Reading the Fund

A Systematic Analysis of IMF Fiscal Advice Using Large Language Models[1]

Jantsankhorloo Amgalan[*], Jérémie Cohen-Setton[*], and Anton Korinek[**]

This version: November 11, 2025

Abstract

This paper estimates the IMF's implicit fiscal reaction function by applying large language models (LLMs) to extract and classify recommendations from over 3,000 Article IV reports over the past 25 years. Extending the reaction function literature from actual fiscal behavior to policy advice, we test whether IMF recommendations follow systematic patterns analogous to monetary policy rules. We find that fiscal advice is systematically countercyclical, tightening when output gaps are positive and loosening during recessions. This countercyclicality was weaker pre-2008 and has strengthened markedly since the Global Financial Crisis, particularly whenever monetary policy was constrained. The estimated reaction function varies predictably: countries with stronger fiscal institutions receive more countercyclical recommendations, while responsiveness decreases with income level. Non-economic factors also matter, with IMF mission chiefs' graduate training influencing the pattern of advice. Our paper demonstrates how LLMs can contribute to the systematic evaluation of economic policy advice.

[1] We thank Chengyu Huang for his invaluable assistance in assembling the corpus of Article IV reports; IMF staff for sharing several internal datasets; Ayhan Kose, Sergio Kurlat, Franziska Ohnsorge, and Naotaka Sugawara for access to their dataset on fiscal space; and Valentin Lang, Lukas Wellner, and Alexandros Kentikelenis for sharing their dataset on IMF mission chiefs. We are also grateful to Barry Eichengreen, Jason Furman, Peter Montiel, and participants in seminars at the IEO and the Peterson Institute for International Economics for their comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily represent those of the IEO, the IMF, or IMF policy. IEO Working Papers report analyses related to the work of the IEO and are published to elicit comments and to further debate. Korinek performed part of this work while he was a Consultant of the IMF's Independent Evaluation Office.

[*] Independent Evaluation Office of the International Monetary Fund.

[**] University of Virginia, NBER, and CEPR.

I. INTRODUCTION

For decades, the International Monetary Fund has faced criticism that its fiscal policy advice suffers from 'austerity bias,' reflexively prescribing fiscal consolidation regardless of economic conditions (Stiglitz, 2002; Krugman, 2013). Critics argued the institution pushed contractionary policies that deepened recessions, while defenders claimed it tailored advice to country circumstances. Yet despite the intensity of this debate and the IMF's influence over 191 countries' fiscal policies, no systematic evidence exists on what the Fund actually recommends across countries and time.

The absence of systematic evidence is striking given the stakes involved. The Fund's recommendations shape national policy debates, anchor lending programs, and influence market perceptions. Previous empirical studies have relied on limited case studies or small hand-coded samples, leaving fundamental questions unanswered: Does the Fund exhibit systematic austerity bias? Is advice procyclical or countercyclical? How does it vary with economic conditions and across country groups? Has it evolved with changing economic paradigms?

Recent advances in large language models (LLMs) make it possible to study IMF advice at an unprecedented scale. This paper applies these tools to more than 3,000 Article IV reports covering 193 countries from 1998 to 2023, the first comprehensive, cross-country analysis of its kind. Building on the breakthroughs described in Korinek (2023, 2024), LLMs can now process and interpret complex policy documents with a level of nuance approaching human judgment while maintaining consistency across thousands of texts. This capability allows, for the first time, a systematic examination of the IMF's fiscal policy advice across its full membership.

We approach this task by estimating the Fund's implicit fiscal reaction function, analogous to how the literature has studied central banks' monetary policy rules or governments' actual fiscal behavior (Bohn, 1998; Taylor, 1993). By treating IMF recommendations as the dependent variable and economic conditions as explanatory factors, we can test whether the Fund's advice follows systematic patterns, how these patterns compare to theoretical benchmarks, and whether they vary across countries and time. This framework transforms qualitative policy text into a quantifiable reaction function that can be empirically analyzed.

This paper makes three main contributions. First, it develops and validates a methodology for using LLMs to analyze IMF policy advice, establishing a framework that could be applied to other aspects of Fund surveillance. Second, it provides the first comprehensive characterization of how the Fund's near-term fiscal policy advice varies with economic circumstances, documenting both systematic patterns and their evolution over time. Third, it offers a framework that researchers could build upon to identify cases where advice deviates from typical relationships, helping ensure that such deviations are reviewed more closely and justified by country-specific factors, and ultimately promoting greater consistency and evenhandedness in Fund advice. Together, these contributions provide the first systematic evidence to evaluate long-standing claims about IMF fiscal policy advice.

We find that IMF fiscal advice indeed places a strong emphasis on consolidation, particularly over the medium term. A large share of recommendations calls for tightening the fiscal stance, over 90 percent in the case of medium-term advice to emerging market and middle-income economies (EMMIEs) and low-income countries (LICs). This reflects the Fund's continued focus on maintaining fiscal sustainability. Among the fiscal indicators shaping this advice, staff assessments of fiscal balance (FB) gaps—measuring the adjustment in the FB required to stabilize the debt-to-GDP ratio—tend to matter more than the debt ratio itself in advanced economies (AEs). This suggests that advice is often guided by the perceived trajectory of fiscal adjustment rather than static debt thresholds. At the same time, stronger fiscal institutions—such as fiscal rules, medium-term frameworks, and debt-anchoring mechanisms—also influence advice, enabling staff to support more countercyclical fiscal policies when credible safeguards are in place.

Yet, we find that the Fund's advice is by no means uniformly contractionary. For the near-term, recommendations are generally countercyclical, more expansionary in periods of economic slack and more contractionary when activity is strong. This countercyclicality has strengthened over time, particularly during episodes when interest rates were low and monetary policy was constrained by the zero lower bound (ZLB), underscoring the Fund's growing recognition of fiscal policy as a key stabilization tool when monetary space is limited. Moreover, in the exceptional macroeconomic environment of the 2010s—characterized by secular stagnation and persistently low interest rates, the Fund increasingly advised countries with ample fiscal space and large current account surpluses to relax their medium-term consolidation plans. In these cases, the Fund's advice on the appropriate medium-term fiscal stance was framed not only around maintaining fiscal sustainability but also around supporting long-term investments, facilitating structural transformation, and addressing external imbalances.

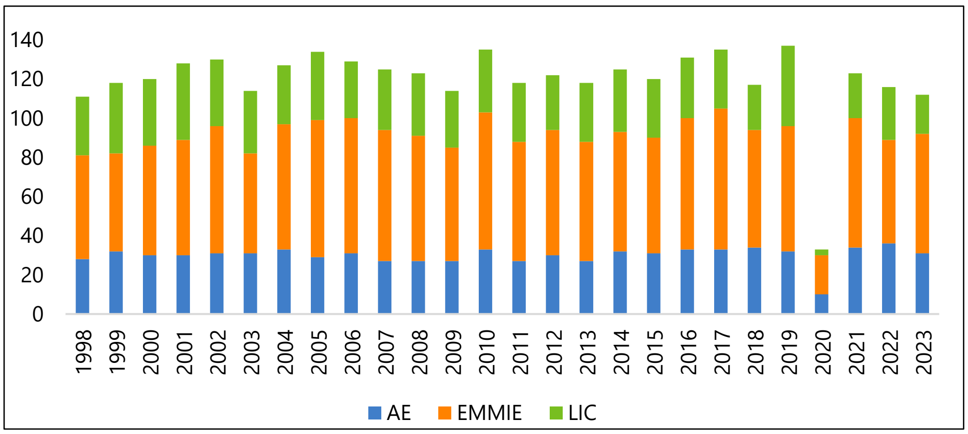
The paper is organized as follows. Section II outlines our methodology for extracting and classifying fiscal policy recommendations from Article IV reports, including our approach to validating the reliability of LLM-based analysis. Section III provides summary statistics on the Fund's fiscal, monetary, and macroprudential advice across income groups. Section IV presents the empirical framework linking fiscal advice for the near-term fiscal stance to economic conditions. Section V presents the main findings on the patterns and evolution of the Fund's advice, and Section VI concludes. An extensive appendix accompanies the paper, offering a practical guide for applying LLMs to the analysis of textual economic data, data description, and a wide range of robustness results.

II. METHODOLOGY AND DATA CONSTRUCTION

The corpus of Article IV consultation reports comprises 3,115 reports covering 193 countries, which we classify by income group—AEs, emerging market and middle-income economies

(EMMIEs), and low-income countries (LICs)—and by IMF area department.[2] On average, each country received about 16 Article IV consultations over the 26-year period, or roughly one every 18 months. Surveillance frequency varies by income groups: AEs undergo consultations every 1.3 years on average, compared with 1.6 years in EMMIEs and 2.0 years in LICs. Figure 1 shows that the number of published Article IV reports doesn't vary much from year to year, with typically 110 to 130 reports published each year. AEs account for a stable base of reports published each year, while most year-to-year variation stems from fluctuations in EMMIEs and LICs. The notable decline in 2020 corresponds to the temporary suspension of Article IV consultations during the COVID-19 pandemic, when IMF staff were redeployed toward crisis response and emergency financing operations (IEO, 2023).

Figure 1. Number of Article IV Reports
(1998–2023)



Sources: IMF Article IV Reports; IEO staff calculations.
Note: AE = advanced economy; EMMIE = emerging market and middle-income economy; LIC = low-income country.

To systematically classify fiscal policy advice, we employ a suite of LLMs—OpenAI's o1, Anthropic's Claude 3.7 Sonnet, and DeepSeek's R1.[3] These models analyze the Staff Appraisal sections of Article IV reports, which are extracted and pre-processed to remove formatting inconsistencies. Each model is prompted with a standardized instruction set designed to guide its reasoning process across three core dimensions of macroeconomic policy advice: the near- and medium-term fiscal stance, the monetary stance, and the macroprudential stance.[4] Each recommendation is categorized as Tighten, Neutral, Loosen, or Unclear (Box 1). The prompt instructs the models to reason explicitly before

[2] With support from the IMF's Strategy, Policy, and Review (SPR) Department, we assembled the full corpus of Article IV reports dating back to 1978. However, reports issued prior to 1998 have not yet been systematically processed and are therefore excluded from the present analysis.

[3] Appendix I presents the full prompt used for classification and provides sample outputs from the three LLMs for illustration. Appendix II offers a detailed practical guide to implementing the LLM-based methodology.

[4] IMF staff reports use the term "near-term" rather than "short-term." However, we use the two interchangeably throughout this paper. "Near-term" refers to the current economic cycle and typically covers the remainder of the year and the following year. "Medium-term" denotes a longer horizon, generally spanning two to five years.

assigning a classification and to produce a self-reported certainty score from 0 (low confidence) to 100 (high confidence).[5] These certainty scores enhance transparency and can be used as regression weights in subsequent analyses. The final classification for each country-year pair is determined through a majority-voting scheme. When at least two models agree, their joint outcome is retained; in the rare case of a three-way disagreement, the classification produced by OpenAI's o1 serves as the tiebreaker, reflecting its superior benchmark performance as of March 2025. The final certainty score equals the mean of the contributing models' confidence levels.[6]

Box 1. What is Meant by the Fiscal Stance in Article IV Reports?

Our analysis adopts a working definition of fiscal stance that aligns with standard practice in the economics literature. We interpret the stance as referring to the directional change in discretionary fiscal policy. Specifically, we classify IMF recommendations based on their implied movement, whether tightening, loosening, or maintaining course, relative to the prior year. This definition maps directly onto the most common empirical proxy for discretionary policy, the change in the cyclically adjusted primary balance (CAPB), as well as to "bottom-up" measures that aggregate the budgetary impact of new tax and spending measures (Romer and Romer 2010; Devries and others, 2011; Alesina and others, 2018; Carriere-Swallow, David, and Leigh, 2021; Adler and others, 2024).

Defining the fiscal stance in directional terms aligns our approach with the empirical literature on fiscal reaction functions and provides a clear basis for analyzing how IMF advice responds to macroeconomic and fiscal conditions. In empirical studies assessing the macroeconomic impact of fiscal consolidations or expansions, the fiscal stance is typically measured as year-on-year changes in the CAPB (Carnot, 2014; ECB, 2016). The same logic holds in the literature on fiscal reaction functions, which estimates how fiscal policy responds to economic developments (e.g., Bohn, 1998), and in theoretical models of optimal fiscal policy, which derive how the stance should evolve to balance output stabilization and fiscal sustainability objectives (Kanda, 2011; Fournier, 2019; Fournier and Lieberknecht, 2020). Across these applications, the focus is not on the level of fiscal support in a given year, but on whether policy is becoming more or less expansionary.

This understanding is also reflected in IMF guidance. IMF (2022) encourages staff to provide a comprehensive assessment of public finances, including a discussion of the fiscal policy stance along the cycle and its appropriateness. Separately, since 2018, the Fund's Consistent Policy Assessment (CPA) database has required country teams to indicate the intended direction of fiscal policy, whether it should be tightened, loosened, or maintained, on semiannual assessments. In practice, however, Article IV reports often use language that is ambiguous. In some cases, it is not clear whether staff refer to the level of fiscal indicators (e.g. the fiscal deficit or the CAPB) or the change in these indicators when referring to the fiscal stance (IMF, 1995). This can pose challenges for classification, particularly when relying on automated tools such as LLMs, which may find it difficult to infer whether staff are referring to the direction or level of fiscal policy support.

Some degree of misclassification is therefore inevitable. However, for the purposes of estimating a fiscal advice reaction function, which is our core empirical objective, it is essential to apply a consistent classification rule. We therefore interpret general language about fiscal tightening, consolidation, or loosening as referring to changes in discretionary policies. This choice ensures internal consistency with the structure of our empirical model, facilitates comparability across countries and time, and mirrors the approach we apply when classifying advice on monetary and macroprudential policy.

Source: Authors.

[5] Appendix I.D presents model confidence across policy domains and income groups. Median certainty scores are high—typically in 80s—indicating that the models classify policy advice with a consistently strong degree of confidence across both policy areas and country groups.

[6] For example, suppose one model classifies the stance as Neutral with 60 percent certainty (assigning only 5 percent to Tighten), a second model classifies Tighten with 50 percent certainty, and a third also classifies Tighten with 60 percent certainty. The majority vote therefore yields Tighten as the final classification. The corresponding certainty score is calculated as the mean of the three models' confidence levels for that category, namely (5 + 50 + 60)/3 in this case. When no majority emerges, the classification from the designated tiebreaker model is adopted, and its certainty score is computed using the same procedure.

