degree of its independence from political pressures, as well as the quality of policy formulation and implementation, and the credibility of commitment to such policies.
- 4. Voice and Accountability (VA):** The extent to which a country's citizens can participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
- 5. Regulatory Quality (RQ):** The ability of the government to formulate and implement sound policies and regulation.
- 6. Political Stability and Absence of Violence (PV):** The likelihood that the government will be destabilized or overthrown by unconstitutional or violent means.

They were not designed for any specific purpose, yet they have been incorporated into a wide range of analytical products and services. The following list provides an overview of some prominent examples, covering the Big Three and smaller CRAs as well as agencies focused specifically on Environmental, Social, and Governance (ESG) scoring.

Sovereign credit ratings and country ceilings:

- **Fitch** uses an average "composite" indicator of the WGIs in its SRM (see Box 1) as well as under the Long-term Institutional Characteristics pillar of the Country Ceiling Model (CCM).
- **Moody's** cites WGIs separately throughout its core criteria, specifically: RQ and GE under the Quality of Legislative and Executive Institutions sub-factor; VA, RL, and CC under the Strength of Civil Society and the Judiciary sub-factor; and VA or PV under the Domestic Political and Geopolitical Risk sub-factor. Moody's also uses the RL and RQ WGIs to determine the scores for the Predictability and Reliability of Institutions and Government Actions sub-factor.
- **S&P Global** does not reference the WGIs in its sovereign criteria or ceilings.
- **DBRS** uses the five-year average of the GE WGI in the Fiscal Management and Policy factor, and the RL and VA WGIs in the Political Environment factor (DBRS 2024). The WGIs therefore feed into two of the six quantitative "building blocks" in its scorecard.
- **Scope Ratings** use the WGIs as inputs into its Sovereign Quantitative Model (SQM), taking the average of RL, CC, GE, VA, and RQ to compile the Governance factor, which has a weight of 12.5% in the model (Scope 2025). It then uses the PV WGI as the only input to calculate the Political Risk factor, which applies to the entire SQM output.

Sovereign ESG scores and ratings (non-exhaustive):

- **FTSE Russell/Beyond Ratings Sovereign Risk Monitor (SRM) Framework:** The WGI is the main input of the Governance Performance theme, which accounts for 20% of the Sovereign Risk Monitor scorecard (LSEG 2021).
- **MSCI ESG Government Ratings** use all the WGI separately under the Governance Risk Exposure and Risk Management Pillars, where they make up approximately 20% and 13% of the scores, respectively (MSCI 2025).
- **RepRisk Country ESG Risk Index (Country RRI)** quantifies business conduct and reputational risk exposure for doing business in a given country, combining the WGI with data on companies' ESG risk incidents (Gratcheva et al. 2020).
- **Morningstar Sustainalytics' Country Risk Ratings** use the WGI under the Institutional Capital factor to assess ESG performance (Gratcheva et al. 2020).

International Monetary Fund (IMF):

- **Sovereign Risk and Debt Sustainability Framework for Market Access Countries (SRDSF-MAC)** uses a simple average of the GE and RQ WGI to derive a Quality of Institutions index, which in turn is used as a variable in the econometric near-term risk assessment model to determine a country's probability of sovereign debt stress (IMF 2021). Lower WGI scores, therefore, translate into lower assessed debt-carrying capacity.
- **Framework for Enhanced Engagement on Governance** references the WGI to complement analysis of macro-critical corruption and governance vulnerabilities to be used in engagements with member countries, although it is explicit in proscribing the use of country ranking according to WGI (IMF 2018a; IMF 2018b).

Private-sector risk models:

- **countryrisk.io**, a platform for financial institutions to develop country and sovereign risk ratings, uses the WGI alongside other variables under the Institutions and Governance scorecard category, which account for roughly 10% of the scorecard (countryrisk.io 2020).
- **The Organisation for Economic Co-operation and Development (OECD) Country Risk Expert group (CRE)**, comprised of major OECD member country export credit agencies (ECAs), uses the WGI as one of the inputs to determine minimum premium rates for export credit insurance. The classification methodology uses the WGI in the Country Risk Assessment Model (CRAM), before applying qualitative adjustments (OECD n.d.).

This points to a more fundamental asymmetry at the heart of sovereign risk analysis: those factors that can be measured numerically tend to be assessed systematically, while those that cannot are often treated in an ad hoc, inconsistent, and sometimes arbitrary fashion. The corollary is that many hard-to-measure, yet credit-related factors are disregarded or unduly discounted, whether they be credit positive (e.g., intact natural capital stock to sustain agricultural output) or credit negative (e.g., unfunded public pension liabilities). This has especially perverse implications for climate change:

whereas flood, drought, and other climate risks can be measured relatively accurately and increasingly in terms of sovereign exposures, the risk-reductive benefits of resilient infrastructure or disaster planning cannot (at least not yet). Rather, investment in climate adaptation or outlays to conserve natural buffers such as watersheds appear purely as fiscal costs. Hence, as climate risks are gradually being integrated into sovereign risk models, vulnerable countries face both a penalty for their exposure and a disincentive to undertake resilience-enhancing interventions.

Improving quantitative measurement and strengthening statistical reporting must remain a priority. Unreliable, erroneous, and missing macro-fiscal data can distort sovereign risk analysis, in part because they put the onus on rating analysts to plug the gaps with their own inferences. Yet the issues surrounding quantitative data reporting are better documented and receive more attention and technical assistance than qualitative considerations.

This paper seeks to remedy the asymmetry between quantitative and qualitative analysis in sovereign risk assessment. Section 2 shines a light on the practice of CRAs, elucidating major issues of accuracy, completeness, consistency, validity, timeliness, and transparency. Section 3 then proposes both practical and ambitious solutions, which are encapsulated in six recommendations:

- 1. Update the rating criteria, qualitative indicators, and scoring guideposts** to be more specific, explicit, clear, and targeted at those qualitative factors that are demonstrably and materially associated with sovereign risk, while at the same time building in sufficient flexibility to take account of credit relevant factors that are not adequately captured elsewhere.
- 2. Update the World Bank's Worldwide Governance Indicators** to remove hidden biases and add inputs that cover resilience measures and public financial management (PFM) considerations, while disclosing all underlying data, metadata, and scoring rationales and improving access.
- 3. Create a new best-in-class qualitative indicator from first principles** and incorporate only those factors and criteria that can be demonstrably associated with sovereigns' willingness and ability to pay.
- 4. Develop performance-based qualitative scoring frameworks** that recognize and reward institutionalized target setting, performance tracking, and implementation as a stand-alone qualitative marker of good fiscal governance and public financial management.

5. Make the qualitative scoring process "radically" transparent to maximize the objectivity of scoring decisions, ensure that scorers are competent and unbiased, and grant full visibility into the review and validation process.

6. Put in place credibility-based qualitative scoring using a super-forecasting approach that crowdsources expertise and makes scoring decisions weighted by credibility.

The last three ideas on the list represent a fundamental departure from how qualitative scoring is currently performed. Fresh ideas and unconventional approaches may be necessary to adapt to the rise of artificial intelligence, the proliferation and weaponization of misinformation, and the escalating complexity and uncertainty of the risk environment due to climate change and geoeconomic fragmentation. CRAs will likely need to update their models and processes to stay relevant and preserve their legitimacy in this rapidly evolving information landscape.

³ Data transparency is formally part of the rating assessment; both Moody's and S&P include specific criteria about the quality of statistical data and reporting (S&P 2017; Moody's 2022; Moody's 2024). Initiatives on fiscal transparency include the ongoing rollout and upgrading of the IMF's Special Data Dissemination Standard (SDDS) and a recent push by the World Bank for "radical debt transparency" (World Bank 2025b), as well as those by private-sector bodies such as the Institute of International Finance's (IIF) Voluntary Principles for Debt Transparency (IIF 2019) and the OECD's Debt Transparency Initiative (OECD 2022). Although these efforts touch on qualitative data, they are generally subordinated to getting the hard data right.



Section 2

**Quality and
transparency gaps
in qualitative
rating factors**

Section 2

Quality and transparency gaps in qualitative rating factors

Data quality can be defined in several ways depending on context and usage. For the purposes of this analysis, the following six criteria are considered:

- 1. Accuracy:** How well do the qualitative factors explain and predict sovereign default risk?
- 2. Completeness:** Do the factors capture all the relevant dimensions of sovereign risk, and exclude those that have no bearing on ability or willingness to pay?
- 3. Consistency:** Are the factors compiled and applied in the same manner across all countries?
- 4. Validity:** Are the factors independently verified against predefined rules, references, or requirements?
- 5. Timeliness:** Do the factors reflect values as close to the time of occurrence as possible, avoiding excessive lags?
- 6. Transparency:** Are the methodologies for compiling the factors clearly defined, reproducible, and visible?

Data quality differs fundamentally between the qualitative and quantitative metrics used in sovereign risk analysis. Whereas quantitative metrics tend to have tangible properties that can be measured precisely on numerical scales (e.g., tons of goods shipped abroad, prices recorded in shops, sums expended on public-sector wages and pensions), most governance concepts are descriptive in nature and can only be categorized (e.g., “country A can be assessed as high/medium/low political risk”) or rank ordered (e.g., “country B is politically riskier than country C”). Furthermore, quantitative metrics can be evaluated against relatively objective benchmarks (e.g., public debt destabilizing primary fiscal balance), whereas qualitative concepts such as “policy effectiveness” and “political transparency” are ambiguous and sometimes contested. Qualitative statements such as “country B is three times politically riskier than C” or “country D is politically stable” are not definitive. Even so, qualitative assessments can still approximate reality if the methodologies and criteria are valid and their application is consistent, timely, and transparent.

Even with all these conditions met, decisions on how to score or rank order countries are inherently subjective. Analysts have extensive training and experience parsing credit relevant signals from political noise. Nevertheless, they are susceptible to bias, human error, and conflicts of interest. Validation by the credit committee process and issuer review provides some quality control. Still, the number of individuals involved in the decision is relatively small (typically two analysts and around five committee members), and distortions can creep in via groupthink, procedural mistakes, or lax oversight.

Appeals to external governance indices and indicators such as the WGIs can mitigate this risk. They benefit from stronger intersubjectivity via the “law of large numbers” by aggregating perceptions and opinions from surveys and expert assessments. Yet they suffer from their own set of conceptual and methodological issues, while also confronting broader challenges for survey research in the digital age such as declining response rates, sampling bias, and misinformation.

The remainder of this section examines attempts to answer each question in turn.

2.1 Accuracy: How well do the qualitative factors explain and predict sovereign default risk?

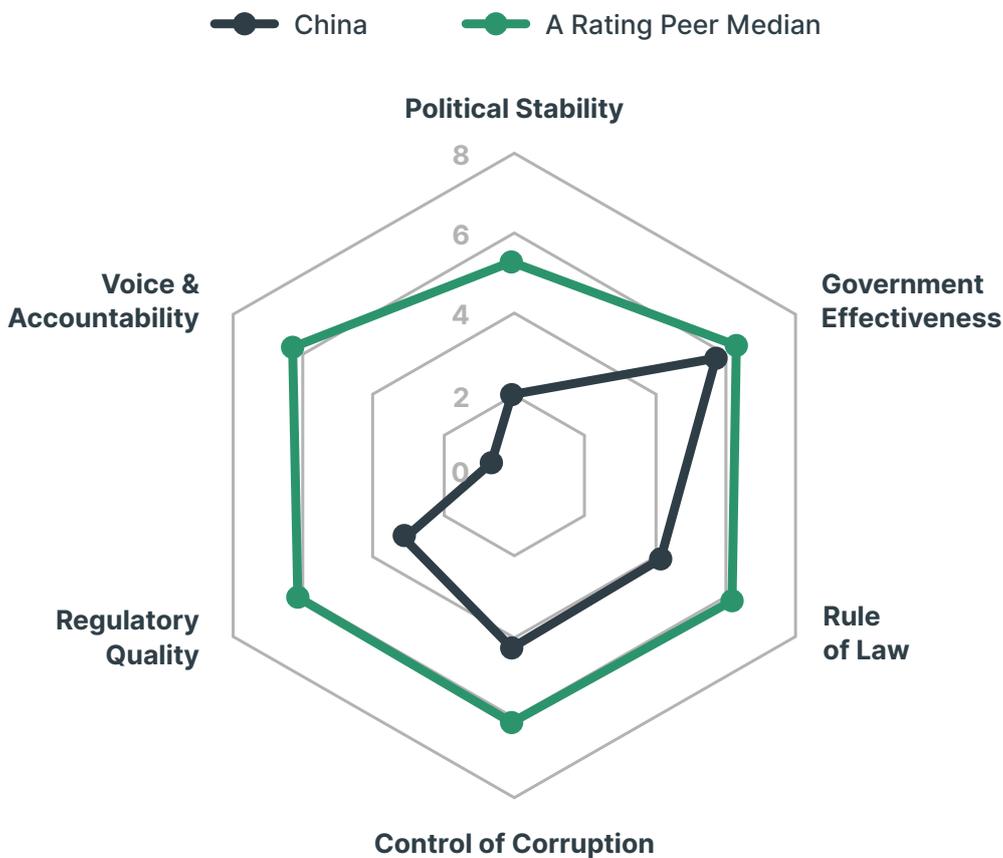
Beyond the intrinsic statistical constraints on qualitative data (i.e., that they can only be measured on nominal and ordinal measurement scales), accuracy is a function of model specification— whether the model incorporates those factors that are closely associated with default risk, while excluding or underweighting the spurious ones. Correct specification is complicated by the lack of a uniform and universally accepted definition of governance as it relates to sovereign risk.⁴ Statistical significance can be hard to establish with “fuzzy” concepts such as institutional quality. Hence, the evidence base for the inclusion of certain qualitative explanatory variables is mixed (Beaulieu et al. 2012; Biglaiser and Staats 2012; Bodea and Hicks 2018; Barta and Johnston 2023).

The empirical evidence does not conclusively support the “democratic advantage” thesis. Recent studies rather show that the relationship is at best conditional on the countries’ level of financial vulnerability and adjustment needs (Hansen 2023). There are also valid theoretical arguments to suggest that democracies are more prone to lapses in fiscal discipline because of political business cycles, common pool problems, and political paralysis caused by veto players (Abbas et al. 2021). Setting aside normative debates about political systems, the implication is that CRAs do not have an *a priori* justification to mechanically penalize sovereigns based on their electoral systems, press freedom, human rights, or other markers of liberal democracy.

A deeper analysis of the “democratic bias” in sovereign ratings is beyond the scope of this paper,

but the impact on sovereign ratings can be illustrated using Fitch’s rating of China. According to Fitch’s SRM, China’s WGI score is the equivalent of 2.5 notches below the A-rated peer median. In other words, China’s percentile rank in the composite WGI drags its model-indicated rating down by 2.5 notches relative to where it should be at an A rating. The reason for can be traced to the Voice and Accountability and the Political Stability and Absence of Violence indicators of the WGI (see Figure 2). Within these pillars, several component surveys focus specifically on liberal democratic principles, including the University of North Carolina’s Political Terror Scale (PTS), Human Rights Measurement Initiative (HRMI), and Freedom House indicators of political rights and civil liberties. To the extent that they reflect normative rather than strictly explanatory factors, the effect is to skew China’s rating to the downside.

Figure 2
China’s ranking in the 2024 WGIs (notch contribution to Fitch’s SRM)



Source: Author’s analysis based on World Bank (2025d).

The direction of causality is also debatable for certain governance factors. Whereas events such as large-scale inter-state war unequivocally increase default risk, others are more ambiguous and can be coded as either credit positive or negative depending on context or the political leaning of the scorer. Fiscal discipline is generally considered a virtue by CRA analysts, but it can be misguided when fiscal policy becomes pro-cyclical during downturns, retards recoveries, or discourages vital productivity- and resilience-enhancing investments that weaken fiscal outcomes in the short run but are credit positive over the longer term.

Miscalibration can also undermine accuracy. Extreme heterogeneity within and between countries means that applying equal weights to factors for all rated sovereigns may not be appropriate. The measures sometimes lack the granularity to distinguish between risks that are material at the national level, and those that are localized. The thresholds for when events become material are also hard to ascertain and require a heavy dose of subjective judgment. For example, political violence and social conflict may be correlated with sovereign risk, but their bearing on creditworthiness can differ depending on whether they occur close to centers of economic and political power, or in a peripheral part of the country.

For instance, the impact of political violence on credit drivers is higher if it occurs in the country of Papua New Guinea (PNG) than in the province of West Papua, Indonesia. Both territories have struggled with armed separatist movements—secessionists in Bougainville in eastern PNG, and the Free Papua Movement (Organisasi Papua Merdeka, OPM) in West Papua. Both PNG and Indonesia are in the 30th percentile of the composite WGI, dragged down by low Political Stability and Absence of Violence scores. Yet Indonesia has a land mass 4 times larger than PNG, a population 26 times larger, and an economy 46 times the size of PNG's. The topic of West Papua has not been mentioned in Indonesia's rating action commentaries, whereas Bougainville's status was cited as a distinct risk in PNG's most recent S&P review (S&P 2025). Assigning the same political risk scores and weights to PNG and Indonesia could be overstating the risks to political stability in Indonesia.

In a similar vein, institutional quality assessments are often too sweeping in treating the government as a monolith, masking significant variation between ministries and levels of administration that can drag down or lift up the sovereign as a whole. The CRAs generally do not use a sum-of-parts approach to institutional analysis. They may argue this would be impractical for all their rated sovereigns, or that they view the sovereign as more than the sum of its parts. However, there may be instances when such a disaggregated approach is appropriate. Indeed, the CRAs do this for the central bank because of the importance of monetary policy institutions for ensuring price and financial stability, which are key fiscal risks. Moody's Institutions and Governance Strength pillar has separate criteria for fiscal and monetary policy effectiveness. A finer-tuned analysis may be warranted for other important agencies, for example the energy ministry in an oil-rich nation, in order to correctly weigh policy effectiveness in credit relevant sectors.

Institutional quality can also vary between levels of government. Applying a single governance score to the "general government"—comprising central, state, and local governments—often does not capture the variegated landscape of sub-national fiscal institutions. High policy effectiveness may not cascade down to lower levels of administration, as evidence from disappointing fiscal decentralization drives shows (Nakatani 2025). Weak sub-national fiscal risks can drag down the sovereign, as exemplified by the string of debt crises caused by insolvencies of Brazilian states and municipalities during the 1980s and 1990s (Bornhorst 2019). Yet the opposite case, where the general government score is improved, is far rarer, even in instances where large local and regional governments have a stronger credit rating. This is the case for the Flemish province of Belgium (AA- by Fitch versus the sovereign rating of A) and Buenos Aires, Argentina (B- by Fitch versus the sovereign rating of CCC+) (Fitch 2025b). Again, a more granular scoring approach may be needed to recognize credit uplift and not just contingent liabilities from the sub-national levels.

The logic also applies to the broader public sector, given that state-owned enterprises (SOEs) and national development banks (NDBs) are often deeply entangled with the sovereign (e.g., as large recipients of public subsidies or major sources of royalties). Fitch states that SOEs are excluded because they obscure fiscal analysis (Fitch 2025a), yet they are also important arms of the policy apparatus, serving as vehicles to implement programs and projects, and therefore vital in considerations of government effectiveness.

Here, too, an asymmetry arises: weak corporate governance at these institutions poses an off-balance-sheet contingent liability for the sovereign, which is often cited as a downside risk or cause for a negative adjustment. For example, Fitch subtracts one notch from Mexico's quantitative SRM-indicated rating on account of contingent liabilities related to the national oil company Pemex, to which it assigns a stand-alone credit rating of ccc- (Fitch 2025c). Yet the inverse scenario, where a well-administered credit relevant SOE lifts up the sovereign, is exceedingly rare. Brazil's NOC Petrobras has a standalone credit rating of bbb—three notches above the sovereign rating and two notches above the country ceiling—due to expectations of continued strong operational and financial performance (Fitch 2025d). Although both NOCs have extensive linkages to their respective sovereigns, Pemex is treated as a distinct rating weakness for Mexico's sovereign, while Petrobras's strong governance and material financial performance—the company contributes roughly 1–2% of government revenues (Fitch 2022)—are not recognized as a commensurate rating asset for Brazil.

2.2 Completeness: Do the factors capture all the relevant dimensions of sovereign risk?

Whereas accuracy requires that sovereign risk analysis exclude spurious and biased qualitative factors, completeness demands that all credit relevant qualitative factors be included. However, the broad scope of sovereign risk analysis means that it could theoretically encompass all aspects of political economy, which is impractical and unparsimonious. The CRA criteria therefore tend to be purposely vague, affording the analysts the freedom to use their expert judgment about which qualitative considerations are most salient for a given credit assessment.

However, there are important qualitative considerations that are either omitted or underweighted in CRA assessments, which can unduly discount credit strengths or exaggerate weaknesses. This may occur for several reasons: errors of oversight, lack of supporting theory or evidence, or because of practical challenges with consistent measurement and evaluation. Failing to consider concepts holistically can also distort the analysis, especially when it concerns assessments about the “quality” of policymaking. For instance, CRAs often make sweeping statements about the “credibility,” “prudence,” or “soundness” of policy frameworks without referencing specific markers or methodologies for how these are evaluated. This has the added downside of depriving issuers of meaningful or actionable insights about how practically to improve these qualitative measures, and by extension their ratings.

The quality of macro-fiscal policy and of public financial management (PFM) frameworks is one example where CRAs' analysis appears incomplete. Although the Big Three reference them in their criteria, underscoring the importance of “robust plans” (Moody's), “proactive” policymaking (S&P), and the “credibility of the fiscal framework” (Fitch), they provide few details about what those qualifiers mean specifically. Given the centrality of macro-fiscal planning and PFM to willingness and ability to pay, a fuller diagnosis should cover topics such as the quality of macro-modeling and forecasting, the effectiveness of disaster planning, the efficiency of treasury management, inter-agency coordination, and investor relations, *inter alia*.

Many of these aspects are already assessed by other organizations including the Country Policy and Institutional Assessments (CPIA) of the World Bank, the African Development Bank (AfDB), and the Asian Development Bank (ADB); the World Bank's Debt Management Performance Assessment (DeMPA); the Institute of International Finance's (IIF) Investor Relations and Debt Transparency Report; and the Public Expenditure and Financial Accountability (PEFA) assessment, which is a joint initiative by the World Bank, the IMF, the European Commission, and several governments. All four provide comprehensive and granular scores on PFM and macro-fiscal frameworks. For instance, the CPIA Debt Policy and Management criterion assigns scores from 1 to 6 based on whether the debt management strategy is conducive to minimizing budgetary risks and ensuring long-term debt sustainability (World Bank 2025b).

However, none of these expert assessments feature explicitly in the CRA criteria. They are included in the WGIs, but not in their entirety (e.g., the Economic Management cluster that deals with monetary and exchange rate policy, fiscal policy, and debt policy is not part of the WGI construction). The CPIA scores also suffer from their own shortcomings (discussed in Section 2.6), including limited country coverage.

Gaps in criteria can arise due to changes in the risk environment. Fiscal governance is not a static concept; it evolves along with shifts in the social and cultural setting, structural transformations of the economy, technological innovations, and political events. The CRAs do periodically update their rating criteria to account for these changes, yet they are often slow in doing so, and they have a record of missing certain major developments (e.g., the degree of sovereigns' exposure to systemic banking sector risks ahead of the Global Financial Crisis).

This is also true for governance indicators such as the WGIs and their constituent surveys. A delay may be deliberate and justified to avoid creating unnecessary breaks in the historical series while the macro-criticality of emerging risks is being assessed. However, an excessive lag can become distortionary when it obscures major accumulations of risk, or if the eventual update fails to adequately capture the salient dimensions. For example, the belated recognition of climate risks by CRAs has failed to account for nature risks in their recent updates (Fitch 2025a), even though these can be equally material (Cevik and Tovar Jalles 2020; Prodani et al. 2023; Agarwala et al. 2022; Ceglár et al. 2024; Kraemer et al. 2015).

The governance-related aspects of climate and nature risks are still a major blind spot for CRAs

and governance indicators. Although the quantitative dimensions of climate risk are being integrated into macroeconomic and credit analysis (Mitra et al. 2025; Fitch 2025e), the qualitative aspects such as food insecurity and population displacement are omitted or not treated systemically.