We focus our analysis on the Staff Appraisal section of Staff Reports because it presents IMF staff's own views and policy advice, rather than describing past fiscal outcomes or the fiscal stance implied by the authorities' budgetary plans. According to IMF (2022), the staff appraisal section of Article IV reports must summarize staff's analysis, policy views, and recommendations and may address only issues discussed both with the authorities and in the main body of the report. It is also encouraged to integrate analyses of risks, spillovers, the overall policy mix, and discussion of policy trade-offs. Once the report is discussed by the Executive Board and published, the appraisal—together with the rest of the report—forms part of the public record, and the Board issues a press release reflecting Directors' endorsement of its main messages. The staff appraisal therefore provides the clearest and most coherent statement of IMF fiscal advice. It consolidates staff's overall assessment into a single, authoritative recommendation that reflects their policy judgment at the time of the consultation. In practice, appraisals typically include explicit assessments of both the near-term fiscal stance—covering the current and following year (t and t + 1)—and the medium-term stance (beyond t + 1), making them particularly well suited for systematic cross-country and time-series classification of IMF fiscal advice.

To validate the accuracy of these machine-generated classifications, we conducted a manual review of 105 randomly selected country-year pairs.[7] Each case was independently coded by the three co-authors based on the full Staff Appraisal text, with the benchmark classification defined by majority human agreement. The comparison between model outputs and manual benchmarks yields an 86 percent match rate for both near- and medium-term fiscal stance classifications, indicating high reliability.[8] Most discrepancies did not reflect clear errors but stemmed from interpretive differences regarding ambiguous or cautiously worded advice—for instance, whether to infer a stance from the overall tone of the staff appraisal or classify it as "Unclear" in the absence of a clearly articulated recommendation. In roughly one-third of disagreement cases, human coders labeled the text as "Unclear" while the LLMs provided a more definite judgment; in several others, differences arose between "Neutral" and adjacent categories. Classifying monetary policy advice proved more challenging, with a 66 percent match rate. This most likely reflects the fact that many countries belong to monetary unions or peg their exchange rates, thereby limiting

[7] The validation sample comprises 105 randomly selected country–year pairs, with some countries appearing more than once for different years. Within this full set, 27 percent of observations correspond to AEs, 47 percent to EMMIEs, and 27 percent to LICs. When restricted to 80 unique countries (excluding duplicates), the distribution shifts to 19 percent AEs, 35 percent EMMIEs, and 22 percent LICs. In terms of membership coverage, the sample includes 49 percent of AE countries, 38 percent of EMMIEs, and 40 percent of LICs. The sample also encompasses 10 small developing states. The results show that the short-term fiscal stance was 74 percent tighten, 10 percent neutral, 16 percent loosen, and 5 percent unclear, while the medium-term stance was 87 percent tighten, 10 percent neutral, none loosen, and 4 percent unclear.

[8] Metrics such as precision, recall, or F1 scores are appropriate for tasks with clearly defined and objective categories, for example, detecting spam emails. In contrast, fiscal stance classifications require interpretive judgment. The distinction between Neutral and Unclear, for instance, hinges on whether staff explicitly recommended maintaining the same stance or simply did not articulate a clear recommendation. Given this overlap, the most transparent and informative measure of model performance is the overall match rate with human coders.

monetary autonomy. In these cases, staff do not generally provide recommendations on the appropriate monetary stance.

III. DESCRIPTIVE EVIDENCE FROM THE CLASSIFICATION

This section summarizes how recommendations on fiscal, monetary, and macroprudential policies vary across income groups, over time, and examines the persistence of advice across consecutive Article IV consultations. Table 1 reports the distribution of classifications across policy areas, while Figures 2 and 3 illustrate the evolution of near-term and medium-term fiscal recommendations over time.

Table 1. LLM-Generated Classification of Recommended Macroeconomic Policy Stances
(In percent)

Policy stance	Income Group	Unclear	Loosen	Neutral	Tighten
Near-term fiscal	All	0	12	16	72
	AE	0	19	19	62
	EMMIE	0	11	13	75
	LIC	0	8	17	75
Medium-term fiscal	All	1	2	4	93
	AE	1	5	7	87
	EMMIE	0	1	3	96
	LIC	0	2	5	94
Monetary	All	23	8	37	32
	AE	36	11	33	19
	EMMIE	16	9	41	34
	LIC	24	3	32	41
Macroprudential	All		1	49	50
	AE		1	45	55
	EMMIE		1	48	51
	LIC		1	57	43

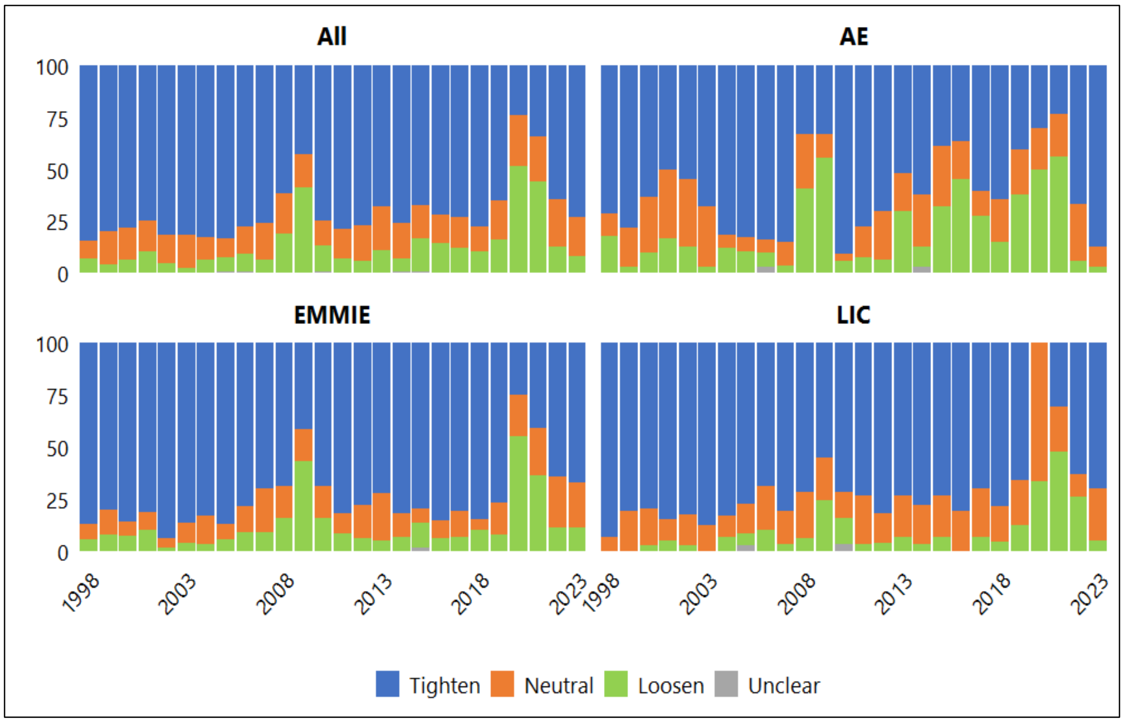
Sources: IMF Article IV Reports; IEO staff calculations.
Note: A substantial share of monetary policy advice is classified as "Unclear," primarily because many countries either participate in a monetary union or operate under fixed exchange rate regimes that constrain independent monetary policymaking. In such cases, IMF staff typically do not provide explicit recommendations on the appropriate monetary stance. Under the monetary stance classification, 23 percent of cases are coded as Unclear, comprising 13 percent with fixed exchange rate regimes, 2 percent with no data on the exchange rate regime, 1 percent with intermediate regimes, and 7 percent with flexible regimes. The horizon for monetary stance is near-term and unspecified for macroprudential stance. The "Neutral" column for the macroprudential stance is a combined "Neutral/Unclear." Macroprudential advice differs from fiscal and monetary guidance because it is often less explicit and more focused on monitoring or structural issues, making it difficult to distinguish a truly "Neutral" stance from "Unclear" cases. Combining these into one category ensures consistent interpretation of macroprudential recommendations across countries and reports.

A. Fiscal Advice: Near- and Medium-Term Stances

The Fund's near-term fiscal advice has been predominantly oriented toward fiscal consolidation. Across all country–year pairs from 1998 to 2023, 72 percent of recommendations call for fiscal tightening, 16 percent call for maintaining the current fiscal stance, and 12 percent advocate fiscal loosening. The emphasis on fiscal consolidation is strongest for EMMIEs and LICs, consistent with the perception that these countries face tighter fiscal constraints and higher debt vulnerabilities. The Fund's push for fiscal support to respond to the Global Financial Crisis (GFC) (2008–09) and COVID-19 pandemic (2020–21) is also evident from Figure 2.

Figure 2. Recommended Near-Term Fiscal Stance

(In percent)



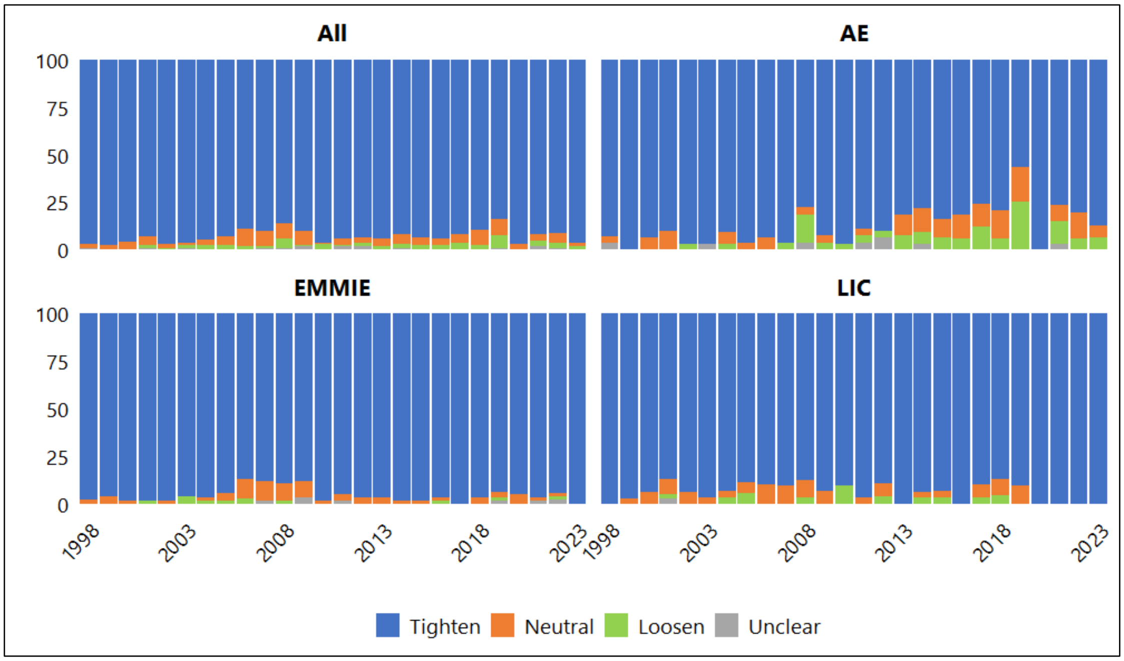
Sources: IMF Article IV Reports; IEO staff calculations.

Medium-term fiscal recommendations place an even stronger emphasis on fiscal consolidation. Across all observations, 93 percent of recommendations call for tightening the fiscal stance, 4 percent for maintaining it, and 2 percent for loosening.[9] Figure 3 nevertheless reveals a noticeable shift around 2013, when staff increasingly advised AEs to relax their medium-term fiscal plans. This evolution reflects the Fund's growing recognition of persistently low growth and interest rates, and its calls for higher public investment in countries with ample fiscal space

[9] The very small share of "Unclear" cases (1 percent) indicates that medium-term guidance is typically articulated with greater clarity than near-term advice. In the few instances where ambiguity is detected, it likely reflects the model's ability to capture genuine uncertainty or mixed signals in the text. Even at the 25th percentile, certainty scores are higher for the medium-term (83) than for the near-term (76).

(Cohen-Setton and Montiel, 2025). In these cases, relaxing medium-term targets was viewed as a way to support long-term growth and help reduce global imbalances.

Figure 3. Recommended Medium-Term Fiscal Stance
(In percent)



Sources: IMF Article IV Reports; IEO staff calculations.

B. Monetary and Macroprudential Advice and the Persistence of Recommendations

Table 2 reports how often the recommended policy stance changes between consecutive Article IV consultations. Near-term fiscal stance recommendations change between consecutive Article IVs in roughly one-third of cases (32 percent), with greater volatility in AEs (39 percent) than in EMMIEs or LICs (30 percent). Medium-term fiscal advice is far more stable, changing in only 8 percent of observations. Monetary advice shows moderate persistence (29 percent), while macroprudential guidance is the most variable (34 percent). This pattern suggests that fiscal advice, especially at the medium-term horizon, is anchored in structural assessments of fiscal sustainability, whereas the monetary, macroprudential, and near-term fiscal recommendations are recalibrated more frequently in response to evolving macroeconomic conditions.

IMF advice on monetary policy is mostly split between tightening and maintaining the stance, while macroprudential advice is largely focused on tightening. Monetary policy advice is relatively evenly distributed, with 32 percent of recommendations calling for tightening, 37 percent for maintaining the stance, and 8 percent for loosening. The large share of cases classified as Unclear reflects that many countries do not set monetary policy independently, either because they are members of a currency union or maintain a hard exchange rate peg, leaving limited scope to adjust the stance. By contrast, macroprudential guidance is far more one-sided: about half of all observations recommend tightening, 49 percent are neutral or unclear, and only 1 percent advocate loosening. Most changes in this category reflect reclassifications rather than substantive shifts in policy direction. In fact, 96 percent of the 34 percent of observed changes represent transitions between "Unclear/Neutral" and "Tighten," rather than reversals between loosening and tightening.

Table 2. Frequency of Changes in Policy Advice Between Consecutive Article IV Reports
(In percent)

	All	AE	EMMIE	LIC
Fiscal near-term stance	32	39	30	30
Fiscal medium-term stance	8	12	5	10
Monetary stance	29	26	31	26
Macroprudential stance	34	33	35	33

Sources: IMF Article IV Reports; IEO staff calculations.

IV. DETERMINANTS OF NEAR-TERM FISCAL POLICY ADVICE

This section outlines the empirical framework used to explain variation in the IMF's near-term fiscal policy recommendations. We interpret the Fund's advice as a fiscal reaction function—that is, as a systematic relationship between economic conditions and the recommended fiscal stance—grounded in the trade-off between supporting economic activity and maintaining fiscal sustainability. Subsection A reviews the theoretical foundations of this framework, drawing on models that formalize how policymakers balance stabilization and debt sustainability objectives. Subsection B then sets out the empirical specification used to estimate this relationship and details the data sources and variables constructed for the analysis.

A. Theoretical Foundations

The choice of the fiscal stance reflects a fundamental trade-off between supporting economic activity and preserving fiscal sustainability. Bianchi, Ottonello, and Presno (2023) formalize this tension in a dynamic model that combines Keynesian stabilization motives with sovereign risk. In their framework, higher public spending during recessions reduces unemployment but increases borrowing costs and default risk, thereby tightening future fiscal constraints. The optimal policy response is therefore state-dependent: when debt levels are low (Keynesian regime), governments pursue countercyclical fiscal expansions to stabilize output; when debt is high, concerns about market access dominate, prompting procyclical austerity to contain risk premia and preserve solvency. This framework captures the central dilemma facing policymakers—whether to stimulate demand at the risk of undermining debt sustainability—and provides a useful normative benchmark for interpreting the IMF's fiscal advice.