This goes beyond generic political risk. For instance, policies that boost agricultural output via intensive export-oriented monoculture in the short run but erode resilience to climate shocks over the medium term may be coded as credit positive due to the GDP impact, even though they may undermine food security over the medium term. Similarly, the management of climate risks, including proactive disaster planning and resilience interventions, do not receive comprehensive treatment in the CRA governance scores or WGIs despite their increasing importance to policy effectiveness (see Box 3).

The situation for nature risks is analogous but made more complicated by the fact that these risks are harder to measure and less widely recognized among financial market actors, despite growing evidence of their financial materiality to sovereign creditworthiness (Kulenkampff 2025). Unlike climate, where mitigation and adaptation have widely accepted quantitative measures (e.g., CO₂ equivalent emissions; economic losses from catastrophic events), nature loss or biodiversity do not have common denominators across all countries and biomes. Initiatives to establish biodiversity baselines and put "nature on the [sovereign] balance sheet" are advancing,⁵ slowly. In the meantime, nature is mainly a qualitative factor, as are considerations around the stewardship of macro-critical natural capital resources.

⁵ See, for example, Capitals Coalition (n.d.).

Box 3. Resilience as a qualitative factor

Many aspects of resilience cannot be measured with a sufficient degree of precision or consistency to incorporate them quantitatively into credit rating analysis, either because the measurement instruments are not (yet) available or up to standard, or because they are intrinsically qualitative. “Resilience assets” such as healthy soil quality and “resilience interventions” such as flood-resistant building codes are too complex, contested, or subjective to lend themselves to uniform definition or standardized measurement. They are examples of credit positive qualitative factors that are material to sovereign creditworthiness (Kulenkampff 2025), yet not adequately reflected in sovereign credit ratings.

The asymmetry arises from the tendency to analyze concepts that can be measured quantitatively in a relatively methodical and objective way, whereas those that cannot are often omitted or treated in an unsystematic, ad hoc, and subjective manner. Physical risks, and to a lesser extent, transition risks, have relatively precise measurement instruments and agreed-upon metrics, such as expected economic losses from natural catastrophes or stranded assets. For instance, the potential hit to agricultural output from floods and droughts can be estimated relatively accurately based on historical loss experience and projected trends under different climate scenarios. By contrast, the potential avoidance of losses due to adaptation investments in climate resilient supply chains or climate smart agricultural practices such as agroforestry is harder to measure. This is because data are often too spotty and low grade, and because the resilience gains tend to crystallize over the medium term. In practice, that means the credit uplift from “soft” measures of resilience is unduly discounted relative to the “hard” data underpinning climate and nature risk. To the extent that this quantitative bias further skews ratings to the downside, it can contribute to the mispricing of sovereign debt.

The asymmetry manifests in multiple qualitative scores, adjustments, and thresholds embedded within the Big Three CRA methodologies. The criteria include language around the capacity of political institutions and the policymaking apparatus to absorb shocks and navigate structural changes, including to physical and transition risks related to climate change. The agencies’ scorecards also allow for adjustments to multiple factors to account for climate and nature risks; for instance, Moody’s highlights how “physical climate risk, water risk or risks related to natural capital can all impair a country’s ability to rely on its natural resources for exports and foreign-currency generation” (Moody’s 2022). Fitch recently introduced a Climate Vulnerability Signals (Climate.VS) assessment to determine whether climate-related risks warrant further scrutiny and discussion in credit committees. The signals are computed from catastrophic risk data and adjusted using “qualitative filters,” which consider only “unmitigated exposures” that “do not incorporate sovereigns’ policy targets or general mitigation and adaptation capacity” (Fitch 2025e).

Even if references to resilience in the Big Three CRA methodologies are cursory, the qualitative criteria are sufficiently elastic to afford the analysts considerable discretion over whether and how much to recognize resilience assets and policies. However, in the absence of clear benchmarks and independent validation, the treatment of resilience can vary considerably between rating reviews and across rated sovereigns. For instance, two small island nations may be equally vulnerable to hurricane perils; Country A is conserving its mangroves and enforcing building codes, while Country B is not. Without hard data to quantify the mitigation effects or prescribed guidelines for how to assess them qualitatively, recognition of Country A’s efforts is ultimately a judgment call by the credit committee.

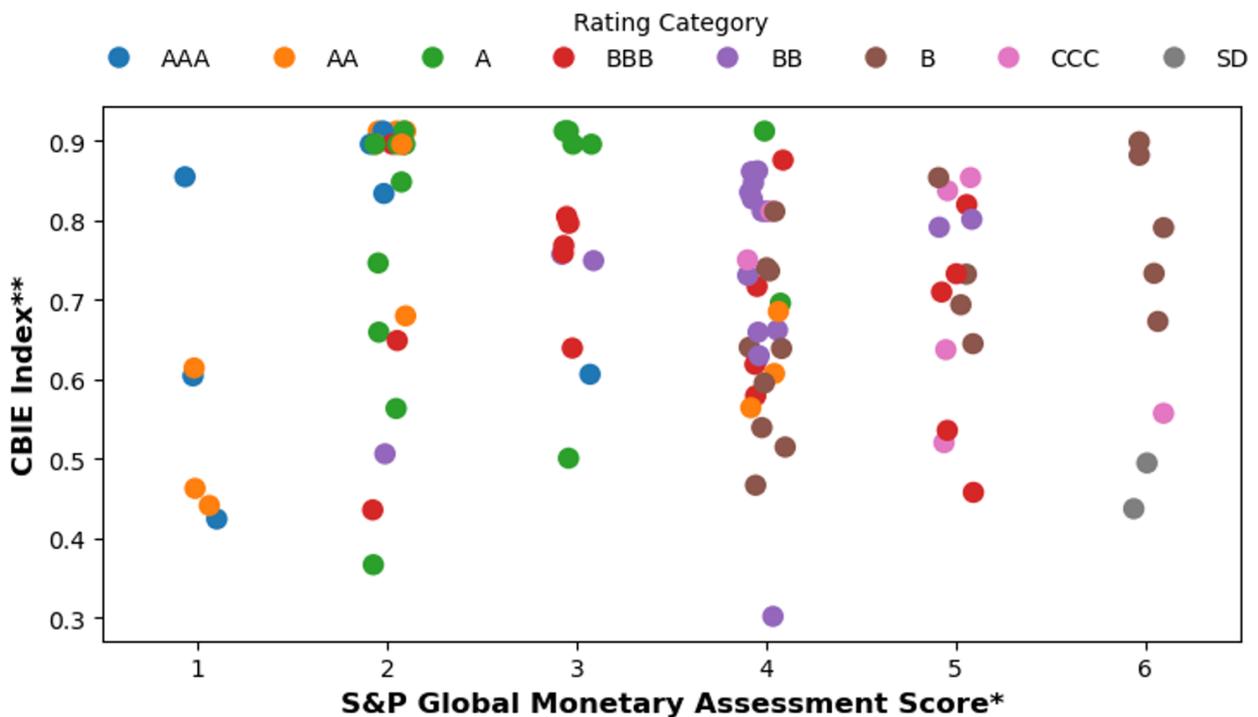
Appeal to the WGI does not correct this shortcoming, since the constituent opinion surveys and expert assessments do not adequately reflect resilience assets or policies. For instance, the WGI Government Effectiveness pillar, which arguably ought to include the ability to absorb shocks and implement resilience plans, does not include any surveys or scores that capture this dimension. It does draw on various parts of the Country Policy and Institutional Assessments (CPIA) performed by the World Bank, the African Development Bank, and the Asian Development Bank. However, the WGIs omit those scores that pertain to Environmental Sustainability and Fiscal Policy, which would capture important dimensions such as “environmental policies and institutions foster the protection and sustainable use of natural resources,” or the “quality” of the fiscal response to natural disaster shocks (World Bank 2021).

2.3 Consistency: Are the factors compiled and applied in the same manner across all countries?

Cross-country comparability is central to sovereign risk analysis. Sovereigns are evaluated on a relative basis precisely because the high qualitative content renders absolute measures of credit risk difficult if not impossible. To be comparable, the rating inputs for all countries must follow as close to a standardized definition and calculation as possible. This is difficult when the concepts are ambiguous, rating criteria are elastic, and the credit analysts have broad latitude to apply their judgment.

Inconsistency can be observed in the high variability of qualitative scores across space and time. The S&P's Monetary Assessment score, for example, encapsulates the effectiveness and credibility of monetary policy as a function of the exchange rate regime, the independence of monetary authorities, the range of monetary instruments at their disposal, the ability to act as lender of last resort for the financial system, and the development of the local financial system. The difficulty of fitting the multiplicity of policy mandates, institutional arrangements, and instruments in operation today to these broad criteria can be appreciated from the high dispersion of scores when plotted against an index of Central Bank Independence developed by Davide Romelli (2022) (see Figure 3).

Figure 3
S&P Global 'Monetary Assessment' Score* vs. CBIE Index** (2023)



*S&P Scores: 1=best, 6=worst

**CBIE Index: Central Bank Independence – Extended index is a measure of central bank independence developed by Davide Romelli (2022). See: <https://cbidata.org/>.

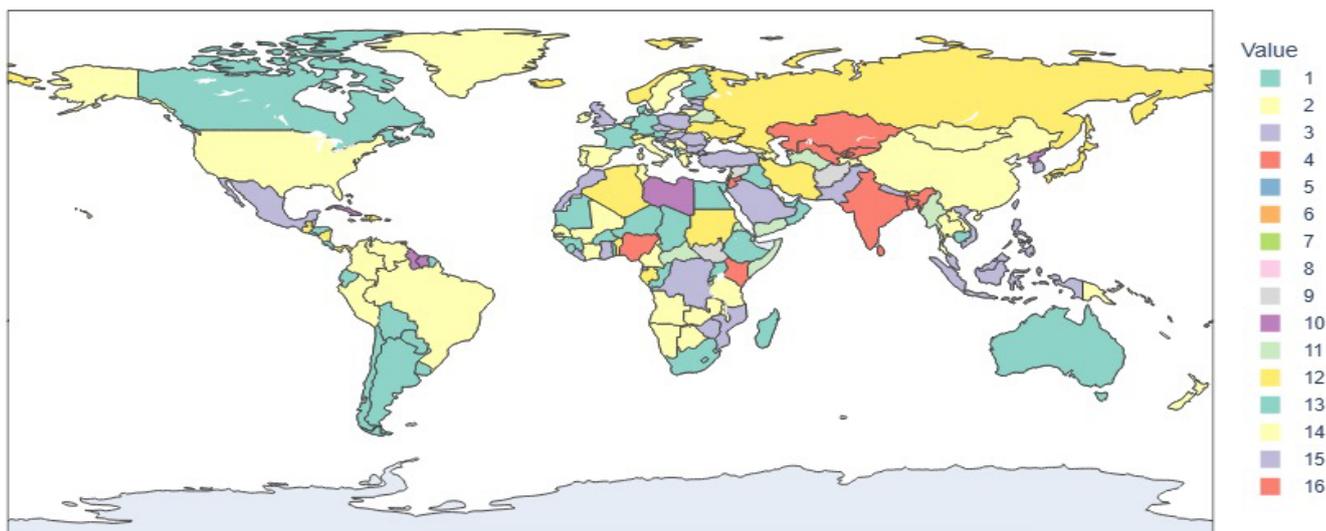
Source: S&P Global latest sovereign ratings, Romelli (2022).

Consistency over time is especially important since long track records are a crucial part of the Big Three’s competitive advantage. Inconsistencies arise when qualitative assessments and scores change frequently between reviews. This may be either because of unforeseen shifts in the risk environment that require frequent updates to the scores or because the quantitative metrics exhibit excessive volatility and the qualitative factors are used as adjustment variables to stabilize the end ratings. In the latter case, CRAs face accusations of arbitrarily “shifting the goalposts” if the qualitative scores do not actually track key credit drivers.