Although integrated models combining these elements are recent, the underlying mechanisms have long been recognized in the literature. The Keynesian and New Keynesian traditions emphasized the gains from countercyclical fiscal policy when monetary policy is constrained—

whether by fixed exchange rates, the zero lower bound, or financial frictions (Christiano, Eichenbaum, and Rebelo, 2011; Farhi and Werning, 2016; Eggertsson, 2011; Woodford, 2011).[10] In contrast, models of sovereign default (Eaton and Gersovitz, 1981; Aguiar and Gopinath, 2006; Arellano, 2008) underscored the costs and risks of additional borrowing, showing how rising debt or deteriorating market confidence can force governments to tighten policy even in recessions. Related work also noted that liquidity shocks can trigger loss of market access even for solvent sovereigns (Cole and Kehoe, 2000). Together, these insights established the conceptual basis for understanding the tension between fiscal stabilization and debt sustainability later formalized in integrated models.

A parallel body of applied work has translated these theoretical insights into operational tools for fiscal surveillance. Models by Kanda (2011), Fournier (2019), and Fournier and Lieberknecht (2020) specify welfare-based objective functions in which governments minimize deviations of output and debt from desired paths, yielding prescriptions for the appropriate adjustment in the structural primary balance given cyclical conditions and debt levels. Institutional frameworks such as those used by the European Commission, the ECB, and national finance ministries (Carnot 2014; European Commission 2015; ECB 2016; Bankowsky and Ferdinandusse 2017) build on similar principles but rely on fixed weights for stabilization and sustainability objectives.[11] While less grounded in micro foundations, these tools have proved valuable for real-time fiscal assessments and for communicating the rationale of fiscal stance recommendations.

Our empirical analysis extends the logic of fiscal reaction function to the Fund's own advice. We estimate an IMF-specific fiscal reaction function linking the recommended stance to cyclical conditions and fiscal vulnerabilities, drawing inspiration from the empirical literature that relates actual policy behavior to these variables (Bohn 1998; Kaminsky, Reinhart, and Végh 2004; Eyraud and others, 2018). Conceptually, we interpret Article IV recommendations as a normative fiscal reaction function, reflecting how staff weigh stabilization needs against sustainability risks. When output gaps are positive or fiscal vulnerabilities are elevated, the Fund's advice should tilt toward tightening; when slack is large and debt risks low, a looser stance should be favored. The next section formalizes this relationship empirically, providing a quantitative counterpart to the theoretical frameworks summarized above.[12]

[10] Countercyclical deficits are also optimal in the tax smoothing literature. In Barro's (1979) framework, governments borrow during downturns to avoid sharp increases in distortionary taxes, implying countercyclical fiscal policy even in the absence of nominal rigidities or unemployment.

[11] For example, the European Commission assesses fiscal plans by quantifying reductions in two indicators: the stabilization gap, defined as the change in the structural primary balance (SPB) consistent with closing 25 percent to 50 percent of the output gap in one year, and the sustainability gap, measured by the cumulative SPB adjustment needed to bring debt to 60 percent of GDP over the next 15 years.

[12] Preliminary evidence supports the theoretical expectation that IMF staff recommend fiscal tightening more often when fiscal vulnerabilities are high and output gaps are positive, indicating a countercyclical stance. This

B. Empirical Specification and Data

To analyze how IMF fiscal advice reflects the trade-off between output stabilization and fiscal sustainability, we use an ordered logit framework linking the recommended fiscal stance to key macro-fiscal variables. The dependent variable is the LLM-based classification of near-term fiscal advice (Loosen, Neutral, or Tighten), which is ordinal by construction. This approach provides a intuitive framework for modeling the probability of tighter or looser advice as a function of cyclical conditions, fiscal indicators, and institutional controls, while preserving the ranking structure of the outcome. The baseline specification assumes that the recommended fiscal stance reflects a combination of stabilization needs, fiscal sustainability pressures, and other relevant factors:

$$\boldsymbol{y^*_{i,t} = \beta Output\ Gap_{i,t} + \gamma FB\ gap_{i,t-1} + \phi Debt_{i,t-1} + \eta X_{i,t-1} + \varepsilon_{i,t}}\ ,$$

where $y^*_{i,t}$ is a latent continuous variable representing the underlying fiscal stance recommended by the IMF for country i in year t.[13] A positive coefficient (β) indicates countercyclical advice—i.e., more expansionary recommendations when economic slack is high (output gap is negative). The $FB\ gap_{i,t-1}$ measures the fiscal adjustment required to stabilize the debt ratio, and $Debt_{i,t-1}$ is the debt-to-GDP ratio. $X_{i,t-1}$ denotes a vector of additional economic or institutional variables that may influence IMF fiscal advice beyond stabilization and sustainability considerations. The error term $\varepsilon_{i,t}$ captures unobserved factors, and in extended specifications, is decomposed as $\varepsilon_{i,t} = \alpha_i + \lambda_t + u_{i,t}$ to account for country fixed effects and time-specific global developments.

Translating these theoretical dimensions into measurable indicators requires careful construction of variables capturing stabilization needs and sustainability pressures. Stabilization needs are proxied by the output gap, while sustainability risks are represented by the debt-to-GDP ratio and the fiscal balance (FB) gap. To enhance coverage and robustness, WEO output gaps are complemented with alternative estimates derived from the Hamilton filter, following Baum and others (2017), Eyraud and others (2018), and Jalles and others (2024). Debt data are taken directly from the WEO, while the FB gap is defined as the difference between the balance required to stabilize debt and the actual balance—so that a positive gap signals the need for fiscal tightening. This measure closely corresponds to the "sustainability gaps" used by international institutions and recent empirical research (Kose and others, 2022; European Commission, 2020; Saxegaard and others, 2012) to quantify the adjustment necessary for stabilizing debt. Observations in the 1st and 99th percentiles of each variable's distribution are excluded to limit

---

pattern is strongest among AEs, less systematic for LICs, and broadly consistent with the FRF framework. See Table AIII.2. in the Appendix for details.

[13] In an ordered logit framework, the observed categorical outcome (Loosen, Neutral, Tighten) is assumed to derive from an unobserved continuous variable, denoted y^*, which represents the underlying intensity of the recommended near-term fiscal stance. This latent variable captures the continuous propensity of IMF staff to advise fiscal tightening or loosening. The observed categories correspond to specific intervals of y^*, separated by threshold parameters that define the cutoffs between Loosen, Neutral, and Tighten recommendations.

the influence of outliers. In specifications with country fixed effects, countries with invariant fiscal stance classifications are dropped to avoid collinearity, though they are retained in pooled regressions without fixed effects.[14] All variables sourced from the World Economic Outlook (WEO) are drawn from the vintage released closest to the date of each Article IV consultation, so that the dataset reflects the economic information set available to IMF staff when formulating their advice.[15]

V. Main Results

A. Baseline Results

Estimation results show that IMF fiscal advice responds systematically to macroeconomic and fiscal conditions, with recommendations becoming more expansionary in periods of economic slack and more restrictive when debt vulnerabilities are elevated. The model links the recommended near-term fiscal stance to three key indicators—the output gap, the overall FB gap, and the debt-to-GDP ratio—which capture cyclical conditions, fiscal adjustment needs, and solvency risks, respectively.[16] All variables are expressed as ratios to GDP or potential GDP, and each observation is weighted by the LLM's classification confidence to give greater influence to recommendations made with higher certainty. Table 3 reports results from four specifications—without fixed effects, with country effects, with year effects, and with both. Across all models, IMF

[14] The dependent variable, the recommended short-term fiscal stance, is an ordinal categorical variable. Estimation instability can arise when the recommended stance remains constant within a country over time—particularly in specifications with country fixed effects—since such cases provide no within-country variation. Countries for which the recommended short-term fiscal stance does not vary over the sample period are Djibouti, Eritrea, Gabon, Ghana, Hungary, Iran, Iraq, Jamaica Lebanon, Malawi, Maldives, Mongolia, Montenegro, South Sudan, Sri Lanka, and Venezuela. In addition, the few cases (five in total) where the majority-voted classification yielded an "Unclear" recommendation are excluded from the estimation sample.

[15] Details on data sources and variable definitions are in Appendix III. Summary statistics for key explanatory variables and the construction of the fiscal balance (FB) gap are also presented. When no contemporaneous WEO vintage is available, we used the 2024 October WEO to maximize coverage, applying its values retrospectively.

[16] Because ordered logit models estimate log-odds rather than linear predictions, model fit is assessed using pseudo-R^2 metrics. McFadden's pseudo-R^2 captures the improvement in log-likelihood relative to a model with no explanatory variables, while Ugba and Gertheiss (2023) propose an ordinal-specific refinement that adjusts the number of outcome categories. McFadden's measure is defined as $R^2_{mf} = 1 - l_p/l_0$, where l_p is the maximum log-likelihood of the full model with p predictors and l_0 is the log-likelihood of the intercept-only model. The Ugba and Gertheiss pseudo R^2 is given by $R^2_J = 1 - \gamma_J^{\lambda(J)}$, where $\gamma_J = l_p(J)/l_0(J)$ and $\lambda(J) = \sqrt{2J}$ is a strictly positive penalty function that depends on the number of ordered categories J.

staff recommendations exhibit clear and statistically significant responsiveness to macro-fiscal conditions.[17]

To interpret the estimated relationships in more intuitive and policy-relevant terms, we complement the coefficient estimates with measures of average marginal effects and predicted probabilities**.** Average marginal effects (AMEs) express the estimated log-odds as the average change in the probability of fiscal tightening recommendations, summarizing each variable's influence on the likelihood of fiscal tightening. Because the ordered logit model is nonlinear, these effects vary across the distribution of covariates, and the AME captures their average impact. The estimates indicate that a one-percentage-point increase in the output gap, FB gap, and debt-to-GDP ratio raises the probability of tightening recommendation by approximately 1.0, 0.3, and 0.2 percentage points, respectively (Table 3). Figure 4 visualizes these relationships, showing how the predicted probabilities of loosening, neutral, and tightening advice evolve as each covariate increases while the others are held constant at their means. The upward slope of the tightening curve across all panels, together with the corresponding decline in neutral and loosening probabilities, underscores that IMF staff systematically weigh both cyclical conditions and fiscal sustainability risks when formulating near-term fiscal advice.

[17] Section IV of the Appendix reports robustness checks for the baseline model. First, it re-estimates the regressions using WEO covariates from a fixed vintage rather than the vintage preceding each Article IV consultation. Second, it replaces the dependent variable with the individual LLM classifications of the near-term fiscal stance, instead of the majority-voted outcome. The results remain qualitatively unchanged across both exercises.

Table 3. Baseline Regression: Determinants of Recommended Near-Term Fiscal Stance

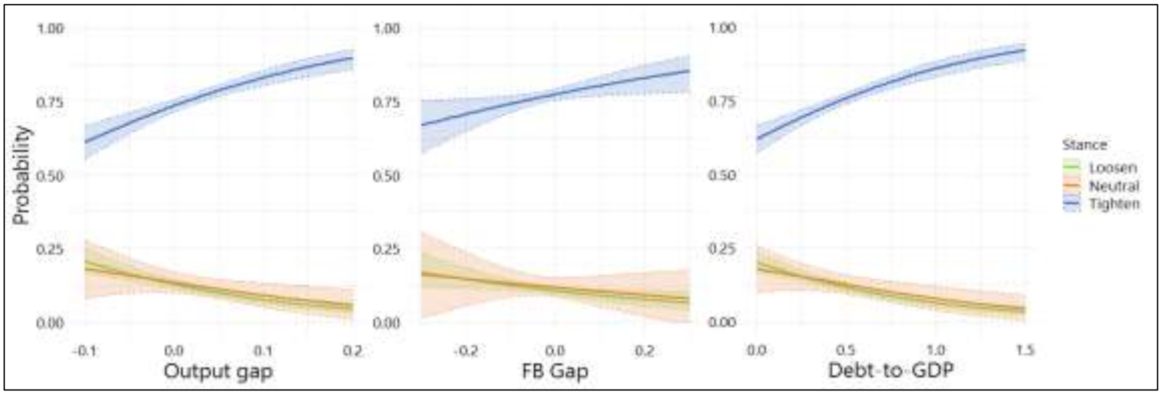
Variables	(1)		(2)		(3)		(4)	
	AME	$\beta(\mathrm{se}_\beta)$	AME	$\beta(\mathrm{se}_\beta)$	AME	$\beta(\mathrm{se}_\beta)$	AME	$\beta(\mathrm{se}_\beta)$
Output gap	1.022	5.671***	1.610	10.448***	1.015	5.689***	0.941	7.001***
		(0.951)		(1.797)		(0.956)		(1.996)
FB gap	0.315	1.752**	0.536	3.482***	0.334	1.873**	1.096	8.149***
		(0.726)		(1.066)		(0.730)		(1.316)
Debt-to-GDP	0.235	1.305***	0.351	2.279***	0.237	1.328***	0.433	3.220***
		(0.182)		(0.377)		(0.183)		(0.459)
τ_1		0.362***		2.399		0.626**		2.548
		(0.104)		(1.400)		(0.287)		(1.451)
τ_2		1.251***		3.494**		1.521***		3.828***
		(0.110)		(1.401)		(0.289)		(1.453)
N	2720		2556		2720		2556	
Country FE	No		Yes		No		Yes	
Year FE	No		No		Yes		Yes	
R^2 (McFadden)	0.145		0.162		0.150		0.248	
R^2 (Ugba & Gertheiss)	0.318		0.352		0.329		0.504	
Proportional odds	No		No		No		No	

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The average marginal effect (AME) columns report the average expected change in the predicted probability of tightening advice given one percentage point increase in the corresponding explanatory variable. As a measure of model fit, we report pseudo R^2 of McFadden and Ugba and Gertheiss. We also test if the proportional odds assumption holds with Brant test.

The model identifies distinct thresholds, denoted τ_1 and τ_2, separating loosening, neutral, and tightening advice, confirming that the underlying ordinal structure captures meaningful shifts in IMF fiscal recommendations. These thresholds—analogous to cutpoints on the latent scale—indicate where the model transitions between outcome categories. Specifically, τ_1 separates the "Loosen" category from the combined "Neutral" and "Tighten" categories, while τ_2 delineates the boundary between "Loosen" and "Neutral" versus "Tighten." Unlike intercepts in linear regressions, they are not interpreted directly but serve to partition the unobserved continuum of fiscal stance recommendations. A Brant test rejects the proportional-odds assumption for the overall model and for key variables, indicating that the effects of explanatory factors are not uniform across thresholds. In particular, the influence of the output gap is stronger when moving away from a loosening recommendation than when shifting from a neutral or tightening stance, whereas the effects of the debt ratio and FB gap remain relatively stable across categories. Alternative specifications—including binary logit, generalized ordered logit, and linear models—produce consistent results, confirming the robustness of the findings. For clarity and comparability, the ordered logit specification is retained as the baseline.[18]

[18] Section IV of the Appendix reports results for these alternative specifications.