The WGIs face the additional challenge of ensuring consistency among the underlying sources. This is complicated by the fact that the surveys and expert assessments have different country coverages

(see Figure 4), ranging from 16 countries (Political and Economic Risk Consultancy Corruption in Asia Survey) to 182 (US Department of State Trafficking in Persons Report). None of the 214 countries and territories in the WGI set are covered by more than half of the total sources, with some smaller territories (e.g., Bermuda, Cayman Islands, French Guiana) relying on only 1–2 sources for all the WGIs (typically Crisi24 Country Security Assessment and S&P Global Country Risk Service). Kaufman and Kraay (2024) note that this does not pose conceptual problems since for each country the scores are the “best possible estimate of governance given the available data for that country.” However, the analysis shared in this report suggests that comparable quality cannot be safely assumed between sources.

Figure 4
2024 WGI country coverage



Source: World Bank (2025d).

2.4 Validity: Are the factors independently verified against predefined rules, references, or requirements?

A closed-door rating process coupled with the lack of objective benchmarks for qualitative metrics complicates the validation of scores. Several checks and balances are built into the rating and review processes, yet they provide imperfect quality control.

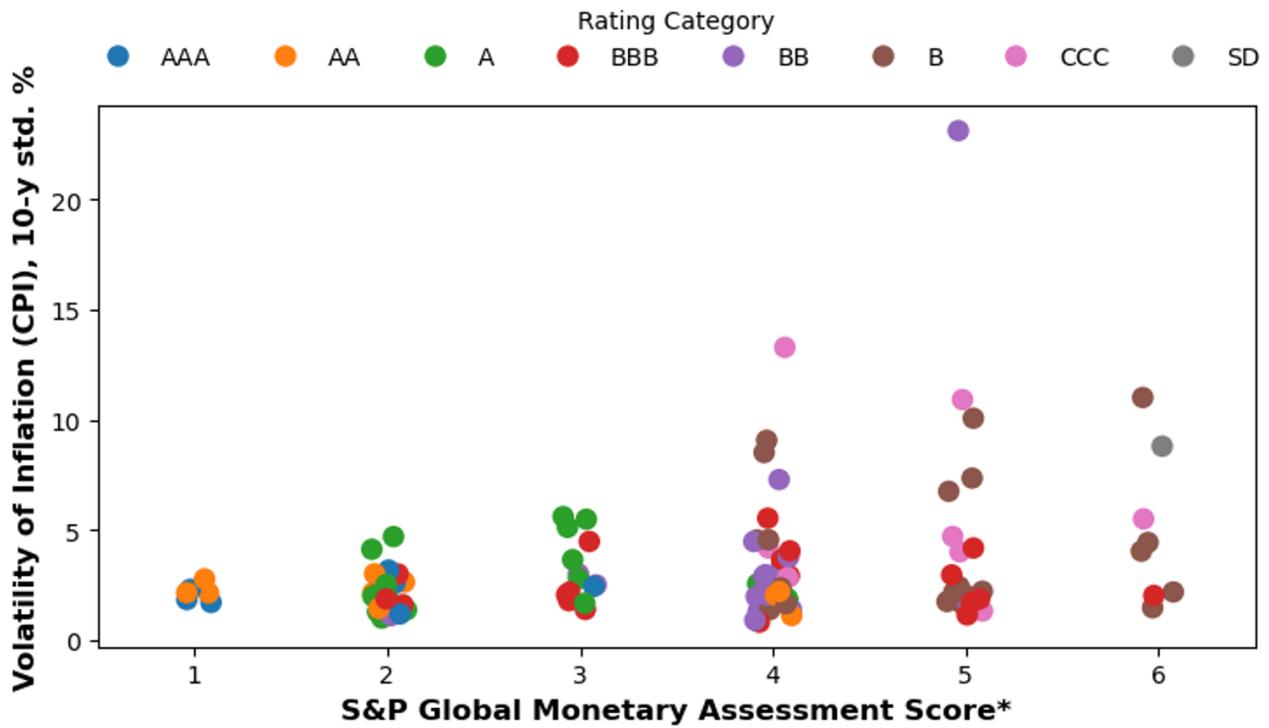
First, a diverse committee of other sovereign credit analysts interrogate the lead analyst’s rating recommendations. These deliberations are not public, so the thoroughness and consistency of

the scrutiny cannot be verified. The panel may be susceptible to groupthink given that it only includes members from within the organization who tend to share similar experiences, values, and mental models. Second, after the committee has reached its decision, the issuer is afforded the opportunity to review and contest the rating decision. However, they typically have a very short time to lodge an appeal, while many sovereigns lack the technical capacity to spot errors or inconsistencies. Third, regulatory supervision of committee outputs and minutes occasionally triggers audits, but these are rare and subject to information asymmetries.

The ambiguity and subjectivity of scoring decisions also makes them difficult to validate with hard evidence. Certain scores with clear policy targets can be assessed in terms of corresponding quantitative metrics. For instance, monetary policy

effectiveness can be measured in terms of long-run price stability. However, plotting S&P's Monetary Assessment scores against the volatility of consumer price index (CPI) inflation (see Figure 5) does not show a strong correlation.

Figure 5
S&P Global Monetary Assessment scores* vs. CPI inflation volatility (2015–2025f avg.)



*S&P Scores: 1=best, 6=worst

Source: S&P Global latest sovereign ratings, IMF World Economic Outlook (IMF 2025).

External governance indicators such as the WGI are meant to provide another independent benchmark for CRA assessments. However, as has been highlighted throughout this report, the constituent surveys and expert assessments suffer from their own validation issues. Problems of declining response rates, sampling bias, and respondents' reticence to provide candid responses to politically sensitive questions can undermine the validity of survey data in the WGIs. Political interference and "gaming" are also concerns in light of the scandal surrounding the World Bank's Ease of Doing Business indicators. And because many of the expert assessments, data, and methodologies are not publicly accessible—e.g., because they are behind a pay wall, not publicly disclosed, etc.—it is often impossible to perform independent validation checks. The University of North Carolina's Political Terror Scale, mentioned in

Section 2.1, illustrates this point. It measures "violations of physical integrity rights carried out by states or their agents" and is coded on a scale of 1–5 based on information from reports published by Amnesty International, Human Rights Watch, and the US Department of State (Haschke 2022). For example, PTS assigned Brazil a score of 4 in its recent release, which it applies to countries where "[c]ivil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life...on this level terror affects those who interest themselves in politics or ideas" (PTS 2025). Other countries in this category include Belarus, Burundi, Central African Republic, Iraq, Kazakhstan, and Mozambique. This scoring decision is questionable. Although much of the effect on the rating washes out in the aggregation of multiple sources, it can still have a bearing on the score.

2.5 Timeliness: Do the governance scores reflect current realities?

Timely information is essential for sovereign risk analysis to be meaningful and actionable. Credit relevant events must be reflected rapidly and accurately in ratings if they constitute a change in credit fundamentals. Long-term foreign currency ratings—the benchmark rating used by investors—are meant to look “through the cycle” and not move with temporary cyclical forces. This gives added weight to structural factors such as institutions, which are “sticky” by nature as social-political transformations and policy reforms take time to unfold. That means that governance scores should be stable over time, requiring fewer updates than for cyclical rating factors such as growth and inflation.

Institutional inertia is also the logic used to justify the lag in the WGIs, which for some of the components is considerable. The WGIs are released annually, typically in September, with the latest available data from the underlying surveys. Several of these are published on a 2-to-3-year cycle (e.g., Afrobarometer, Bertelsmann Transformation Index, European Quality of Government Survey), and sometimes after the WGI cutoff date. But given that they mean to capture structural features and slow-moving governance trends, this is not considered a problem. As the architects of the WGIs themselves note: “the WGI are not designed to be a tool to evaluate specific governance reforms in individual economies—for this purpose, the WGI should be supplemented with more granular country-specific data and analysis” (Kaufman and Kraay 2024).

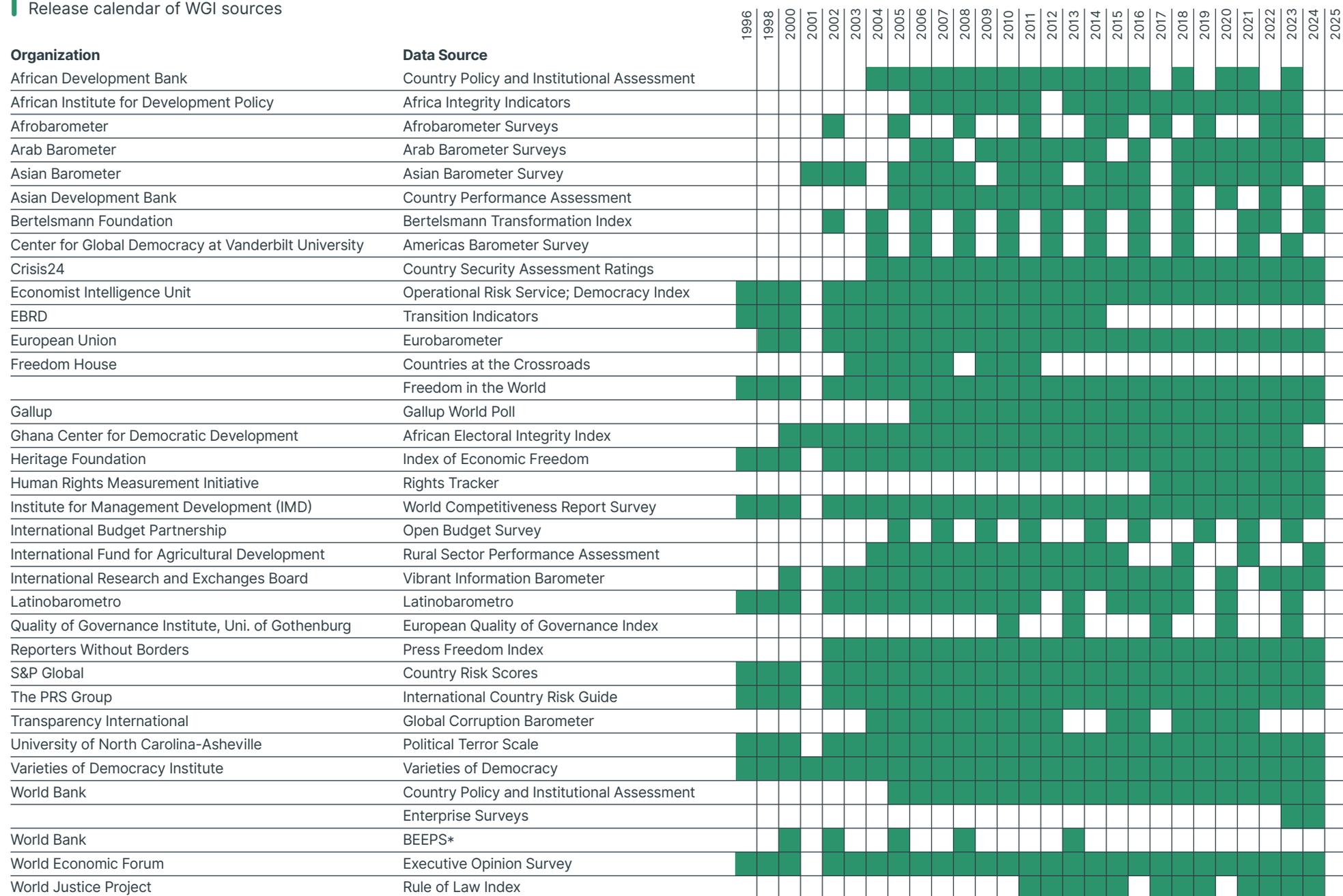
However, a closer look at some WGI components reveals a significant blurring of lines between structural and cyclical factors. For instance, the Bertelsmann Transformation Index (BTI) is published biannually towards the beginning of the year.

The current edition was published in March 2024 with data collected during 2021–2023. This will remain in the WGIs until the next update in late 2026—a lag of more than three years. In Brazil’s case, much of the commentary accompanying the BTI 2024 Brazil Country Report makes references to the prior administration’s policy stances, rather than strictly structural considerations. Based on their assessment of those policies and the prevailing political climate at the time, the BTI assigns Brazil low scores for key sub-factors under the Government Effectiveness pillar, including a score of 1 out of 10 for “conflict management”—below countries such as Haiti (3), Libya (2), and Somalia (2) (Bertelsmann Stiftung 2024). While the decision to score Brazil lower than several failed states testifies to a problem of accuracy and validation in this instance, the point here is that the structural nature of governance scores cannot be taken for granted. And where the scores are actually tracking cyclical trends, extra caution must be applied to ensure at a minimum they reflect prevailing realities. Thus, despite ongoing challenges of political polarization in Brazil, conflict was effectively managed during the transfer of power between administrations, making BTI’s coding of this sub-factor appear both outdated and misguided.