Figure 4. Predicted Probabilities of Near-Term Fiscal Stance Recommendations
(Decimals)



Source: IEO staff calculations.
Note: The lines show the predicted probabilities of each fiscal stance given macroeconomic conditions, with shaded areas representing 95 percent confidence intervals. The figure is based on the estimates in Table 3 column (1). Covariates are presented in decimal.

IMF fiscal advice on the near-term stance is countercyclical, tightening as economic slack narrows and easing when activity falls below potential**.** The baseline estimates show that as the output gap increases, the probability of a tightening recommendation rises (Table 3; Figure 4). Interpreted through average marginal effects, a 1 percentage point increase in the output gap raises the likelihood of a tightening call by about 1 percentage point, with corresponding declines in neutrality or loosening. This pattern holds across alternative real time measures of cyclical conditions—including WEO-only output gaps, Hamilton-filter estimates, deviations of current growth from its five-year average, and deviations from the five-years-ahead WEO growth projection—all of which yield positive and statistically significant coefficients of economically meaningful magnitude (Appendix Table AV.1; Figure AV.1). Taken together, these results confirm that IMF staff systematically lean against the cycle when formulating near-term fiscal advice.

IMF fiscal advice also responds systematically to fiscal vulnerabilities, tightening as solvency and liquidity risks intensify. Both the FB gap—the adjustment needed to stabilize debt—and the debt-to-GDP ratio are positively and significantly associated with tightening recommendations. These messages remain robust when using broader sustainability indicators—such as long-run "sustainability gaps" and alternative debt stock measures à la Kose and others (2022)—and when extending the analysis to liquidity and market-access pressures (gross financing needs, borrowing costs, ratings, CDS spreads), balance-sheet exposures (nonresident holdings, foreign-currency debt shares, maturity), and external vulnerabilities (current account deficits, private external debt). In short, IMF staff advice becomes more conservative as fiscal risks mount or market confidence weakens. Full robustness checks and results by income group are reported in the Appendix.[19]

B. Heterogeneity by Income Group

The responsiveness of IMF fiscal advice to macro-fiscal conditions varies systematically across income groups. This heterogeneity is expected, as countries differ in institutional capacity, policy credibility, and exposure to external constraints. AEs typically possess stronger fiscal frameworks, deeper financial markets, and more reliable real time data, which enhance their ability to conduct countercyclical policy. By contrast, EMMIEs and LICs often face tighter financing constraints, greater output volatility, and weaker capacity to assess and implement discretionary measures. To capture these structural differences, we interact income group dummies (AEs, EMMIEs, and LICs) with the key explanatory variables.

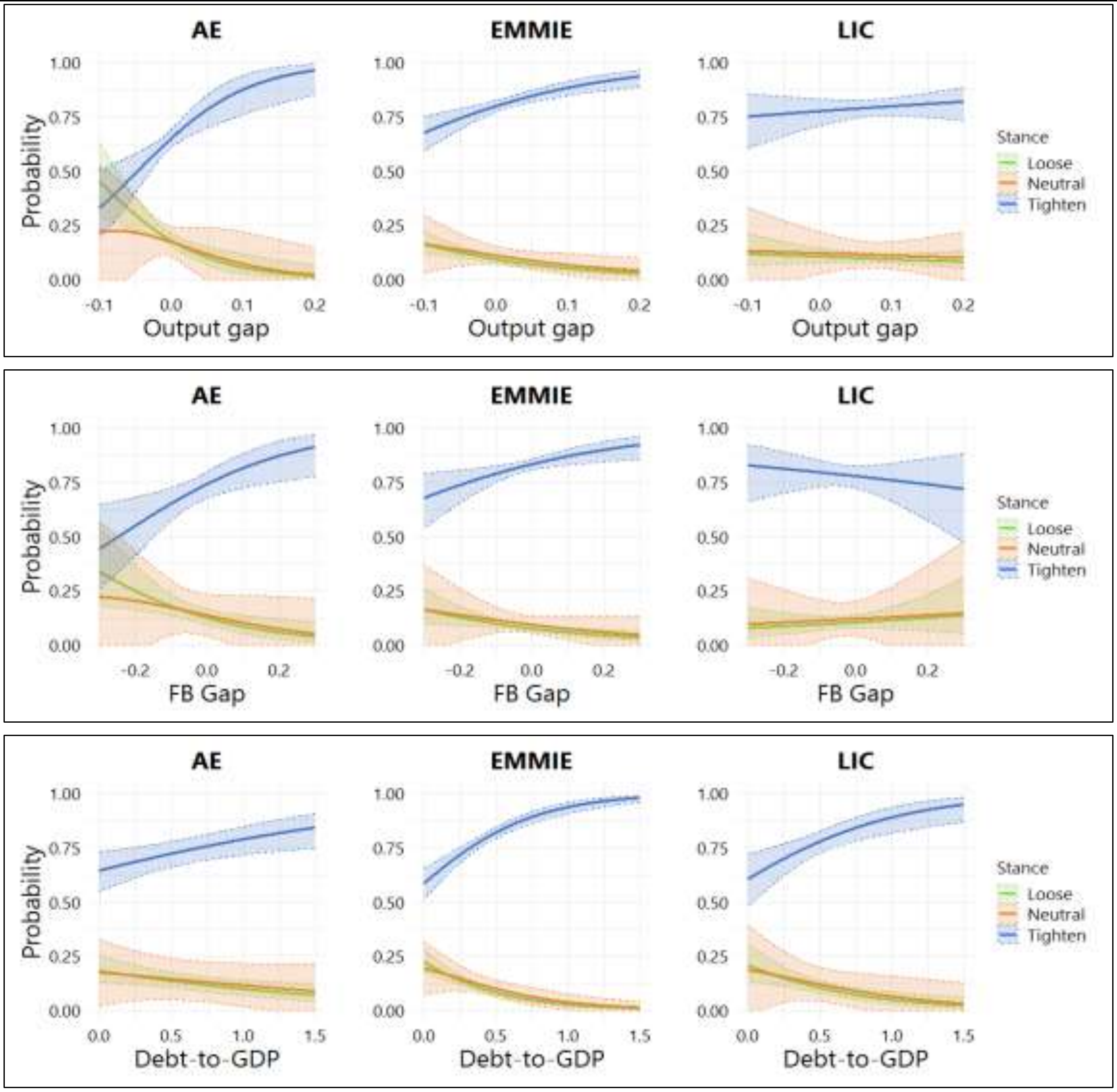
IMF fiscal advice tends to be more countercyclical in AEs, with the strength of the response diminishing at lower income levels**.** Figure 5 shows that staff recommendations for AEs display a strong and consistent sensitivity to cyclical conditions: the coefficient on the output gap is large and highly significant, implying a clear tendency to recommend tightening when output exceeds potential and loosening when economic slack prevails. For EMMIEs, the output gap remains a significant predictor, but the coefficients are smaller, suggesting that fiscal advice is only moderately countercyclical. This can be seen graphically by the flatter slope. In LICs, by contrast, the estimated coefficients are small and generally insignificant, pointing to a weaker link between cyclical conditions and advice. This attenuation likely reflects both greater uncertainty in measuring potential output and limited fiscal or institutional capacity to implement discretionary countercyclical policies.[20] Overall, these results indicate that IMF fiscal advice aligns most closely with macroeconomic stabilization objectives in AEs, while structural and operational constraints reduce the scope for countercyclical advice in lower-income settings.

Figure 5. Predicted Probabilities of Near-Term Fiscal Stance Recommendations by Income Group

(Decimals)

[19] Section V.B. of the Appendix broadens the analysis of fiscal vulnerability beyond traditional solvency indicators by incorporating the IMF's Vulnerability Exercise and complementary measures from Kose and others (2022). These indicators capture solvency, liquidity, balance sheet, and external vulnerabilities. Together, they provide a multidimensional view of fiscal risks and help assess how IMF fiscal stance advice varies across different vulnerability measures.

[20] For countries without WEO-provided output gap estimates, output gaps were derived using the Hamilton filter. Overall, 34.7 percent of all observations come from these Hamilton-based estimates. These Hamilton-filtered values account for 30 percent of observations among EMMIEs and 84 percent among LICs.



Source: IEO staff calculations.
Note: The lines show the predicted probabilities of each fiscal stance given macroeconomic conditions, with shaded areas representing 95 percent confidence intervals. Figures are based on the estimates in Table AVI.1. column (1) of the Appendix. Covariates are presented in decimal.

C. Heterogeneity Over Time

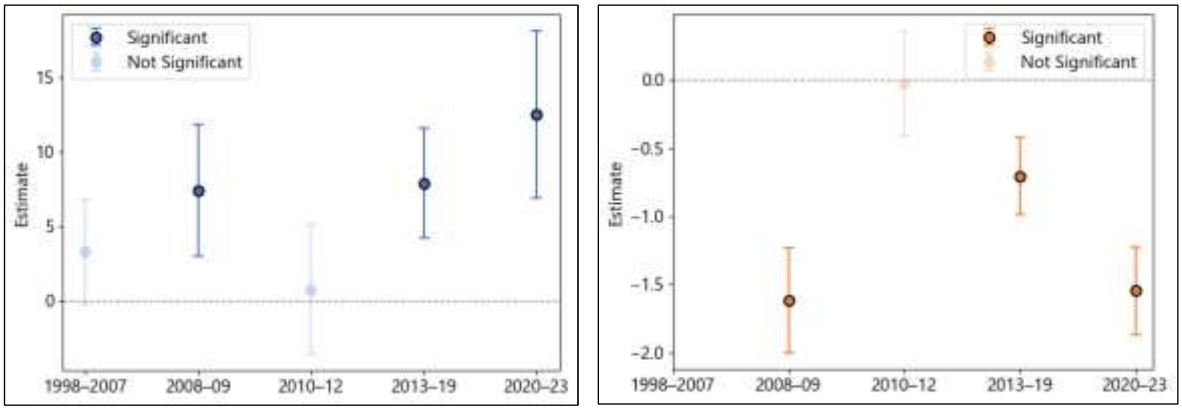
Since the GFC, the IMF's approach to fiscal policy has evolved toward a more explicitly countercyclical orientation, reflecting both a transformed macroeconomic environment and a reassessment of fiscal trade-offs. As noted by Cohen-Setton and Montiel (2025), the combination of exceptional economic slack, disinflationary pressures, and persistently low interest rates shifted the policy debate from fiscal retrenchment toward active demand support. These conditions reduced earlier concerns about crowding out and debt sustainability, while new empirical evidence on fiscal multipliers reinforced the case for fiscal stimulus when monetary policy was constrained. Without abandoning its focus on fiscal prudence, the Fund increasingly advocated expansionary fiscal measures—particularly public investment—anchored in credible medium-term consolidation plans and supported by stronger fiscal institutions to preserve confidence while sustaining demand.

Empirical evidence confirms that this intellectual reorientation translated into increasingly countercyclical bilateral fiscal advice in practice. Results from the extended ordered-logit analysis show that, since the GFC, IMF staff have become more likely to recommend fiscal tightening when output is above potential and more cautious about consolidation when economic slack persists (Figure 6). The negative and significant time trend points to a gradual shift away from the Fund's earlier default presumption of consolidation, while the interaction between the output gap and the post-2009 dummy indicates that fiscal advice has become more sensitive to cyclical conditions. When the sample is divided into multiyear subperiods, the coefficient on the output gap rises steadily, showing that fiscal advice has become progressively more responsive to measures of economic slack. The left panel of Figure 6 illustrates this evolution, with the magnitude of the output-gap coefficient increasing from the early 2000s to the most recent period, while the right panel shows a parallel decline in the baseline probability of recommending tightening—particularly after 2015—consistent with the more accommodative stance adopted in a low-interest-rate environment. Taken together, these results indicate that the Fund's evolving view of fiscal policy—emphasizing its stabilization role alongside sustainability—has been reflected not only in analytical frameworks and policy papers but also in the day-to-day advice delivered through bilateral surveillance.[21]

[21] Detailed estimates corresponding to these results are reported in Table AVI.2. of the Appendix.

Figure 6. Evolution of the Countercyclicality of IMF Fiscal Advice

A. Output Gap Interacted with Year Buckets | **B. Year Bucket Intercepts**

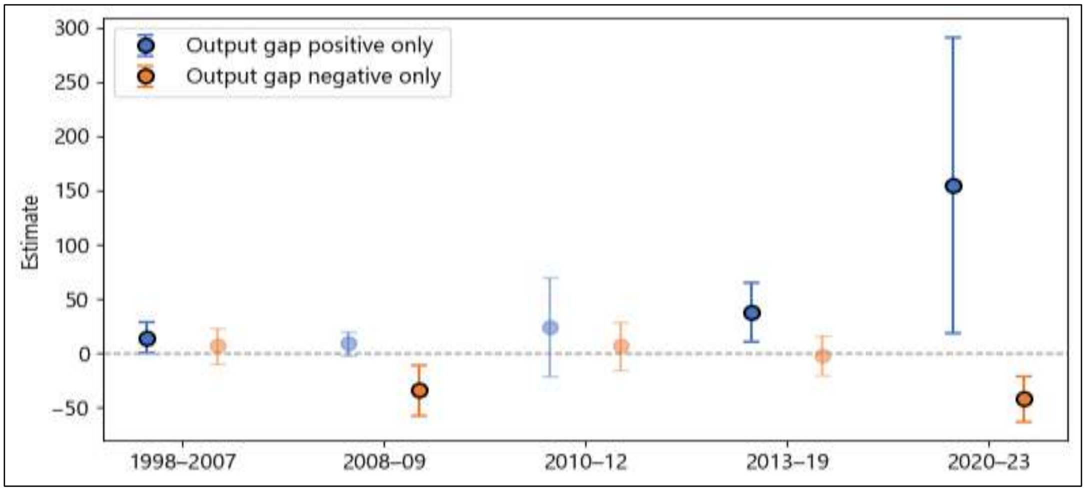


Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: The figure is based on the estimates in Table AVI.2, column (3) of the Online Appendix. It shows coefficient estimates from an ordered logit model where the output gap is interacted with period dummies. Each point represents the estimated coefficient for a given subperiod, and vertical lines indicate 95 percent confidence intervals. Panel A displays how the responsiveness of fiscal stance advice to the output gap has changed over time, while Panel B shows the baseline effect for each period, capturing shifts in the underlying likelihood of recommending fiscal tightening. Brighter colors denote statistically significant estimates; lighter shades indicate insignificant results. All regressions use contemporaneous data.

IMF fiscal advice has remained consistently countercyclical when economies operate above potential but more differentiated when they operate below it. As shown in Figure 7, staff typically recommend tightening in the presence of positive output.[22] When output gaps are negative, however, the relationship is less uniform. Fiscal advice was strongly expansionary during the two major periods of widespread slack—the GFC (2008–09) and the COVID-19 pandemic (2020–23)—but less systematically accommodative in the intervening years. In the immediate post-crisis period (2010–12), the point estimate for negative output gaps is positive but statistically insignificant, suggesting that fiscal advice was less countercyclical than in other post-GFC subperiods. Between 2013 and 2019, recommendations became more clearly differentiated across countries: staff encouraged those with fiscal space to maintain or expand support while advising consolidation in more constrained cases. This evolution is consistent with the pattern documented by Cohen-Setton and Montiel (2025), underscoring the Fund's growing emphasis on tailoring near-term fiscal recommendations to country-specific fiscal space.

[22] We estimate the following specification: $y^*_{i,t} = \sum_{k\in K} I_{k,t}\{\delta_k + Output\ Gap_{i,t}[\beta_{1,k}D_{i,t} + \beta_{2,k}(1-D_{i,t})]\} + \gamma FB\ gap_{i,t-1} + \phi Debt_{i,t-1} + \varepsilon_{i,t}$, where D is an indicator variable that takes value 1 when the output gap is positive and 0 otherwise, $I_{k,t}$ is a dummy variable for year bucket k, and δ_k is the main effect of the year bucket k.

Figure 7. Evolution of Asymmetric Responsiveness of Near-Term Fiscal Advice



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: The figure shows how the asymmetry of fiscal advice with respect to the output gap varies over time. Each point represents the estimated relationship between the output gap and the likelihood of recommending fiscal tightening, conditional on whether the output gap was positive or negative. The analysis relies exclusively on output gap estimates from the WEO, without incorporating Hamilton-filtered values. Vertical lines indicate 95 percent confidence intervals. "Output gap positive only" (blue) refers to cases where the economy was operating above potential, while "Output gap negative only" (orange) captures cases below potential. Brighter colors indicate statistically significant estimates, while lighter shades represent insignificant results. The estimates are grouped into multi-year intervals to highlight changes across key periods.

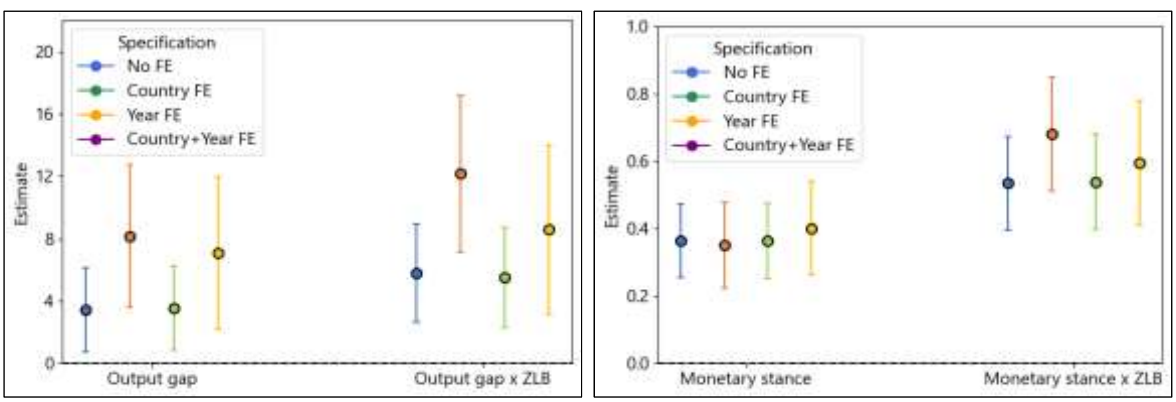
D. Macroeconomic Policy Mix

A central question in assessing IMF fiscal advice is how it relates to recommendations on other macroeconomic instruments. The Fund's advice on fiscal policy rarely operates in isolation: effective macroeconomic management typically requires coordination across fiscal, monetary, and macroprudential levers. Theoretical and empirical research highlights that the optimal configuration of this policy mix depends on the nature of the underlying shock. During negative demand shocks, fiscal and monetary policies are typically complementary—both easing to support demand (Woodford, 2011; Christiano, Eichenbaum, and Rebelo, 2011)—especially when monetary policy is constrained by the zero lower bound (ZLB). Outside the ZLB, when monetary authorities retain room to adjust rates, fiscal policy is often advised to remain neutral or consolidate, particularly in high-debt contexts. By contrast, supply shocks can generate trade-offs between inflation and output stabilization, with fiscal loosening offsetting monetary tightening, implying a substitutive relationship (Clarida, Galí, and Gertler, 1999; Gali, 2008). The interaction between macroprudential and other policy levers remains less codified: macroprudential tools may complement monetary policy in containing financial imbalances or act as substitutes when interest rates cannot be adjusted freely (Korinek and Simsek, 2016; Jeanne and Korinek, 2013; IMF, 2013).

Empirically, IMF fiscal advice appears closely coordinated with monetary policy, particularly when monetary space is constrained. To test this, we extend our baseline ordered logit specification by adding indicators of the Fund's advice on monetary and macroprudential stances as explanatory variables. These indicators take the value of 1 for easing, 3 for maintaining the current stance, and 5 for tightening, and monetary stance is interacted with a ZLB dummy to assess how fiscal advice adjusts when conventional monetary tools are constrained.[23] The results reveal a clear pattern of complementarity between fiscal and monetary advice over the evaluation period. Across all specifications, tighter monetary recommendations are associated with tighter fiscal advice, with the coefficient on the monetary stance positive and highly significant. This complementarity strengthens when monetary policy is constrained: the interaction term between the monetary stance and the ZLB is also positive and larger in magnitude, indicating that fiscal and monetary advice become more closely aligned when policy rates approach zero (Figure 8, Panel B). In addition, the interaction between the output gap and the ZLB is positive and significant, suggesting that fiscal advice becomes more sensitive to cyclical conditions when monetary policy is less effective (Figure 8, Panel A). Finally, the negative and significant main effect of the ZLB supports this interpretation—fiscal advice is generally more accommodative when interest rates are constrained, consistent with the view that fiscal policy should play a greater stabilizing role under such circumstances.[24]

Figure 8. Fiscal Countercyclicality and Policy Complementarity at the ZLB

A. Fiscal Countercyclicality | **B. Fiscal-Monetary Policy Complementarity**



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Panel A shows how the responsiveness of IMF near-term fiscal advice to the output gap changes when monetary policy is constrained by the ZLB, while Panel B illustrates the relationship between fiscal and monetary policy advice. Vertical lines indicate 95 percent confidence intervals. Detailed estimates are provided in Table AVI.3. of the Appendix.

[23] The results remain virtually unchanged when these variables are recoded as binary indicators equal to 1 when IMF staff recommend policy tightening.

[24] The years 2009–15 and 2020–21 are classified as ZLB periods, corresponding to episodes when the U.S. federal funds rate was at or near the effective lower bound. Detailed estimates for these results are reported in Table AVI.3. of the Appendix.

By contrast, fiscal advice shows no systematic relationship with contemporaneous macroprudential recommendations. Across specifications, the coefficient on the macroprudential stance is small and statistically insignificant, suggesting that the Fund does not treat fiscal and macroprudential policies as jointly calibrated instruments. This may reflect the more indirect transmission channels of macroprudential tools, the absence of established analytical frameworks linking them to aggregate demand, or the Fund's tendency to view them primarily through a financial-stability lens rather than a cyclical one. Taken together, these results indicate that IMF fiscal advice tends to be coordinated with monetary policy—particularly when monetary space is limited—but remains largely independent of macroprudential considerations.[25]

E. External Imbalances

IMF fiscal advice generally responds to external imbalances, even after controlling for domestic fiscal conditions. To test whether external positions influence fiscal recommendations independently of traditional fiscal indicators, we augment the baseline model with the current account balance as a share of GDP. The rationale is straightforward: fiscal imbalances contribute to current account imbalances through their impact on saving–investment dynamics, and vice versa. If the Fund considers these broader macroeconomic linkages, the current account should influence the recommended fiscal stance. The results strongly support this hypothesis. Across specifications (Figure 9), the coefficient on the current account deficit is positive, large, and highly significant, indicating that countries with larger external deficits are systematically more likely to receive advice for fiscal tightening, while those with surpluses tend to face more accommodative recommendations. Including country fixed effects yields similar results, suggesting that even within countries, a deterioration of the external balance is associated with tighter fiscal advice.

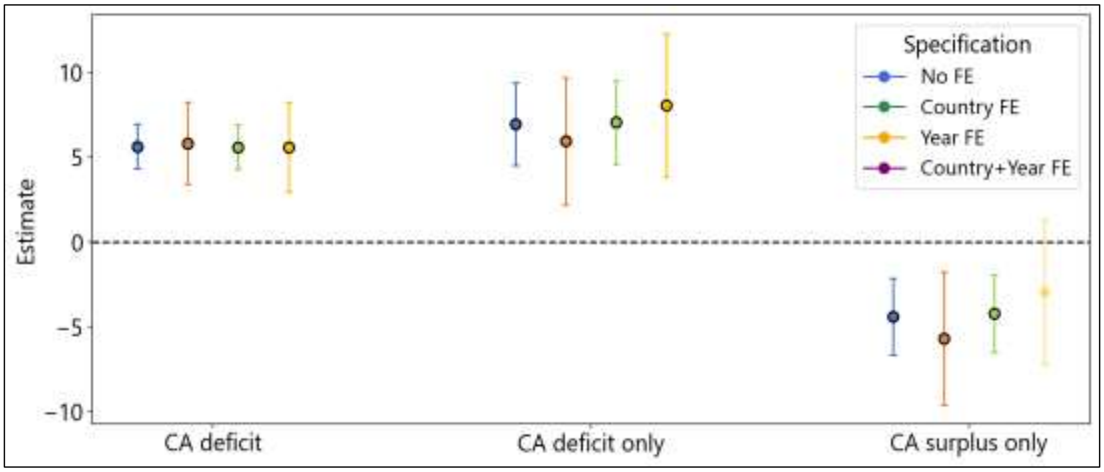
Fiscal advice displays a modest asymmetry in its response to external imbalances. When distinguishing between current account deficit and surplus cases,[26] the results show that fiscal tightening recommendations are positively associated with current account deficits across all specifications (Figure 9, center). For current account surpluses, the coefficients are negative and mostly statistically significant—except in the specification with both country and time fixed effects—indicating that fiscal advice tends to become more accommodative when external positions are stronger. However, the magnitude of this difference is relatively limited: without fixed effects, the coefficient on surpluses is around –4.4 compared to about 7 for deficits, and the asymmetry largely disappears once country effects are introduced. On balance, these results suggest that while IMF staff place somewhat greater emphasis on correcting external deficits than on offsetting external surpluses, the aggregate asymmetry is modest.

---

[25] Detailed estimates corresponding to these results are reported in Table AVI.4. of the Appendix.

[26] We estimate the following specification: $y^{*}_{i,t} = \beta Output\ Gap_{i,t} + \gamma FB\ gap_{i,t-1} + \phi Debt_{i,t-1} + \eta_1 CA_{i,t-1} \times D_{i,t-1} + \eta_2 CA_{i,t-1} \times (1 - D_{i,t-1}) + \varepsilon_{i,t}$ where D is an indicator variable that takes value 1 when the current account balance as percent of GDP is positive and 0 otherwise.

Figure 9. External Imbalances



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: All regressions use contemporaneous data. "CA deficit" reports the baseline coefficient on the current account deficit (percent of GDP), while "CA deficit only" and "CA surplus only" present asymmetric effects when allowing the impact to differ by the sign of the current account. Vertical lines indicate 95 percent confidence intervals. Detailed estimates are provided in Table AVI.5. of the Appendix. Brighter colors denote statistically significant estimates; lighter shades indicate insignificant results.

The asymmetry becomes more pronounced when the analysis is disaggregated by income group.[27] The responsiveness of fiscal advice to external imbalances is strongest among AEs, where coefficients on current account deficits are large and highly significant, while those on surpluses are smaller and often insignificant once country effects are included. Among EMMIEs, both deficit and surplus coefficients are significant in several specifications, but the asymmetry remains smaller in magnitude. In LICs, fiscal advice is less sensitive overall, though the coefficient on deficits becomes significant in the model with country and time fixed effects. Taken together, these results indicate that while the global asymmetry is limited, the Fund's fiscal advice reacts more strongly to external deficits than surpluses within individual income groups—particularly among AEs.

F. Exchange Rate Regime

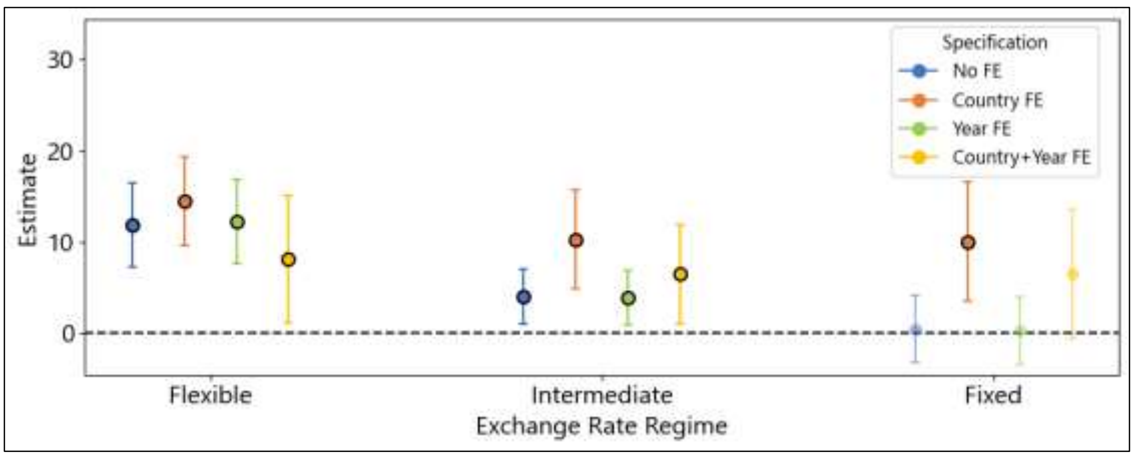
The relationship between exchange rate regimes and the countercyclicality of IMF fiscal advice can be assessed through the lens of the Mundell–Fleming model. In the textbook framework, fiscal policy plays a stronger stabilization role under fixed exchange rates, where monetary policy is constrained by the need to maintain the peg. By contrast, under flexible exchange rates, monetary policy retains autonomy, and fiscal policy is expected to play a more limited role in demand management. Everything else being equal, one would therefore expect fiscal recommendations to be more countercyclical under fixed than under flexible regimes. To test this prediction, we interact the output gap with indicators for flexible, intermediate, and fixed

[27] Detailed estimates corresponding to these results are reported in Tables AVI.6. and AVI.7. of the Appendix.

exchange rate arrangements, using the IMF's de facto classification from the Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER), grouped following Bal Gunduz and Darius (2021).[28]

The results reveal the opposite pattern of that predicted by the Mundell–Fleming model. As shown in Figure 10, fiscal advice is most countercyclical under flexible exchange rates, where the output-gap coefficients are large and highly significant across specifications, and least responsive under intermediate and fixed regimes.[29] This divergence from theoretical expectations likely reflects the fact that "all else" is not equal across exchange rate regimes. Countries maintaining fixed exchange rate arrangements often face deeper structural and credibility constraints that limit their ability to deploy fiscal policy countercyclically. Many adopt pegs precisely to import monetary credibility or to anchor expectations in the presence of weak policy frameworks and limited market confidence. These same characteristics—restricted fiscal space, vulnerability to capital outflows, and a stronger emphasis on external balance—make discretionary fiscal expansion less feasible. By contrast, countries with flexible exchange rates typically enjoy greater policy autonomy and credibility, allowing the Fund to recommend more active fiscal responses to cyclical conditions.