Reflecting current conditions is better, but projecting future conditions is best. That is because credit ratings are *forward-looking* measures of creditworthiness, and indicators with exceedingly long lags can create disconnect between the readings in governance indicators, realities on the ground, and realistic projections of conditions over the rating horizon. This is especially problematic in situations where reform-minded leaders come into power after a crisis with promising momentum to address institutional weaknesses in the credit profile. An excessive lag in the governance indicators can in effect burden the country with odious “institutional debt” in the form of stale scores that impose an undue drag on risk perceptions and sovereign ratings.

Figure 6

Release calendar of WGI sources



Source: World Bank (2025d).

2.6 Transparency: Are the methodologies for compiling the factors clearly defined, reproducible, and open to scrutiny?

Transparency encapsulates many of the data quality issues covered above, as the shortcomings in many respects reflect the lack of clarity about how qualitative data are compiled, scoring decisions are made, and outcomes are communicated. Transparency is therefore more about visibility into the process of generating and disseminating qualitative data than the substance of the information *per se*.

The criteria and methods used to create qualitative data are generally well documented and publicly available. There was a strong push in the aftermath of the Global Financial Crisis for greater regulatory disclosure and reporting by CRAs (e.g., see the European Union's CRA3 Regulation of 2013),⁷ including requiring all licensed CRAs to publish their methodologies and rating action commentaries. These commentaries are crucial for understanding the rationale behind rating decisions. Yet like the criteria, they are light on detail and drafted in boilerplate, vague language without benchmarks such as peer scores or category medians for reference. More detail is provided in longer credit reports, but these are generally behind a paywall. The deliberations of the credit committee, where scoring decisions are scrutinized by peers from different parts of the CRA organization, are not public, and even regulators are only privy to summarized committee minutes. Consequently, it is difficult to ascertain from either public or published commentary why certain scores were assigned.

The transparency of the WGs is more mixed. Although the World Bank provides extensive methodological documentation and commentary about how the indicators are constituted as well as the underlying input data and metadata, access to the original data varies greatly by source. Survey methodologies, codebooks, and raw data are generally available, but scoring rationales are not. In some cases this may be deliberate since anonymity protects respondents against reprisals by clients or the governments they are reviewing, especially for the more sensitive topics such as corruption and human rights. Yet given how consequential their decisions are for

sovereign ratings, further disclosure of the credentials, competence, and potential conflict of interests of the scorers may be warranted on top of more justification for scoring decisions and source selection.

Transparency concerns also extend to the choice of “guideposts.” For example, the researchers at the University of North Carolina who compile the PTS must adhere to the US State Department's Country Reports to determine their scoring. Arguably, this is not a sufficiently impartial source to render objective verdicts about the United States' geopolitical rivals.

Transparency is also a function of accessibility. The commercial nature of credit ratings means that the CRAs keep much of the key information behind paywalls. This restricted content includes not just the detailed credit opinions and data, but also the inputs to the scorecard (Moody's) and SRM (Fitch). Only S&P provides all relevant scorecard information in their public rating action commentaries, including the qualitative scores and quantitative data inputs, albeit with limited details on rationales. The cost of a subscription for paid content is not advertised for such enterprise solutions with bespoke pricing, but informal interviews suggest that costs are broadly in line with the standard ranges for financial data aggregators and market intelligence providers, which is in the order of USD10,000–USD30,000. This puts it out of reach for many budget-constrained issuers since the research service is not included in the cost of obtaining a rating.

Accessibility is an issue for the WGs. Of the 35 data sources, four are commercial business information providers—the Economist Intelligence Unit, Political Risk Services, Crisis24, and S&P Global Country Risk Service. For these, the World Bank publishes source data used to compute the indicators along with metadata, yet neither the methodologies nor the analysis is publicly available. This opacity is problematic given their considerable weight in the Government Effectiveness and Political Stability indicators—45% and 64% of Brazil's scores for these WGs, respectively.

⁷ See <https://eur-lex.europa.eu/eli/reg/2013/462/oj/eng>.

Accessibility also refers to the intelligibility of the scores and methodologies, which is crucial for issuers to identify and contest potential errors and omissions. The limited direct access to the models and the input data has created business opportunities for commercial sovereign rating advisors. The emergence of this cottage industry of specialized firms and investment banks testifies to the challenges of issuers in understanding sovereign rating dynamics and model mechanics, which can be attributed in part to the problems outlined above. Advisors help issuers demystify their ratings—often by reverse engineering the rating models—and develop CRA relations and credit positioning strategies with the aim of securing upgrades. These services can be expensive, which adds to the cost of obtaining rating for an issuance, although recently the creation of public interest platforms such as the African Peer Review Mechanism (APRM) have helped to defray costs.

Again, intelligibility is a problem for the WGI too. Although the World Bank has made significant strides in presenting the data and analysis behind the WGI in digestible and user-friendly formats on their website, that does not extend to the underlying surveys and expert assessments. Drilling down into the drivers of individual scores—for the data that are not proprietary or paywalled—requires analytical resources that few issuers have at their disposal. Those who do have the wherewithal to probe their WGI can dedicate resources towards improving the underlying scores, which can take the form of “gaming” when certain attributes are deliberately misrepresented for this purpose. On the flipside, those who lack the resources to interrogate or challenge the scoring decisions are at a disadvantage in trying to improve their standing in international capital markets.



Section 3

Solutions and recommendations

Section 3

Solutions and recommendations

The solutions set to tackle the qualitative data quality and transparency issues in sovereign risk assessment outlined in this report can be broken into three basic groups: (i) fix the factors, (ii) find an alternative, (iii) fix the process. This section lays out a menu of options of varying complexity and scale.

Recommendation #1

Update the qualitative rating scorecard factors

The updates to the Big Three credit rating agency (CRA) methodologies should correspond to and address the major gaps and shortcomings identified in this report. In general, this entails making the criteria, qualitative indicators, and guideposts much more specific, explicit, clear, and targeted at those qualitative factors that are demonstrably and materially associated with sovereign risk, while at the same time building in sufficient flexibility to take account of credit relevant factors that are not adequately captured by existing factors.

Accuracy

- Scrutinize past rating records and run statistical tests to isolate only those qualitative variables that are materially and demonstrably associated with sovereign creditworthiness.
- Correct the “democratic advantage” by removing any spurious criteria, indicators, and guideposts that explicitly or implicitly reference any factors that have a liberal democratic bias.
- Recalibrate model and scorecard weights to distinguish between countries with fundamental differences in structure (e.g., small island states, etc.).
- Create weighted governance scoring for different levels of the general government (i.e., central, state, local, or municipal) to reflect the importance of fiscal decentralization for government effectiveness.
- Create weighted scoring for credit relevant and financially material entities in the public sector, ensuring that both downside risk and credit uplift can be recognized in the sovereign rating.

Completeness

- Introduce criteria and/or scores related to the quality and transparency of fiscal policy planning and public financial management, covering both *de jure* and *de facto* features.
- Introduce criteria and/or scores related to climate resilience and stewardship of natural capital, including levels of disaster risk preparedness and protection of critical watersheds.

Consistency

- Create more granular scoring across qualitative factors to facilitate cross-country comparison and ensure more consistent treatment.
- Create more adaptable scoring methodologies to allow regular updates to the composition of sub-factors or sub-components in the Worldwide Governance Indicators (WGIs).

Validity

- Require more detailed commentary on scoring rationales, including with peer comparisons and supporting evidence (e.g., external guideposts used).
- Develop a new sovereign governance indicator with stronger validation checks hosted by a reliable third party.
- Increase scrutiny of the credit committee validation process by enabling external review and input on scoring decisions.

Timeliness

- Develop a new sovereign governance indicator that is updated on a more regular cadence than the CRA reviews and WGI releases.
- Include only high-frequency (e.g., at most one-year lag) or purely structural indicators.

Transparency

- Publish all relevant data, methodologies and code books, commentaries, and regulatory disclosures.
- Provide greater visibility into the committee review process, including potentially live streams.

Recommendation #2

Update the Worldwide Governance Indicators

The preceding sections revealed the pitfalls of relying too heavily on the WGIs for sovereign risk analysis. To summarize the main issues, complementing the findings of the 2024 *External Review of The Worldwide Governance Indicators* (World Bank 2024a):

- The WGIs only measure performance of countries relative to each other, and not absolute changes of governance quality within countries over time.
- They mask significant variation in governance within countries by assigning a single headline score per country.
- They do not provide a constructive road map for how to improve governance owing to the lack of absolute measures of governance.
- They rely heavily on perceptions-based data that are susceptible to bias, lags, and myopia, as well as the general challenges of research design in the internet era.
- They lose statistical significance when averaged due to correlated measurement errors (Kaufman and Kraay 2024; World Bank 2025b).

These critiques do not invalidate their utility for other applications, in particular Environmental, Social, and Governance (ESG) assessments and ratings where an explicit normative stance favoring liberal democratic institutions can be appropriate. Yet in their current form they are not fit for the purpose of assessing governments' willingness or ability to pay. For that, several updates are possible to correct some of the issues highlighted above.

- Strip out surveys and expert assessments that have a clear normative stance on the qualitative factor—e.g., “democratic advantage”—unless these have been independently tested and demonstrated to be statistically significant. The resulting indicator could be a recalibrated set of WGIs with different inputs and parameters, or a standalone “sovereign risk” WGI that combines a special mix of WGI sub-components.
- Introduce additional opinion surveys and expert assessments that cover qualitative factors relevant to climate and nature resilience, either under

existing pillars such as Policy Effectiveness or Regulatory Quality, or via a new dedicated WGI pillar. This could include the Environmental Sustainability indicator of the Country Policy and Institutional Assessments (CPIA).

- Introduce additional opinion surveys and expert assessments that cover fiscal planning and public financial management (PFM) considerations, including the broader set of World Bank's CPIA—including the Fiscal Policy and Debt Policy indicators, which cover critical matters related to disaster risk planning and debt sustainability, but are currently excluded from the WGIs—or Debt Management Performance Assessment (DeMPA) scores, or the Public Expenditure and Financial Accountability (PEFA) scores, which assess the strengths and weaknesses of PFM across 125 countries. These are valuable sources of credit relevant data, although they too may need to be revamped for use in the WGIs by increasing frequency, country coverage, transparency, and quality control.
- Introduce additional surveys and expert assessments from Global South countries that faithfully represent distinct ideological, cultural, and political perspectives.
- Allow only sources that disclose all underlying data, metadata, and scoring rationales, or else secure a separate agreement with commercial providers that they furnish all relevant information used for the WGIs.
- Improve access to the underlying sources—not just the data, but all applicable methodologies and code books, commentaries, and regulatory disclosures (e.g., funding sources).

These updates should be accompanied by robust empirical analysis to ensure goodness-of-fit with measures of sovereign default risk. The 2025 revisions to WGI methodology go some way towards addressing that gap by introducing stricter screening protocols and placing emphasis on “demonstrated relevance for economic and development outcomes” (World Bank 2025e). This led the World Bank to strip out 27 indicators from the 2024 WGI vintage, including removing three entire data sources (see Box 4 for details). However, as argued above, economic and development outcomes do not necessarily equate to sovereign creditworthiness, while the empirical record on “economic relevance” remains mixed.

Box 4. 2025 revision to the World Bank WGI

The 2024 vintage of the WGI was published in December 2025 under a revised methodology (see World Bank 2025e). This follows moves by the World Bank to take greater ownership of the WGI by moving its institutional home to the Development Economics Vice Presidency (DECVP), making it a formal World Bank product with the enhanced quality assurance that entails (World Bank 2024b). After several external and internal consultations, the group made six targeted changes:

1. Stricter screening protocols were applied to all data sources, with an emphasis on relevance, originality, cross-country coverage, regularity, transparency, and independence.
2. Three additional sources were added, including two to increase Global Source representations.
3. “Politically linked indicators,” mainly within the Voice and Accountability and the Political Stability and Absence of Violence WGI, were assessed based on their “demonstrated relevance for economic and development outcomes,” with some indicators removed.
4. More than 400 individual indicators were remapped to strengthen their conceptual alignment.
5. The methodology now allows the global average of governance over time.
6. The new methodology computes governance scores on an absolute 0–100 scale, adding to the prior annual governance represented in statistical units.