Figure 10. Countercyclicality by Exchange Rate Regime



Sources: WEO; Kose and others (2022); IMF's AREAER Database; IEO staff calculations.
Note: All regressions use contemporaneous data. Vertical lines indicate 95 percent confidence intervals. Detailed estimates are provided in Table AVI.8. of the Appendix. Brighter colors indicate statistically significant estimates, while lighter shades represent insignificant results.

[28] Fixed regimes include conventional pegs, currency boards, and exchange arrangements with no separate legal tender. Intermediate regimes include managed floats without a predetermined path, stabilized or other managed arrangements, crawl-like arrangements, crawling pegs, pegged exchange rates within horizontal bands, and crawling bands. Flexible regimes encompass floating, free floating, and independently floating rate arrangements.

[29] Detailed estimates corresponding to these results are reported in Table AVI.8. and AVI.9. of the Appendix.

G. Fiscal Institutions

Fiscal institutions shape not only countries' policy capacity but also how the IMF frames its fiscal advice. The Fund has long emphasized that credible medium-term fiscal frameworks can expand short-term policy space by strengthening confidence and anchoring expectations (Cohen-Setton and Montiel, 2025; Eichengreen and Gupta, 2025). As a result, Article IV reports often argue that temporary fiscal support is more effective and sustainable when embedded in credible medium-term consolidation plans and supported by robust institutions, such as fiscal rules and independent fiscal councils, that help preserve confidence while enabling countercyclical action. At the same time, fiscal rules typically constrain how much stimulus can be undertaken. Whether the presence of fiscal institutions ultimately expands or restricts the scope for countercyclical fiscal policy therefore remains an empirical question.

To explore this question, we extend the baseline framework to incorporate the role of fiscal institutions. Specifically, we introduce a dummy variable that equals one when a fiscal rule or an independent fiscal council is in place, using data from the IMF Fiscal Rules and Fiscal Councils Datasets (Alonso and others, 2025a; 2025b).[30] This variable is interacted with the output gap, the FB gap, and the debt-to-GDP ratio to test whether the presence of fiscal institutions alters the responsiveness of IMF fiscal advice to cyclical and sustainability conditions. The resulting specification allows us to assess both the overall effect of institutional strength on the likelihood of fiscal tightening and whether advice becomes more countercyclical when fiscal credibility is institutionally anchored.

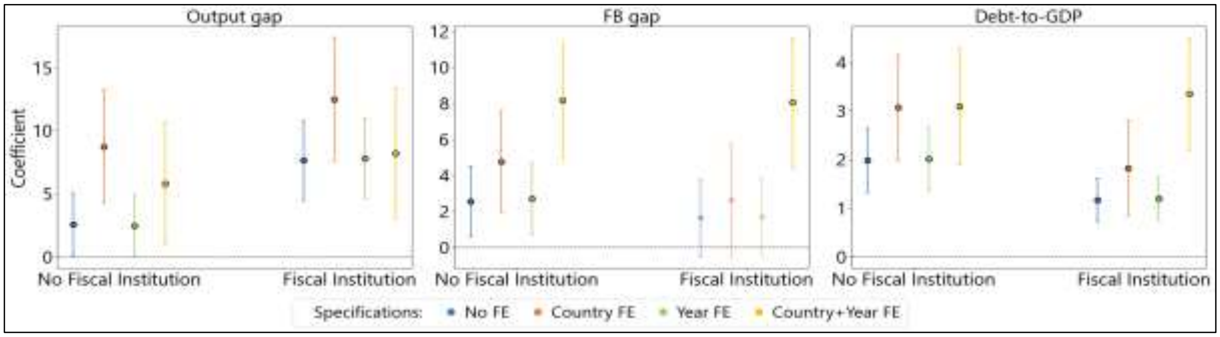
The results show that countries with fiscal institutions receive more countercyclical and less consolidation-oriented IMF advice. As illustrated in Figure 11,[31] the interaction between the output gap and the presence of fiscal institutions is large and highly significant across all specifications, indicating that Fund advice is more responsive to cyclical conditions where fiscal frameworks are in place. This pattern supports the view that credible institutions allow staff to emphasize short-term stabilization without undermining confidence in long-term sustainability. In such settings, fiscal rules and councils act as enablers rather than constraints—consistent with the Fund's argument that well-designed frameworks can help reconcile countercyclical flexibility with medium-term discipline. Consistent with this pattern, fiscal advice in countries with stronger institutions is less influenced by fiscal sustainability indicators. The coefficients on the FB gap and the debt-to-GDP ratio are smaller when fiscal institutions are present, suggesting that staff place less weight on fiscal sustainability variables once credibility mechanisms are in place. These findings echo Bianchi, Ottonello, and Presno (2023), who show that credible commitments to future adjustment allow policymakers to respond more forcefully to cyclical conditions in the short term. Fiscal institutions, by signaling future discipline, thus permit staff to advocate more

[30] Following the classification in Alonso and others (2025a; 2025b), we construct country-year dummies for the presence of fiscal rules and fiscal councils.

[31] Detailed estimates corresponding to these results are reported in Tables AVI.10. and AVI.11. of the Appendix.

accommodative recommendations during downturns and to calibrate consolidation more gradually over the cycle.

Figure 11. Fiscal Institutions



Sources: WEO; Kose and others (2022); Alonso and others (2025a, 2025b); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. Fiscal institution variables are derived from the IMF Fiscal Rules and Fiscal Councils Datasets (Alonso and others, 2025a; 2025b). For each country-year, a dummy equals 1 if any fiscal rule or council is present and 0 otherwise. Vertical lines indicate 95 percent confidence intervals. Detailed estimates are provided in Table AVI.11. of the Appendix. Brighter colors indicate statistically significant estimates, while lighter shades represent insignificant results.

H. Non-Economic Factors

While grounded in economic analysis, IMF fiscal advice may also reflect non-economic influences linked to institutional, political, and individual characteristics. Political economy research has long emphasized that international organizations are not purely technocratic actors: their policy advice is shaped by global power dynamics, bureaucratic norms, and the personal beliefs of staff. This literature raises a central question for IMF surveillance—whether fiscal recommendations reflect not only objective macroeconomic and fiscal conditions but also the incentives, constraints, and worldviews of those delivering them. One strand of research highlights geopolitical influences, arguing that major shareholders use international organizations to advance strategic interests through processes of bargaining and “horse-trading” (Copelovitch, 2010; Stone, 2011; Dreher and others, 2022). A second strand focuses on bureaucratic culture, suggesting that institutional histories and internal norms foster relatively stable modes of thinking that shape how evidence is interpreted across country cases (Chwieroth, 2010; Weaver, 2022). A third line of inquiry underscores the agency of individual staff, particularly mission chiefs, whose personal ideologies, professional networks, and educational backgrounds can influence how data are framed and policy advice formulated (Lang, Wellner, and Kentikelenis, 2024; Heinzel, 2022).

We investigate whether the well-known divide between “freshwater” and “saltwater” schools of thought helps explain differences in IMF fiscal advice. This intellectual split, popularized by Krugman (2009), captures enduring differences in how economists view fluctuations and the appropriate role of fiscal policy. Freshwater institutions, such as the University of Chicago and the University of Minnesota emphasize frictionless markets and rational expectations, interpreting recessions as efficient adjustments to real shocks and viewing fiscal activism as potentially distortionary. Saltwater schools, by contrast—exemplified by Harvard, MIT, and UC Berkeley—adopt New Keynesian frameworks that highlight market imperfections and nominal rigidities, providing a rationale for countercyclical fiscal intervention. If mission chiefs internalize these paradigms through their graduate training, their fiscal recommendations may systematically differ in both the intensity of their response to cyclical conditions and their underlying predisposition toward fiscal activism.

To test whether these intellectual traditions shape IMF fiscal advice, we link the macroeconomic views associated with staff training to the fiscal recommendations made in Article IV consultations. Using the dataset compiled by Lang, Wellner, and Kentikelenis (2024), which covers 835 IMF staff serving as mission chiefs across 190 countries between 1980 and 2016, we classify mission chiefs according to their graduate education at institutions traditionally associated with either "freshwater" or "saltwater" schools of thought. This classification serves as a proxy for exposure to distinct macroeconomic paradigms regarding the desirability and effectiveness of countercyclical fiscal policy. We extend our baseline fiscal reaction framework by including indicator variables for freshwater and saltwater backgrounds and interact them with the output gap. This strategy allows us to test whether mission chiefs trained in different academic traditions differ both in their responsiveness to cyclical conditions and in their average propensity to recommend fiscal tightening or loosening.

The results indicate that staff educational background meaningfully influences the formulation of fiscal advice.[32] The interaction between the output gap and the saltwater dummy is positive and statistically significant across all specifications, with larger coefficients than those associated with freshwater backgrounds. This finding suggests that saltwater-trained mission chiefs place greater weight on macroeconomic stabilization when formulating near-term fiscal advice. The freshwater interaction term is likewise positive, though smaller and statistically significant only in some specifications, consistent with a more cautious approach to countercyclical fiscal policy. In addition to these interaction effects, we detect direct differences in the thresholds that determine the switch between loosening, neutral, and tightening advice: staff trained at freshwater institutions are, on average, more likely to recommend tighter fiscal stances irrespective of cyclical conditions. Taken together, these results suggest that the intellectual traditions transmitted through graduate education continue to shape the way fiscal policy advice is framed within the IMF, introducing a non-economic source of heterogeneity into surveillance outcomes.

VI. CONCLUSION

This paper has developed and implemented a novel methodology to analyze the IMF's fiscal policy advice systematically across its membership over a 26-year period (1998–2023). Using state-of-the-art LLMs, the study extracts and classifies fiscal stance recommendations from more

[32] Detailed estimates corresponding to these results are reported in Table AVI.12. of the Appendix.

than 3,000 Article IV reports and links them to contemporaneous macroeconomic conditions. By combining multiple models—OpenAI's o1, Claude 3.7 Sonnet, and DeepSeek-R1—with carefully designed prompts, majority voting, and confidence weighting, the approach provides a reliable and transparent framework for identifying fiscal policy advice in a consistent, replicable way. This methodology demonstrates how recent advances in natural-language processing can be applied to institutional documents to support systematic policy evaluation within and beyond the Fund.

The analysis also develops a framework for relating near-term fiscal advice to economic circumstances at the time of each consultation. By constructing contemporaneous indicators of cyclical conditions, fiscal sustainability, and external imbalances, the paper evaluates whether the Fund's recommendations reflect a balance between output stabilization and debt sustainability concerns. Our results show that near-term advice is generally countercyclical and that this relationship strengthened after the GFC. Advice on the fiscal stance also responds to external imbalances: countries with current-account deficits are more likely to receive tightening recommendations, while those with surpluses are somewhat more likely to be advised to ease. The analysis further highlights that the presence of stronger fiscal institutions—such as fiscal rules or medium-term frameworks— receive advice that is more responsive to cyclical conditions and less constrained by fiscal sustainability concerns, indicating that credible institutions enable Fund staff to tailor fiscal guidance more flexibly to country-specific contexts.

Taken together, these results provide a foundation for more systematic evaluation of IMF fiscal advice. The framework offers staff, evaluators, and external stakeholders a practical tool to identify cases where advice deviates from typical relationships with economic fundamentals—helping ensure that such deviations are reviewed more closely and, when warranted, justified by country-specific factors. Beyond its immediate findings, the paper demonstrates the potential of LLMs to enhance transparency and replicability in IMF surveillance analysis and provides a template for applying similar techniques to other policy areas.

APPENDIX I. LLM-BASED DATASET CONSTRUCTION AND DIAGNOSTICS

A. Prompt

The following is the prompt used to determine the IMF staff's advice on policy stance by classifying the text from the Staff Appraisal section into near- and medium-term fiscal, monetary, and macroprudential stances.

Box AI.1. Prompt Used to Extract Fiscal Stance from Article IV Reports

You will be given the staff appraisal section of an IMF Article IV report. The staff appraisal section summarizes the IMF staff's analysis of recent economic developments and policies, views on the outlook and of risks, and policy advice.

Based on the provided text, your task is to identify and analyze the IMF staff's advice on fiscal policies for both the near-term and medium-term, as well as the advice on monetary policy and macroprudential policy.

Near-term refers to the current economic cycle and the immediate future, typically the current and next year. It relates to addressing current economic challenges or imbalances, for example, the recovery from an ongoing recession or crisis or the response to overheating. Near-term advice typically focuses on immediate policy responses, measures to stabilize the economy in the short run, actions to support or moderate current economic growth, and actions to address urgent fiscal imbalances.

Medium-term generally refers to a longer time horizon, typically 2 to 5 years into the future or more. It often relates to structural reforms and long-term fiscal sustainability. Medium-term advice typically focuses on fiscal consolidation efforts to ensure long-term debt sustainability, structural reforms and public investments to improve economic efficiency and potential growth, measures to address long-term demographic or economic challenges, or policies to build fiscal buffers for future economic shocks.

Classify the fiscal and monetary advice into one of four categories: Tighten, Neutral, Loosen, or Unclear. Categorize the macroprudential advice into one of three categories: Tighten, Neutral/Unclear, or Loosen.

Report your certainty weight (0–100 percent) for each category. If your assessment falls between categories, pick the likelier one but assign positive certainty weights to all relevant categories.

Categories for fiscal stance: The fiscal stance reflects the impact of changes in governments' discretionary spending and revenue measures on the budget. It is measured by the change in the cyclically adjusted primary balance, i.e., the fiscal balance after stripping out the effects of macroeconomic developments and interest payment fluctuations, or equivalently by summing the net budgetary impact of discretionary fiscal actions. The recommended fiscal stance should be categorized as follows:

Tighten: The staff recommends or endorses policies leading to an increase in the cyclically adjusted primary balance. Examples: consolidation (reducing deficits, increasing surpluses), continued effort to narrow the structural deficit.

Neutral: The staff recommends or endorses no significant change in the cyclically adjusted primary balance. Examples: maintaining current spending and revenue policies without additional consolidation or expansion; letting only automatic stabilizers operate. If the "current policy" is on a path of tightening (or loosening) beyond automatic stabilizers, that is effectively a tightening (or loosening) stance and should be classified accordingly, rather than "Neutral".

Loosen: The staff recommends or endorses a net increase in the cyclically adjusted primary balance. Examples: stimulus spending, higher deficits, lower surpluses.