Separately, there is a valid case for stricter guidelines or regulatory provisions to limit the reliance on third-party indicators such as the WGI in credit ratings and other critical financial market signals. This is analogous to the rules that govern usage of credit ratings themselves in prudential regulation and investment advisory. The World Bank prominently displays a “Usage Advisory” on the WGI website, yet this does not seem to have discouraged the widespread application for credit risk metrics. To remedy this, securities regulators in key sovereign debt jurisdictions alongside associations such as the International Organization of Securities Commissions (IOSCO) should consider updating CRA regulations, codes of conducts, and other guidelines accordingly.

Recommendation #3

Create new qualitative indicators

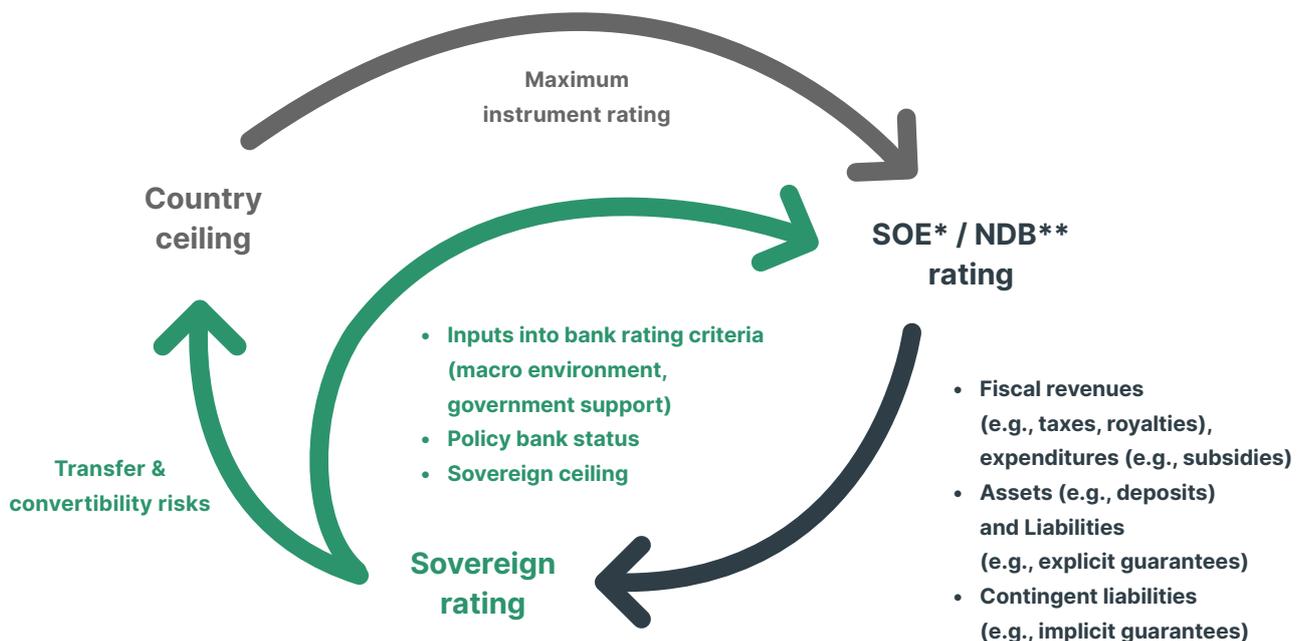
A new qualitative indicator could be designed from first principles and incorporate only those factors and criteria that can be demonstrably associated with sovereigns’ willingness and ability to pay. The aim would be to create a best-in-class qualitative indicator that the CRAs could use as an input variable or guidepost, substituting and augmenting the WGIs. It would retain the independence and quality control of the WGIs but use a tighter definition of governance and more granular scoring of factors such as institutional quality of sub-national governments and corporate governance of state-owned enterprises (SOEs). It would also create

a fresh opportunity to incorporate emerging issues surrounding climate risks and stewardship of natural capital, while introducing new types and innovative methods for compiling qualitative data (covered in the following recommendations). The drawback would be the lack of historical data and uptake by market participants, but this could be addressed through endorsement and co-creation by stakeholders including the CRAs.

A more expansive definition of fiscal governance that encompasses SOEs and national development banks (NDBs) would acknowledge their critical role in delivering policy outcomes and acting as a shock absorber. Policy banks are integral to policy effectiveness since they are a major conduit of financing investments into macro-critical growth- and resilience-enhancing infrastructure. Their ability to modulate government spending via countercyclical investment and lending operations can enhance policy flexibility. This function within the larger policy apparatus is not captured in the definition of general government and is typically only recognized by the CRAs when it impairs creditworthiness. The quality of capital and risk management and other operational considerations should be reflected symmetrically in sovereign scoring. By directly and explicitly linking fiscal and corporate governance in this way, the uplift to the sovereign from good corporate governance and effective capital deployment by SOEs and NDBs could feed back to their issuer ratings via the sovereign ceiling and systemic risk factors. Figure 7 depicts these interlocking relationships between the sovereign and public-sector entities.

Figure 7

Sovereign-SOE-NDB linkages



*SOE = State-owned enterprise. **NDB = National Development Bank.

Source: Author's analysis.

Recommendation #4

Develop performance-based qualitative scores

A new type of qualitative data would complement (a) *de-jure* descriptions of institutions and (b) perceptions of *de facto* implementation with (c) performance against credit relevant targets. The last are considered qualitative since it is not the values of the key performance indicators (KPIs) *per se* that count—these may be tracked via the quantitative factors. Rather, what matters is the institutionalized target setting, performance tracking, and implementation mechanisms. If the KPIs are credit relevant and financially material, especially if they are targeted at key credit weaknesses, then performance tracking systems can help to signal fiscal responsibility and anchor policy commitments through political and business cycles, bolstering credibility, predictability, and effectiveness. For example, targets trained on physical climate risks with KPIs that track adaptation measures such as

investments in climate proof infrastructure demonstrate greater willingness to pay by their implementation of performance-based measures intended to strengthen resilience to shocks.

Performance-based instruments in public financial management are mechanisms that link funding to results (outputs and/or outcomes). They can be instituted both on the spending side with program budgets and bonus funding, or on the financing side with interest relief or other concessional terms linked to outcomes. Commitments are enshrined in budget system laws or loan contracts, which together with robust KPIs and strict reporting requirements provide a degree of accountability. This is reinforced by performance incentives in the form of financial penalties and rewards, for instance step-downs in interest rates or higher budgetary allocations if targets are achieved. Grounding the targets and KPIs in macro-fiscal plans and national strategies with strong government backing can bolster the credibility of commitments. Ideally, they are also based on sound modeling and scenario analysis.

In a sovereign risk context, KPI-linking is analogous to fiscal rules and targets, which are widely employed by sovereigns as an institutional mechanism to limit fiscal policymakers' discretion over actions that undermine debt sustainability (IISD 2025). Although not a requirement, CRAs view fiscal rules positively, as "consistent compliance with those targets over a number of political cycles generally signal stronger fiscal policymaking and implementation" (Moody's 2022). Moody's criteria give highest scores under its Institutions and Governance Strength pillar for countries where "Fiscal targets or expenditure ceilings are observed or outperformed" (Moody's 2022). Fitch considers the "prudence and transparency of fiscal rules and the credibility of fiscal framework" when evaluating whether to apply a qualitative adjustment under the Public Finances pillar (Fitch 2025a).

Following a similar logic, performance-based frameworks and systems should be recognized in CRA criteria as markers of good fiscal governance and coded accordingly, especially if they incorporate financially material targets that generate credit uplift when achieved. And because ratings assessments are forward looking—i.e., the ratings reflect both the current and near-term projections of credit drivers—their adoption should be afforded some benefit upfront in qualitative scoring.

Aside from reinforcing credibility of commitment, the KPIs can themselves target governance outcomes. For example, Sri Lanka issued a state-contingent debt instrument as part of its 2024 sovereign debt restructuring that included a coupon step-down linked to (i) the level of government revenues to GDP in 2026 and 2027 and (ii) the annual publication of a detailed Fiscal Strategy Statement in line with the country's new Public Financial Management Act (Lalo 2025). Other fiscal or policy governance KPIs can be envisioned for such performance-based financing arrangements that map more closely to the CRA qualitative factors.

Recommendation #5

Enforce "radical" transparency in qualitative scoring processes

Many of the data quality and transparency issues cited above pertain to the process of generating and compiling qualitative data, rather than the content of the scores or criteria. Upgrades should maximize the objectivity of scoring decisions, ensure that scorers are competent and unbiased, and grant full visibility into the review and validation process.

Sunlight is the simplest remedy: all inputs to the ratings and WGIs should be available in their entirety and made publicly accessible; the profiles and credentials of the decision-makers should be disclosed; and committee deliberations should be open to scrutiny, for instance, by publishing detailed minutes. Fully open sourcing sovereign ratings would have financial repercussions for CRAs' paid subscription services and the commercial business information providers feeding the WGIs. CRA analysts may be less candid about their views in committee if they fear disclosure of material non-public information or reprisals from their clients.

However, there is a strong public interest argument for making sovereign rating analysis—not the rating themselves—a public good, given the repercussions of the decisions on countries' cost of capital and by extension businesses whose ratings are constrained by the sovereign ceiling. Furthermore, CRAs would still collect fees for rating sovereign bond issuances, while the impact on subscriptions would be cushioned by non-sovereign research. The hit to commercial data providers in the WGI would likely be bigger if sensitive intellectual property is exposed, so they should have the option to opt out of the WGI.

Radical transparency will only partially address the problem of subjectivity and bias. Improved access to information after the fact does not necessarily alter rating decisions, even if public scrutiny makes analysts more guarded in their judgment calls. Opening up the committee process to external reviewers before ratings are finalized would provide stronger accountability and quality control. One possibility would be an independent platform or hub such as the African Peer Review Mechanism (APRM) that convenes outside stakeholders, coordinates their engagement with the CRAs, and consolidates inputs and feedback for each rating review. At the same time, the platform would scrutinize the rating rationale, raise objections to fallacious reasoning, and call out inconsistencies or lapses in data quality. It could also offer sovereign advisory services, ideally free of charge or at nominal rates, to issuers who are struggling to engage on an equal footing with the CRAs.

A disadvantage of such an approach is that it risks clogging up the rating process without meaningfully influencing scoring decisions or fundamentally solving the problems of subjectivity or concentration of decision-making power. Unless required by regulation, the CRAs are free to ignore or dismiss the challenges if they are not factual errors or material omissions. Claims about biased or incorrect

qualitative scoring are difficult to substantiate, as the first section of this report has explained. Plus, the reviewers would need to demonstrate their credentials and competence to opine on matters of sovereign risk to avoid amplifying distortions and to be taken seriously by the CRAs. In other words, they need to be credible.

Recommendation #6

Put in place credibility-based qualitative scoring

In the absence of clear benchmarks for evaluating ability and willingness to pay, the validity of qualitative scoring depends on the credibility of the CRAs' analysts. That, in turn, is a function of the issuers' credibility in terms of delivering macro-fiscal stability and debt sustainability over the rating horizon. The credibility gap is what an issuer commits in terms of policies, plans, and programs, and what the CRA analysts believe based on the track record of fiscal prudence and the strength of their implementing institutions. The wider this gap, the more scope for analysts to insert their subjective opinions, and the more weight is accorded to the credibility of the analysts and committee. One way of narrowing this gap is for the issuer to strengthen its commitments and clearly document its record. The performance-based scoring approach outlined in Recommendation #4 can help with that.

The irony is that whereas CRAs demand a track record of good performance from issuers for their commitments to be considered credible, the CRAs—and likewise the scorers in the WGI—are not held to the same standard. This depends partly on what is being tracked. If it is their record of correctly calling default, then CRAs get a passing grade, although this is an imperfect yardstick given the rare occurrence of sovereign defaults and the potential for self-fulfilling prophecies. If it is about the accuracy of their quantitative forecasts, then these can be backtested against actual outcomes, although large errors are a common hazard of the macroeconomic forecasting trade, and the CRAs do not stand out as much worse than their peers. That leaves qualitative scoring, for which inter-subjective validation checks offer a suboptimal fix, and backtesting has not been possible—until now.

Superforecasting offers a practical way to close this asymmetry and introduce genuine quality-based scoring (see Box 5). In these structured exercises, forecasters, whether human experts or artificial intelligence (AI) models, assign explicit probabilities to future events such as fiscal distress, policy reversal, or reform success. Once outcomes occur, forecasts are scored for accuracy and calibration. This simple discipline, testing judgment against reality, creates objective measures of analytical quality. It exposes biases, highlights effective reasoning patterns, and builds an evidence base for what kinds of qualitative assessments actually align with real-world outcomes.

The tournament method has gained prominence with the work on “superforecasting” popularized by Tetlock and Gardener (2015). Applied to sovereign ratings, such tournaments would allow credit rating agencies and international institutions to test and refine their qualitative factors empirically. Assessments of policy effectiveness or institutional strength, for example, could be reframed as probabilistic forecasts estimating the likelihood that a government maintains fiscal discipline, avoids arrears, or implements announced reforms. Repeated testing over time would show which qualitative dimensions truly predict sovereign performance, allowing rating frameworks to evolve through data rather than doctrine.

Beyond improving quality control, forecasting tournaments also make possible an epistemic change in how qualitative analysis is conceived. The current model begins with pre-set categories such as political stability, governance quality, and rule of law, and evaluates countries against them as if those constructs inherently determine creditworthiness. The new model inverts this logic. It starts instead with the outcome that matters most, the probability of default, and lets the process of forecasting reveal which dimensions of governance and policy actually matter. Because forecasts are scored and compared, analysts are incentivized to focus on what improves predictive accuracy, not on what sounds convincing or aligns with familiar narratives. Over time, categories themselves emerge from demonstrated predictive value rather than from normative assumptions.

This epistemic shift also advances transparency, accountability, and timeliness. Forecasts and rationales are explicit and trackable, making the reasoning behind qualitative judgments open to review and replication. Analysts and institutions can be held responsible for their performance because the quality of their qualitative assessments becomes measurable. Moreover, forecasting-based evaluation aligns sovereign analysis with the time horizon of investors. It focuses on the future, the very domain where investment decisions and risk perceptions operate, rather than on backward-looking indicators.

In areas where no clear empirical benchmark exists, reciprocal scoring can extend this logic. Here, forecasters predict not hard outcomes but the distribution of forecasts produced by their peers, essentially anticipating what the most accurate forecasters will predict. This creates an intersubjective

scoring rule similar in spirit to a prediction market, where the best way to anticipate the consensus is to reason as accurately as possible oneself. Even without hard data, this structure still rewards disciplined reasoning and enables transparent evaluation of judgment quality.

Together, these innovations mark a shift from categorical evaluation to forecast-based learning. Qualitative analysis becomes a dynamic, self-correcting process, adaptive to evidence, measurable over time, and grounded in real predictive performance. In this sense, forecasting tournaments are not only a methodological improvement but the foundation of a new epistemic architecture in which qualitative scoring becomes transparent, testable, and empirically anchored. They offer a credible pathway toward a system of sovereign assessment that is accountable not only to markets but also to truth itself.

Box 5. Mapping Pathways for Security (MPS)—forecasting tournaments as a tool or counterfactual policy evaluation

The Mapping Pathways for Security (MPS) initiative⁸ illustrates how forecasting tournaments can be used to evaluate policy effectiveness and governance credibility—the same qualitative dimensions that underpin sovereign risk assessments. It applies the tools of structured forecasting not only to anticipate events but also to estimate the likely impact of alternative policy choices, creating a quantitative framework for learning about “what works” in fragile and uncertain environments.

In conventional forecasting tournaments, participants predict what will happen and are later scored on accuracy once outcomes are observed. MPS extends this idea to counterfactual reasoning by posing questions about what would happen if particular interventions were implemented. Analysts and AI systems are asked to assign probabilities to conflict or instability under two contrasting scenarios: **a baseline trajectory** and a **with-intervention scenario** reflecting a proposed reform, investment, or diplomatic action. The difference between these forecasts represents the estimated marginal effect of the policy as a forward-looking measure of its expected contribution to stability or risk reduction. When real-world data eventually become available, these judgments can be scored and compared, closing the feedback loop between analysis, action, and evaluation. When this is not possible, reciprocal scoring rules are used to maintain incentives and calibration (see above).

By embedding these mechanisms in a structured process (forecasting, policy/project design and convenings) MPS offers a model to reduce the probability of conflict. It does so by eliciting patterns, indicators, and levers associated with accurate predictions.

As large language models and machine learning systems now perform at human level in forecasting tasks, initiatives like MPS make it possible to scale such exercises globally. They point toward a future where qualitative assessments of governance are not only explained but empirically validated through continuous testing, enabling more credible, transparent, and adaptive sovereign evaluation.

⁸ See <https://mpsglobal.org/>.

Conclusion

The aim of this report was to demystify the role of qualitative analysis in sovereign credit ratings and highlight major issues with how the Big Three CRAs apply qualitative criteria and use third-party indicators such as the World Bank's WGI. In drawing attention to the soft side of sovereign risk analysis, it plugs a major gap in the existing literature on credit ratings, which to date have focused on the quantitative drivers of ratings and the quality of "hard" data.

A deeper inspection of the qualitative factors and third-party indicators reveals material problems with the way that both sets of inputs are defined and compiled under current methodologies. It casts doubt on whether many of them are even relevant for assessing sovereign creditworthiness given the weak empirical evidence on which to base their inclusion in the rating models and scorecards. To the extent that a statistically significant relationship cannot be conclusively demonstrated between certain qualitative factors such as type of political system or degree of openness and sovereigns' likelihood of default, the report points to a systematic bias embedded within the inputs and criteria.

To be clear, this does not imply a normative position on whether these institutional features are beneficial or desirable in other contexts, but rather that they have limited predictive power for assessing a government's willingness and ability to pay their sovereign debt obligations. Nor does the analysis suggest that the WGIs are not good measures of governance in general or useful for other purposes outside of sovereign risk analysis—just that they should be avoided or de-emphasized in sovereign credit analysis in favor of narrower and more targeted definitions of fiscal governance.

The report calls for closer scrutiny of these issues by relevant policymakers, regulators, and other stakeholders in sovereign debt markets. It pinpoints key areas for improvement and suggests concrete proposals for how to change the rating methodologies and processes accordingly.

The proposed solutions range from those that entail relatively modest technical upgrades, to others that imply a more fundamental shift in approach through the use of artificial intelligence. The latter options may be necessary to adapt to an increasingly complex and uncertain sovereign risk environment under growing strain from geoeconomic fragmentation, AI-driven technological disruption, and climate-induced physical destruction. In practical terms, these changes impose a far heavier analytical workload on CRAs and demand a wider breadth of expertise from their sovereign analysts in order to parse signals from noise. Agentic AI can be a valuable tool to augment analytical capabilities and check human decision-making, provided the appropriate safeguards are in place.

The first step in addressing these issues should be additional research to further substantiate the claims of the report, and to isolate the core set of institutional drivers of sovereign creditworthiness. That should ideally include quantifying the potential distortionary effects of erroneous or biased qualitative inputs in terms of excess risk premia and opportunity costs for EMDE sovereigns. In parallel, prototyping and experimentation are needed to test whether an AI-based approach is effective and feasible.

There is an urgent need to tackle these issues. The disproportionate weight of qualitative factors in sovereign rating means that the distortionary effect on sovereign risk pricing and EMDE capital flows could be material. In the context of historic EMDE public debt levels, escalating financing needs for resilience and development, and volatile financial conditions, getting institutional assessments right could determine whether many EMDE debts are sustainable or not.

References

- Abbas, A., A. Pienkowski, and K. Rogoff, eds. 2020. *Sovereign Debt: A Guide for Economists and Practitioners*. Oxford University Press.
- Agarwala, M., M. Burke, P. Klusak, M. Kraemer, and U. Volz. 2022. *Nature Loss and Sovereign Credit Ratings*. NatureFinance and Bennett Institute for Public Policy. <https://www.naturefinance.net/wp-content/uploads/2022/09/NatureLossSovereignCreditRatings.pdf>.
- Barta, Z., and A. Johnston. 2023. *Rating Politics: Sovereign Credit Ratings and Democratic Choice in Prosperous Developed Countries*. Oxford University Press.
- Beaulieu, E., G. W. Cox, and S. Saiegh. 2012. "Sovereign Debt and Regime Type: Reconsidering the Democratic Advantage." *International Organization*, 66 (4): 709–38.
- Bertelsmann Stiftung. 2024. *BTI 2024 Country Report: Brazil*. https://bti-project.org/fileadmin/api/content/en/downloads/reports/country_report_2024_BRA.pdf.
- Biglaiser, G., and J. L. Staats. 2012. "Finding the 'Democratic Advantage' in Sovereign Bond Ratings: The Importance of Strong Courts, Property Rights Protection, and the Rule of Law." *International Organization*, 66 (3): 515–35.
- Bodea, C., and R. Hicks. 2018. "Sovereign Credit Ratings and Central Banks: Why Do Analysts Pay Attention to Institutions?" *Economics & Politics*, 30 (3): 340–65.
- Bornhorst, F. 2019. "Strengthening the Subnational Fiscal Framework." In *Brazil: Boom, Bust, and the Road to Recovery*, edited by A. Spilimbergo and K. Srinivasan. International Monetary Fund. <https://www.elibrary.imf.org/display/book/9781484339749/ch012.xml>.
- Capitals Coalition. n.d. "Nature on the Balance Sheet Initiative." <https://capitalscoalition.org/project/nature-on-the-balance-sheet/>.
- Cash, D., and H. Khan. 2024. *Rating the Globe: The Role of Credit Rating Agencies in the Global Financial System*. United Nations University. https://collections.unu.edu/eserv/UNU:9832/rating_the_globe.pdf.
- Ceglar, A., M. Stevanović, P. von Jeetze, et al. 2024. *Climate-Nature Scenario Development for Financial Risk Assessment*. NatureFinance. <https://www.naturefinance.net/wp-content/uploads/2024/02/InterimReport2024-6.pdf>.
- Cevik, S., and J. Tovar Jalles. 2020. *This Changes Everything: Climate Shocks and Sovereign Bonds*. Working Paper No. 20/79. International Monetary Fund. <https://www.imf.org/en/Publications/WP/Issues/2020/06/05/This-Changes-Everything-Climate-Shocks-and-Sovereign-Bond-s-49476>.
- CountryRisk.io. 2020. "Sovereign Rating Model (Version July 2020)." <https://www.countryrisk.io/methodology/sovereign-rating-model/version-july-2020>.
- DBRS. 2024. *Global Methodology for Rating Sovereign Governments*. <https://prod.dbrs.com/research/436000/global-methodology-for-rating-sovereign-governments-archived>.
- Economist Intelligence Unit. 2024. *Democracy Index 2024*. <https://www.eiu.com/n/campaigns/democracy-index-2024/>.
- Fitch Ratings. 2021. *ESG Relevance Scores for Sovereigns 2.0*. <https://www.fitchratings.com/research/sovereigns/esg-relevance-scores-for-sovereigns-20-07-2021>.
- Fitch Ratings. 2022. *What Investors Want to Know: Petrobras Pressure Points Post-presidential Elections*. <https://www.fitchratings.com/research/corporate-finance/what-investors-want-to-know-petrobras-pressure-points-post-presidential-elections-03-11-2022>.
- Fitch Ratings. 2023. *Country Ceiling Criteria*. <https://www.fitchratings.com/research/sovereigns/country-ceiling-criteria-24-07-2023>.