Unclear: Reflects that there are conflicting or ambiguous recommendations without a clear overall direction, or that the text offers no explicit fiscal advice, or that the advice is too vague to categorize definitively.

Categories for monetary stance:

Tighten: Refers to recommendations for more restrictive monetary conditions, higher interest rates, reduced liquidity, or other contractionary measures to control inflation. Sample terms: rate hikes, monetary tightening, liquidity absorption, less accommodative stance.

Neutral: Refers to recommendations to maintain the current monetary policy stance, suggesting the current calibration of policy instruments is appropriate for economic conditions. Sample terms: maintain current stance, hold steady.

Loosen: Refers to recommendations of accommodative monetary conditions, lower interest rates, increased liquidity provision, or other expansionary measures to support growth and financial conditions. Sample terms: rate cuts, monetary easing, liquidity injection, accommodative stance.

Unclear: Reflects that there are conflicting or ambiguous recommendations without a clear overall direction, or that the text offers no explicit monetary advice, or that the advice is too vague to categorize definitively.

Categories for macroprudential stance:

Tighten: Refers to recommendations to tighten MPP measures, to restrict credit growth or strengthen financial stability buffers through macroprudential measures, higher capital/liquidity requirements, stricter lending standards, or other tools affecting cyclical credit conditions. Sample terms: tighten macroprudential policy, strengthen prudential measures, build additional buffers.

Neutral/Unclear: Applies when the current stance is deemed appropriate, no explicit cyclical financial policy recommendation is made, or the focus is on structural rather than cyclical financial policy aspects. Also covers cases where recommendations are ambiguous or purely monitoring-focused. Sample terms: maintain current MPP stance, continue monitoring risks.

Loosen: Refers to recommendations to ease credit conditions or release financial buffers to support economic activity. May include relaxing macroprudential measures, releasing countercyclical buffers, or easing lending standards. Sample terms: ease lending conditions, release buffers, support credit provision.

If the staff explicitly agrees with and endorses the stance of country authorities, then classify the stance of the IMF staff according to how you would classify the stance of country authorities. For instance, if authorities' plan is to reduce deficits over time or tilts towards fiscal consolidation, then a staff recommendation to continue would be classified as "Tighten". For monetary policy, apply similar logic: for example, depending on the context, "continue the current path" could effectively mean continued rate hikes, implying "Tighten", or ongoing accommodative measures, implying "Loosen", or no further changes, implying "Neutral".

If staff advice is conditional (e.g., “Loosen if downside risks occur”), decide whether that baseline or the conditional scenario is more likely and classify based on the more likely scenario. If no clear or coherent direction is given, choose Unclear.

Required Output Format: After reviewing the staff appraisal excerpt, provide your analysis using this structure:

<near_term_reasoning>[Your reasoning for the near-term fiscal stance]</near_term_reasoning>

<near_term_advice>[Choose one: Loosen, Neutral, Tighten, or Unclear]</near_term_advice>

<near_term_certainty>Loosen: [0–100], Neutral: [0–100], Tighten: [0–100], Unclear: [0–100]</near_term_certainty>

<medium_term_reasoning>[Your reasoning for the medium-term fiscal stance]</medium_term_reasoning>

<medium_term_advice>[Choose one: Loosen, Neutral, Tighten, or Unclear]</medium_term_advice>

<medium_term_certainty>Loosen: [0–100], Neutral: [0–100], Tighten: [0–100], Unclear: [0–100]</medium_term_certainty>

<monetary_reasoning>[Your reasoning for the monetary stance]</monetary_reasoning>

<monetary_advice>[Choose one: Loosen, Neutral, Tighten, or Unclear]</monetary_advice>

<monetary_certainty>Loosen: [0–100], Neutral: [0–100], Tighten: [0–100], Unclear: [0–100]</monetary_certainty>

<macropru_reasoning>[Your reasoning for the macroprudential policy stance]</macropru_reasoning>

<macropru_advice> [Choose one: Tighten, Neutral/Unclear, Loosen] </macropru_advice>

<macropru_certainty> Tighten: [0-100], Neutral/Unclear: [0-100], Loosen: [0-100] </macropru_certainty>

===

Staff Appraisal:

Here is the relevant section from the IMF Article IV report on {country}, dated {publication date}:

<report_extract>

{staff appraisal}

</report_extract>

B. Coverage of Article IV Reports

Table AI.1 summarizes the coverage of Article IV consultation reports included in the dataset. The sample comprises 3,115 reports across 193 countries from 1998–2023, averaging roughly one consultation every 1.6 years per country. AEs are covered most frequently, with an average interval of 1.3 years between consultations, followed by EMMIEs at 1.6 years LICs at 2 years. Coverage is well distributed across IMF area departments, with the European Department accounting for the largest share of reports (826), reflecting the relatively high concentration of AEs in that region.

Table AI.1. Coverage of Article IV Reports by Income Group and Area Department
(1998–2023)

	Number of AIVs	Number of Countries	Number of AIVs Per Country	Average Frequency (Years between AIV)
Income Groupings				
Advanced Economies (AEs)	779	39	20	1.3
Emerging Market and Middle-Income Economies (EMMIEs)	1565	96	16.3	1.6
Low-Income Countries (LICs)	771	58	13.3	2.0
IMF Departments				
European Dept (EUR)	826	45	18.4	1.4
Western Hemisphere Dept (WHD)	579	35	16.5	1.6
Asia-Pacific Dept (APD)	633	36	17.6	1.5
African Dept (AFR)	603	45	13.4	1.9
Middle East & Central Asia Dept (MCD)	469	31	15.1	1.7
Total	3115	193	16.1	1.6

Sources: IMF Article IV Reports; IEO staff calculations.
Note: Average frequency is calculated as the number of years in the sample (1998–2023, i.e., 26 years) divided by the number of Article IV reports per country.

C. Sample LLM Classification

Below, we present the full outputs from the three LLMs for a sample AIV report.

Box AI.2. Ireland 2019

Near-Term Fiscal Stance

Majority classification: Tighten

OpenAI: Tighten (Loosen: 0, Neutral: 10, Tighten: 85, Unclear: 5)

Staff explicitly calls for small budget surpluses in 2019–20 to alleviate demand pressures and build buffers, underscoring a recommendation to avoid further spending overruns and save any corporate tax windfalls. In the baseline (i.e., absent a disorderly Brexit), this guidance amounts to fiscal tightening.

Claude: Tighten (Loosen: 20, Neutral: 0, Tighten: 80, Unclear: 0)

The IMF staff clearly recommends fiscal tightening for the near-term baseline scenario. The text explicitly states "Fiscal policy should be tightened to alleviate demand pressures and build buffers against potential shocks." Furthermore, they recommend "pursuing small budget surpluses in 2019–20, including by avoiding further spending overruns and saving any corporate tax windfalls." These statements point to increasing the cyclically adjusted primary balance. However, there is a conditional recommendation for fiscal loosening in case of a "disorderly no-deal Brexit" where the IMF advises allowing "automatic fiscal stabilizers to operate freely and provide targeted, temporary, and effective support to hard-hit sectors" and preparing for "a fiscal stimulus, depending on the severity of the downturn." Since this is a contingency scenario rather than the baseline, I'm categorizing the main advice as tightening while acknowledging the conditional loosening recommendation.

DeepSeek: Tighten (Loosen: 0, Neutral: 5, Tighten: 90, Unclear: 5)

The IMF explicitly recommends tightening fiscal policy to alleviate demand pressures and build buffers, including pursuing small budget surpluses in 2019–20, avoiding spending overruns, and saving corporate tax windfalls. While conditional advice for fiscal stimulus is provided in case of a disorderly Brexit, the baseline recommendation is tightening.

Medium-Term Fiscal Stance

Majority classification: Tighten

OpenAI: Tighten (Loosen: 0, Neutral: 20, Tighten: 70, Unclear: 10)

Staff emphasizes reducing the public debt ratio below 50 percent over the medium term and recommends measures such as streamlining VAT, reforming income taxation, and gradually increasing property taxes to reduce reliance on volatile corporate taxes. Moderating expenditure growth and using any bank-share disinvestment proceeds to reduce public debt also point to a consolidation path. Hence, the overall advice implies a tighter position over the medium term.

Claude: Tighten (Loosen: 0, Neutral: 0, Tighten: 95, Unclear: 5)

The medium-term fiscal advice clearly points toward tightening. The staff recommends aiming at "reducing the public debt ratio below 50 percent over the medium term." There are several specific tightening measures mentioned: streamlining VAT, reforming income taxation, gradually increasing property taxes, and "moderating expenditure growth while increasing its efficiency." The IMF also advises that "any proceeds from government disinvestments in the financial sector should be used for public debt reduction." Additionally, to tackle long-term challenges like population aging, they recommend "a review of social security contributions and benefits" to safeguard long-term viability and avoid future budgetary pressures. All these recommendations collectively point to fiscal consolidation in the medium term.

DeepSeek: Tighten (Loosen: 0, Neutral: 10, Tighten: 85, Unclear: 5)

The report emphasizes medium-term fiscal consolidation: reducing public debt below 50%, reforming taxes (VAT, income, property), moderating expenditure growth, and using disinvestment proceeds for debt reduction. These structural measures aim to ensure sustainability and reduce reliance on volatile corporate taxes.

Monetary Stance

Majority classification: Unclear.

OpenAI: Unclear (Loosen: 0, Neutral: 10, Tighten: 0, Unclear: 90)

No explicit guidance is offered on monetary policy. The report does not advise changes to interest rates or the broader monetary stance, so the direction is unclear.

Claude: Unclear (Loosen: 0, Neutral: 0, Tighten: 0, Unclear: 100)

There are no explicit recommendations regarding conventional monetary policy in the staff appraisal. This is likely because Ireland is a member of the Eurozone, and monetary policy decisions are made by the European Central Bank rather than national authorities. The only remotely monetary-related statement refers to releasing "the countercyclical capital buffer" in the event of a credit contraction, but this is a macroprudential tool rather than conventional monetary policy. Given the absence of any concrete monetary policy advice, the stance must be classified as unclear.

DeepSeek: Unclear (Loosen: 0, Neutral: 20, Tighten: 0, Unclear: 80)

The report does not explicitly address monetary policy (e.g., interest rates or liquidity measures). Ireland's monetary policy is set by the ECB, and the IMF staff does not provide direct recommendations. The mention of releasing the countercyclical capital buffer pertains to macroprudential policy, not monetary stance.

Macroprudential Stance

Majority classification: Tighten.

OpenAI: Neutral/Unclear (Tighten: 10, Neutral/Unclear: 80, Loosen: 10).

Staff finds current macroprudential settings "appropriately calibrated," while recommending expansion of the toolkit (e.g., a systemic risk capital buffer). They would also release the countercyclical capital buffer if credit contracted sharply. These proposals do not clearly signal an immediate tightening or loosening, so the stance is best described as neutral/unclear.

Claude: Tighten (Tighten: 70, Neutral/Unclear: 25, Loosen: 5).

The IMF provides mixed guidance on macroprudential policy. On one hand, they state that "macroprudential policies appear to be appropriately calibrated," suggesting a neutral current stance. On the other hand, they recommend enhancing the toolkit with additional measures: complementing "existing limits on loan-to-value and loan-to-income ratios with debt-based measures" and "expanding the toolkit with a systemic risk capital buffer" to bolster system resilience. These recommendations suggest strengthening the macroprudential framework, which leans toward tightening. There is also a conditional recommendation to release the countercyclical capital buffer in case of a "sharp contraction in bank credit," but this is contingent on a specific scenario. Overall, the dominant recommendation is to maintain current calibration while expanding the toolkit with additional prudential measures, which points toward a tightening bias.

DeepSeek: Tighten (Tighten: 70, Neutral/Unclear: 30, Loosen: 0).

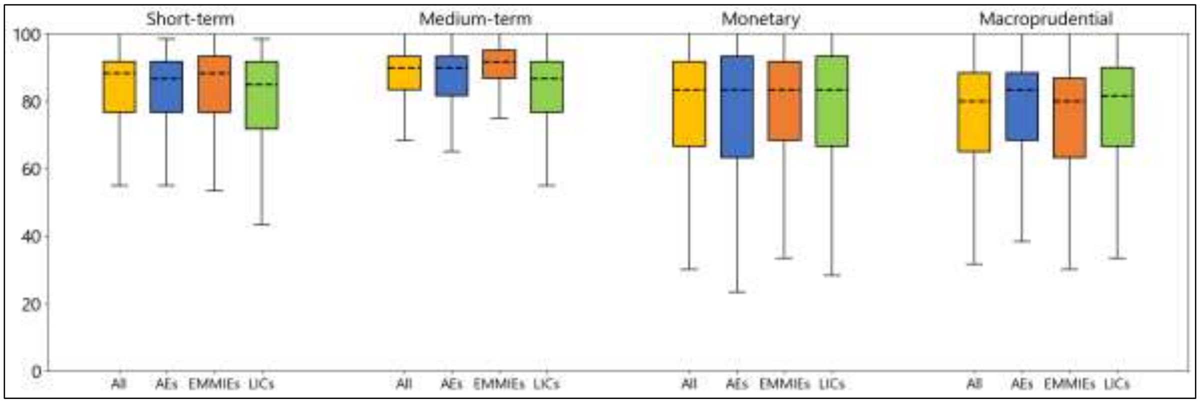
The IMF advises enhancing macroprudential tools by introducing debt-based mortgage measures and a systemic risk capital buffer, which constitutes tightening. While current policies are deemed appropriate, the recommendations to expand the toolkit signal a need for stricter measures.

Sources: Ireland 2019 Article IV report; IEO staff calculations.
Note: The table presents full outputs from OpenAI's o1, Anthropic's Claude 3.7 Sonnet, and DeepSeek's R1 models. Each policy stance classification includes the majority voted final classification, individual model's classification, certainty score (totaling 100), and reasoning provided by the model.

D. Model Classification Confidence

Figure AI.1 displays the distribution of LLM classification certainty scores across four policy areas—near-term fiscal stance, medium-term fiscal stance, monetary policy, and macroprudential policy—for all countries and by income group. The certainty scores represent the model's self-reported confidence (on a 0–100 scale) in the accuracy of its classification, generated after reasoning through each case. We explicitly instruct the model in the prompt to report a certainty score for each classification. The boxplots show the median (black dashed line), interquartile range, and overall dispersion of certainty values. Certainty levels are generally high across income groups, with medians in the mid-80s to 90s range. Differences across groups are modest: AEs and EMMIEs tend to show slightly higher median certainty in some areas, while LICs display somewhat greater variability.

Figure AI.1. Certainty Weights Across Policies Areas and Income Groups



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Boxplots show the distribution of LLM classification certainty weighs across four policy areas (short-term, medium-term, monetary, and macroprudential) for all countries and by income group (AE, EMMIE, LIC). The black dashed line within each box represents the median certainty, while the boxes and whiskers indicate the interquartile range and overall dispersion of certainty scores.