- Fitch Ratings. 2025a. Sovereign Rating Criteria. <https://www.fitchratings.com/research/sovereigns/sovereign-rating-criteria-15-09-2025>.
- Fitch Ratings. 2025b. Rating Subnationals above the Sovereign: Institutional Recognition and Autonomy Protect from Sovereign's Default. <https://www.fitchratings.com/research/international-public-finance/rating-subnationals-above-sovereign-28-10-2025>.
- Fitch Ratings. 2025c. "Fitch Affirms Mexico at 'BBB-'; Outlook Stable." April 16. <https://www.fitchratings.com/research/sovereigns/fitch-affirms-mexico-at-bbb-outlook-stable-16-04-2025>.
- Fitch Ratings. 2025d. "Fitch Affirms Petrobras' IDRs at 'BB'; Outlook Stable." April 3. <https://www.fitchratings.com/research/corporate-finance/fitch-affirms-petrobras-idrs-at-bb-outlook-stable-03-04-2025>.
- Fitch Ratings. 2025e. Sovereigns Climate Vulnerability Signals: Discussion Paper. <https://www.fitchratings.com/research/sovereigns/sovereigns-climate-vulnerability-signals-discussion-paper-27-01-2025>.
- Gbohoui, W., R. Ouedraogo, and Y. M. Some. 2023. Sub-Saharan Africa's Risk Perception Premium: In the Search of Missing Factors. Working Paper No. 2023/130. International Monetary Fund. <https://www.imf.org/en/publications/wp/issues/2023/06/23/sub-saharan-africas-risk-perception-premium-in-the-search-of-missing-factors-534885>.
- Gratcheva, E. M., T. Emery, and D. Wang. 2020. Demystifying Sovereign ESG. World Bank Group. <https://documents1.worldbank.org/curated/en/842671621238316887/pdf/Demystifying-Sovereign-ESG.pdf>.
- Griffith-Jones, S., and M. Kraemer. 2021. Credit Rating Agencies and Developing Economies. DESA Working Paper No. 175. United Nations Department of Economic and Social Affairs. <https://desapublications.un.org/working-papers/credit-rating-agencies-and-developing-economies>.
- Hansen, D. 2023. "The Democratic (Dis)advantage: The Conditional Impact of Democracy on Credit Risk and Sovereign Default." *Economics & Politics*, 35 (1): 356–410. <https://doi.org/10.1111/ecpo.12218>.
- Haschke, P. 2022. The Political Terror Scale (PTS) Codebook v2.00. University of North Carolina Asheville. <https://www.politicalterrorsscale.org/Data/Files/PTS-Codebook-V200.pdf>.
- Institute of International Finance. 2019. Voluntary Principles for Debt Transparency. <https://www.iif.com/portals/0/Files/Principles%20for%20Debt%20Transparency.pdf>.
- Institute of International Finance. 2025. 2025 IIF Investor Relations and Debt Transparency Report: Unlocking Affordable Capital For Emerging Markets—The Vital Role of Investor Relations. <https://www.iif.com/Publications/ID/6219/2025-IIF-Investor-Relations-and-Debt-Transparency-Report>.
- International Institute for Sustainable Development. 2025. Design Choices for Debt Sustainability. <https://www.iisd.org/system/files/2025-06/design-choices-debt-sustainability.pdf>.
- International Monetary Fund. n.d. "Special Data Dissemination Standard: SDDS Overview." <https://dsbb.imf.org/sdds/overview>.
- International Monetary Fund. 2018a. Review of 1997 Guidance Note on Governance—A Proposed Framework for Enhanced Fund Engagement. <https://www.imf.org/-/media/files/publications/pp/2018/pp030918govpaper.pdf>.
- International Monetary Fund. 2018b. Guidance Note for the Use of Third-party Indicators in Fund Reports. <https://www.imf.org/-/media/files/publications/pp/2018/pp071718guidance-note-for-the-use-of-third-party-indicators-in-fund-reports.pdf>.
- International Monetary Fund. 2021. Review of the Debt Sustainability Framework for Market Access Countries. <https://www.imf.org/en/-/media/files/publications/pp/2021/english/ppea2021003.pdf>.
- International Monetary Fund. 2025. October 2025 World Economic Outlook Database. <https://data.imf.org/en/datasets/IMF.RES:WEO>.
- Kaufmann, D., and A. Kraay. 2024. The Worldwide Governance Indicators: Methodology and 2024 Update. World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099005210162424110>.

- Kraemer, M., M. Msrnik, A. Petrov, and B. S. Glass. 2015. "Storm Alert: Natural Disasters Can Damage Sovereign Creditworthiness." S&P Global Ratings, September 10.
<https://www.spglobal.com/ratings/en/research/articles/150910-storm-alert-natural-disasters-can-damage-sovereign-credit-worthiness-9327571>.
- Kulenkampff, A. 2025. Nature as a Shock Absorber: A Financial Materiality Assessment of Forestry-linked Sovereign Indicators in Ghana. NatureFinance.
<https://www.naturefinance.net/resources-tools/financial-materiality-assessment-framework/>.
- Lalo, E. 2025. "Important Innovations in Sovereign Finance: The Arrival of Macro-Linked Bonds (MLBs) and Governance-linked Bonds (GLBs)." Rothschild & Co., December 13.
<https://www.rothschildandco.com/en/newsroom/insights/2024/12/ga-important-innovations-in-sovereign-finance/>.
- London Stock Exchange Group. 2021. Sovereign Risk Monitor Methodology. FTSE Russell.
https://www.lseg.com/content/dam/ftse-russell/en_us/documents/research/sovereign-risk-monitor-methodology.pdf.
- Lysenko, T. 2025. Tracing the Drivers of SSA's Low Sovereign Credit Ratings. Finance for Development Lab.
<https://findevlab.org/tracing-the-drivers-of-ssas-low-sovereign-credit-ratings/>.
- Mitra, P., M. Raissi, B. Versailles, et al. 2025. Integrating Climate Change into Macroeconomic Analysis: A Review of Impact Channels, Data, Models, and Scenarios. Working Paper No. 2025/087. International Monetary Fund.
<https://www.imf.org/en/publications/wp/issues/2025/08/26/integrating-climate-change-into-macroeconomic-analysis-a-review-of-impact-channels-data-569996>.
- Moody's Investors Service. 2020. Country Ceilings Methodology. <https://ratings.moodys.com/rmc-documents/68640>.
- Moody's Investors Service. 2022. Rating Methodology: Sovereign. <https://ratings.moodys.com/rmc-documents/395819>.
- Moody's Investors Service. 2024. Better Data Transparency Would Boost Credit Quality of Frontier and Emerging Markets.
- MSCI. 2025. MSCI ESG Government Ratings Methodology.
<https://www.msci.com/downloads/web/msci-com/data-and-analytics/sustainability-solutions/MSCI%20ESG%20Government%20Ratings%20Methodology.pdf>.
- Nakatani, R. 2025. Preventing Fiscal Crises under Decentralization: Intergovernmental Policies and Institutions. Working Paper No. 2025/087. International Monetary Fund.
<https://www.imf.org/-/media/Files/Publications/WP/2025/English/wpia2025087-print-pdf.ashx>.
- Organisation for Economic Co-operation and Development. n.d. "Participants' Country Risk Classification."
<https://www.oecd.org/en/topics/sub-issues/country-risk-classification.html>.
- Organisation for Economic Co-operation and Development. 2022. OECD Debt Transparency Initiative: Trends, Challenges and Progress. https://www.oecd.org/en/publications/oecd-debt-transparency-initiative_66b1469d-en.html.
- Political Terror Scale. 2025. "Political Terror Scale Levels."
<https://www.politicalterroryscale.org/Data/Documentation.html#PTS-Levels>.
- Prodani, J., S. Gallet, D.-J. Jansen, I. Kearney, G. Schotten, G. Brouwer, and W. Jan van Zeist. 2023. "The Economic and Financial Stability Repercussions of Nature Degradation for the Netherlands: Exploring Scenarios with Transition Shocks." DNB Occasional Studies, 21 (2).
- Romelli, D. 2022. "The Political Economy of Reforms in Central Bank Design: Evidence from a New Dataset." *Economic Policy*, 37 (112): 641–88. <https://doi.org/10.1093/epolic/eiac011>.
- Ryder, H., Aggad, F., Mutize, M., Lopes, C., Naidoo, C., Kaboub, F., Nkosi, R., and Braganza, J. 2024. "Letter: Little Surprise African Governments Reject Credit Rating Agencies." Letter to the editor. *Financial Times*, February 5.
<https://www.ft.com/content/f973cc4b-2764-4ec8-b0e6-a255d2edb4d9>.
- S&P Global Ratings. 2009. "General Criteria: Methodology: Criteria For Determining Transfer And Convertibility."
<https://www.spglobal.com/ratings/en/regulatory/article/-/view/sourcelid/5402435>.

S&P Global Ratings. 2017. "Sovereign Rating Methodology." <https://www.spglobal.com/ratings/en/regulatory/article/-/view/sourceId/10221157>.

S&P Global Ratings. 2025. "Research Update: Papua New Guinea 'B-/B' Ratings Affirmed; Outlook Stable." June 27. <https://www.spglobal.com/ratings/en/regulatory/article/-/view/type/HTML/id/3397301>.

Schulz, K. A., and B. R. Weingast. 2003. "The Democratic Advantage: Institutional Foundations of Financial Power in International Competition." *International Organization*, 57 (1): 3–42.

Scope Ratings. 2025. Sovereign Rating Methodology: Sovereign and Public Sector. <https://www.scopegroup.com/ScopeGroupApi/api/methodology?id=01508950-119c-4ab5-9182-54ffdc1003f>.

Tinubu, B. 2026. "Africa Needs Its Own Credit Rating Agency." *Financial Times*, February 16. <https://www.ft.com/content/0671e4f5-e0d3-47e3-ad7c-509b9e785687>.

Tshabalala, S. 2025. "The Global South Does Not Need a New Credit Rating System." *Financial Times*, November 12. <https://www.ft.com/content/49f4c8b5-7d69-455d-9262-c97973c7ad53>.

Tetlock, P., and D. Gardener. 2015. *Superforecasting: The Art and Science of Prediction*. Penguin Random House.

United Nations Development Programme. 2025. Sovereign Credit Ratings: Perspectives for Africa's Development. <https://www.undp.org/africa/credit-ratings-resource-platform/publications/sovereign-credit-ratings-perspectives-africas-development>.

United Nations Trade and Development. 2025. Credit Rating Agencies, Developing Countries and Bias. https://unctad.org/system/files/official-document/gds2024d3_en.pdf.

Wheatley, J. 2021. "World Bank Inquiry Finds IMF Chief Pushed Staff to Boost China Rankings." *Financial Times*, September 16. <https://www.ft.com/content/029b75dc-86a0-4ebb-956b-a3135b6c510d>.

World Bank. 2021. CPIA Criteria 2021. <https://thedocs.worldbank.org/en/doc/69484a2e6ae5ecc94321f63179bfb837-0290032022/original/CPIA-Criteria-2021.pdf>.

World Bank. 2024a. External Review of The Worldwide Governance Indicators. <https://documents1.worldbank.org/curated/en/099951310312414964/pdf/IDU-433971de-1de5-4d9f-835a-026fc163b3b0.pdf>.

World Bank. 2024b. External Review of the Worldwide Governance Indicators: Review Panel Recommendations and Management Responses. <https://www.worldbank.org/content/dam/sites/govindicators/doc/wgireviewresponse.pdf>.

World Bank. 2025a. International Debt Report 2025. <https://openknowledge.worldbank.org/entities/publication/7c0dbf75-2bd7-4ae3-9db9-91318290c781>.

World Bank. 2025b. CPIA Africa, July 2025: Policies For Better Service Delivery. <https://openknowledge.worldbank.org/server/api/core/bitstreams/f506bc30-2670-4ddd-adb0-3ae8f554aefb/content>.

World Bank. 2025c. Radical Debt Transparency. <https://documents1.worldbank.org/curated/en/099439106192540687/pdf/IDU-a86e3bb4-3446-4716-9131-085141b81950.pdf>.

World Bank. 2025d. Worldwide Governance Indicators. <https://www.worldbank.org/en/publication/worldwide-governance-indicators>.

World Bank 2025e. The Worldwide Governance Indicators: Revised Methodology for Measuring Governance Using Perception Data. <https://www.worldbank.org/content/dam/sites/govindicators/doc/The%20Worldwide%20Governance%20Indicators%202025%20Methodology%20Revision.pdf>.