APPENDIX II. ANALYSIS OF TEXTUAL ECONOMIC DATA WITH LLMS

This appendix describes how we analyzed textual economic data to perform the evaluation reported in this paper. We explain the initial steps for setting up and running LLM-based textual analysis in bulk through application programming interfaces (APIs) and how to think about designing prompts. We draw on our hands-on experience from working on this paper and aim to give readers a clear sense of our approach and the broader lessons for textual economic data analysis that we learned. We also hope that our description can serve as a practical guide for researchers who are new to using LLMs in economic research.[1]

A. Textual Data Selection

The central objective of the economic and financial literature on natural language processing (NLP) is to quantify and systematically extract information from natural language produced by humans to capture expectations, sentiment, and signals that are not directly contained in traditional quantitative data (Gentzkow, Shapiro, and Taddy, 2019; Hassan and others, 2019; Shapiro and Wilson, 2022; Kalyani and others, 2025). This literature has analyzed textual sources, such as political speech, job postings, patents, news articles, social media posts, earnings call transcripts, corporate filings, and central bank communications.

In our research, we analyze the policy advice provided by IMF staff to country authorities. We classify the overall advice on fiscal, monetary, and macroprudential policies into broad categories of Loosen, Neutral, Tighten, or Unclear, based directly on the textual content of the Staff Appraisal sections in AIV reports. This approach places emphasis on extracting the policy stance embedded in the text rather than on measuring tone or thematic composition. Comparable work in this direction includes Hansen and Kazinnik (2024) and Jha and others (2024).

Our starting point is the Article IV reports[2] that are regularly produced by IMF staff for its member countries. Article IV reports cover a wide range of economic and financial topics from different angles and perspectives. If a human analyst were tasked with extracting the IMF staff's advice on, e.g., the near-term fiscal stance, a common approach would be to review all sections that relate to fiscal issues, read through the relevant paragraphs carefully, and then synthesize the information to form a single classification such as Tighten or Loosen. While feasible for a small number of cases, this process becomes extremely time-intensive and impractical when scaled across many countries and years. LLMs can automate this repetitive and labor-intensive task. The goal is to replicate the human process of identifying, reading, and synthesizing the relevant content in a consistent and systematic manner that can be applied across the entire collection of Article IV reports.

---

[1] See Korinek (2023, 2024, and 2025) for more on using generative AI in economic research.

[2] Article IV reports are published in English, which is an important consideration when selecting data sources, particularly given the composition of pre-training data used to train LLMs.

To focus directly on the IMF staff's views and recommendations, we provided only the Staff Appraisal section of each Article IV report to the model for evaluation. This section typically consists of just a few pages compared to the full Article IV reports, which can easily exceed a hundred pages. While this focus was a necessity in 2024 when we started work on this to avoid overwhelming the limited processing capacity of LLMs at the time, we found that it is still useful today as the remainder of the report occasionally proves distracting for LLMs' ability to assess Fund advice.

The raw text data can come in many formats such as Word, PDF, or plain text files. Because these formats are unstructured, a preprocessing step is needed to convert them into clean text that the model can use as input. Preprocessing means selecting only the relevant parts of the document, removing or separating extra material such as tables, figures, headers, fixing formatting issues, and ensuring the text contains the information necessary for the research. This step is crucial because unnecessary or messy text can confuse the model and reduce the quality of its answers. In practice, most of this work can be done using Python, which has powerful libraries for handling text. For example, the "pandas" library can be used to read and filter the text data stored in a CSV file, while the "re" library can help remove unwanted characters or patterns using regular expressions.

In our case, the textual data we employed was prepared by the IMF's SPR Department and was delivered in CSV format. Each row in the CSV file corresponds to a single paragraph taken from the main body of an Article IV report and includes an identifier indicating the section of the report from which the paragraph originated. The text was extracted directly from the PDF version of each report, which preserves the original order of paragraphs and allows for better mapping of content to its location in the document. This paragraph-level structure made it possible to filter sections and build a clean and consistent input string for the LLM.

We performed a basic consistency check by counting the words in each extracted Staff Appraisal sections and examined the distribution to spot unusually short or long excerpts. Taking a closer look at these outliers helped us find cases where the text extraction or separation process had failed and fix these issues manually to ensure that the final input data was clean, consistent, and ready for use with the LLM.

B. Prompt Engineering

When working with a large number of documents or long texts, the back-and-forth between user and LLM that is typical of chatbots quickly becomes impractical. Instead, it is more efficient to repeatedly call the application programming interface (API) of language models to process all questions in bulk. During each model call, we provide a single, complete prompt that includes both the instructions and the text to analyze.[3]

---

[3] See Appendix I to see the full prompt in detail.

A helpful way to begin is to test a prompt in a chat interface using a few sample cases. This allows the researcher to see whether the model interprets the instructions as intended. The prompt itself can also be improved with the help of an LLM by asking the model to suggest clearer or more specific wording. In this paper, the prompt was refined through repeated testing to ensure that the model followed the intended reasoning process. To align the model with the economic meaning of policy stance[4] and the distinction between near-term and medium-term, these concepts were defined clearly within the prompt. The model was asked to perform its internal reasoning and then provide a short justification, a categorical label for fiscal, monetary, and macroprudential stances, and a confidence score for each label.

Requiring a short justification encouraged the model to pay attention to details in the text and created a record that could be reviewed later. For texts that were ambiguous, the prompt allowed the model to choose an "Unclear" category or to divide its confidence between categories when the interpretation was uncertain. As Kalai and others (2025) note, LLMs are trained to provide an answer rather than to say, "I do not know." Because of this, prompts should be written in a way that gives the model permission to express uncertainty when the evidence in the text is not strong enough to justify a clear conclusion.

Figure AII.1. LLM Processing Pipeline

Source: IEO staff.
Note: Orange cells represent steps that typically involve coding or scripting, while blue cells indicate conceptual or design stages. Together, these steps form the full workflow from raw text preparation to the generation and organization of model outputs.

C. API Calls

Application programming interfaces (APIs) are tools that let computers talk directly to other services. Many LLM companies, such as OpenAI and Anthropic, provide access through their APIs. To use one, a user must first create an API key on their website. This key is a unique code that identifies the user akin to a credit card and allows for payment to use the model. A good starting point for learning how to use an API is to go through the cookbooks[5] or sample codes[6] provided by the model companies. These resources show how the programming structure should look, how to organize API call functions, how to store model outputs, and how to troubleshoot

[4] See Section III on how we define fiscal stance.

[5] For example, see https://cookbook.openai.com/ or https://github.com/anthropics/claude-cookbooks.

[6] For example, see https://platform.openai.com/docs/quickstart.

common errors. Both OpenAI and Anthropic provide extensive documentation and well-written examples that serve as practical guides for setting up the coding environment for any LLM-based project. Another useful approach is to "vibe-code" on the LLMs' chatbot itself. This method helps researchers understand the logic of API calls and get comfortable with the coding workflow before writing full scripts. A third approach is to use the Playground tools offered by most LLM providers. The Playground provides an easy-to-use interface where users can adjust model settings (such as temperature, max tokens, and response style) and immediately see how these changes affect the output. It also has an option to generate the equivalent Python code that mirrors the selected settings.

In our setup, each classification task is processed as a separate instance, meaning that every Staff Appraisal is handled independently from others. For each document, we submit one chat completion request to the model that includes the extracted text as part of the prompt. Treating each classification as a separate instance ensures that the model focuses on the content of that document alone, without being influenced by other submissions.

When working with many such individual tasks, batch processing can make the workflow much more efficient. Batch processing groups many API requests into one large job that is sent to the model provider's servers to be processed together. This method saves time and cost because it reduces the overhead of sending and managing each request separately. Many AI companies, including OpenAI and Anthropic, already offer built-in batch processing as part of their APIs.[7] A similar effect can be achieved by writing code for parallel asynchronous API calls. In this approach, multiple requests are launched simultaneously from the local environment rather than waiting for one to finish before starting the next.

If the goal is for the model to generate answers only from the text provided in the prompt (as we did in the paper) and to produce the most consistent output possible, the researcher can adjust certain inference parameters in the API call. These parameters control how much randomness or variation is allowed in the model's responses. For example, OpenAI's models allow the temperature parameter—which captures randomness—to be set to zero, making the model's responses more deterministic

It is also important to understand several technical and cost-related limits that affect how LLMs can be used. One of these is a rate limit, which controls how many requests can be sent to the model within a certain time frame, usually measured per minute. Exceeding the rate limit causes errors or rejected requests.

A key concept is the size of a model's context window, which is the maximum amount of text that the model can process at one time. The context window is measured in tokens, where a token represents a small piece of text, roughly equal to four characters or about three-quarters

[7] For example, see https://platform.openai.com/docs/guides/batch.

of a word. Both the input prompt and the model's reply count toward this limit. If the total text exceeds the context window, it causes an error. A helpful way to estimate the number of tokens before sending a request is by using the "tiktoken" Python library, which can calculate token usage for a given piece of text and help users design prompts that stay within the model's capacity or control cost.

Because tokens directly determine cost, it is important to track their usage carefully. Each model provider charges a specific amount per million tokens, so longer texts input and output increase expenses. Most LLM providers offer dashboards that display real-time usage, token consumption, and total spending. Checking these dashboards regularly helps identify unexpected increases in cost and ensures that projects remain within budget.

When working with LLMs through public APIs, it is crucial to ensure that the materials sent are appropriate for external processing and exclude, for example, confidential documents. Many institutions have specific policies that govern how data can be shared with third-party services. Reviewing and following these guidelines before running LLMs ensures that research remains secure and compliant with institutional standards.

APPENDIX III. DATA DESCRIPTION

A. Data Sources

To align macroeconomic data with the timing of IMF staff recommendations, we matched each Article IV report to the World Economic Outlook (WEO) dataset that was current when the report was prepared. Specifically, each country–year observation was linked to the WEO vintage released immediately prior to the publication of the corresponding Article IV report, ensuring that the analysis draws on information available to staff at the time. When no preceding vintage was available, we systematically backfilled using earlier vintages until a non-missing observation was identified. If no earlier vintages existed, the latest available vintage was used instead, corresponding to the default (October 2024) assignment in the dataset. This procedure ensures that, to the greatest extent possible, each observation relies on contemporaneous macroeconomic data, resorting to earlier or default vintages only when necessary to fill remaining gaps. The Hamilton-filtered output gap is constructed by applying the Hamilton filter to the real GDP series across all WEO vintages. For each observation, the vintage closest to the Article IV report's publication date is selected, and the resulting values are used to supplement the WEO output gap in the analysis.

Table AIII.1. Data Description

Indicator	Description	N	Unit	Source	Vintage	Transformation	Time index
Macro policy advice							
ST stance	Majority vote (OpenAI if no agreement) for short-term fiscal stance	3115	Ordinal	AIV, Authors' calculations	No	None	T
MT stance	Majority vote (OpenAI if no agreement) for medium-term fiscal stance	3115	Ordinal	AIV, Authors' calculations	No	None	T
Monetary stance	Majority vote (OpenAI if no agreement) for monetary policy stance	3115	Categorical	AIV, Authors' calculations	No	None	T
Macroprudential stance	Majority vote (OpenAI if no agreement) for macroprudential policy stance	3115	Categorical	AIV, Authors' calculations	No	None	T
Cyclical conditions							
Output gap	Output gap series as a combination of WEO and Hamilton filter estimates	3090	Percent of potential GDP	WEO	Yes	Winsorized	T
5-year average growth deviation	Real GDP growth deviation from the preceding 5-year average growth	3103	Percent	WEO, Authors' calculations	Yes	Winsorized	T
Growth deviation from 5-year ahead WEO projection	Real GDP growth deviation from the 5-year ahead WEO projection	2944	Percent	WEO, Authors' calculations	Yes	Winsorized	T
Fiscal sustainability							
FB gap	Required adjustment in the fiscal balance to stabilize the debt-to-GDP ratio	2897	Percent of GDP	Kose and others (2022), Authors' calculations	No	Winsorized	T-1
PB gap 0	Sustainability gap, primary balance, country-specific conditions	2027	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
PB gap 1	Sustainability gap, primary balance, historical conditions	2027	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
PB gap 2	Sustainability gap, primary balance, current conditions	1688	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
PB gap 3	Sustainability gap, primary balance, stressed conditions	2005	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1

PB gap 4	Sustainability gap, primary balance, benign conditions	2005	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
VE Fiscal crisis	Vulnerability exercise's percentiles for fiscal sector crisis rating (low, medium, high)	3096	Categorical	SPR	No	None	T
Debt sustainability							
Debt-to-GDP	Ratio of general government gross debt to fiscal year GDP	3041	Percent of GDP	WEO	Yes	Winsorized	T-1
Debt-to-revenue	General government gross debt, % of average tax revenues	2921	Percent	Kose and others (2022)	No	Winsorized	T-1
Debt-to-average GDP	General government gross debt, % of 10-year moving average GDP	2937	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
Gross financing need	Sum of the public sector fiscal deficit and maturing debt over the following 12 months	2478	Percent of GDP	DSA, DSF	No	Winsorized	T-1
Effective nominal interest rate	Average effective interest rate the public sector pays on its debt stock	2527	Percent	DSA, DSF	No	Winsorized	T-1
Debt service	Total debt service paid as a ratio of general government revenue	2301	Percent	WEO	Yes	Winsorized	T-1
General government debt in foreign currency, % of total		556	Percent	Kose and others (2022)	No	Winsorized	T-1
Debt securities held by nonresidents, % of total		698	Percent	Kose and others (2022)	No	Winsorized	T-1
General government debt held by nonresidents, % of total		984	Percent	Kose and others (2022)	No	Winsorized	T-1
Concessional external debt stocks, % of external public debt		1718	Percent	Kose and others (2022)	No	Winsorized	T-1
Sovereign debt average maturity, years		658	Years	Kose and others (2022)	No	None	T
Central government debt maturing in 12 months or less		998	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
Total external debt stocks		2549	Percent	Kose and others (2022)	No	Winsorized	T-1
External debt in foreign currency, % of total		479	Percent	Kose and others (2022)	No	Winsorized	T-1
Private external debt stocks		2054	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
Domestic credit to private sector, % of GDP		2979	Percent of GDP	Kose and others (2022)	No	Winsorized	T-1
Short-term external debt stocks, % of total		2535	Percent	Kose and others (2022)	No	Winsorized	T-1
Short-term external debt stocks, % of reserves		2286	Percent	Kose and others (2022)	No	Winsorized	T-1
Total external debt stocks, % of reserves		2300	Percent	Kose and others (2022)	No	Winsorized	T-1
Total external debt stocks, % of reserves excluding gold		2300	Percent	Kose and others (2022)	No	Winsorized	T-1
Foreign currency long-term sovereign debt ratings		2159	Index from 1-21	Kose and others (2022)	No	None	T-1
5-year CDS spread		924	Basis points	Kose and others (2022)	No	Winsorized	T-1
External balance							
Current account deficit	Balance of Payments, Current Account	3098	Percent of GDP	WEO	Yes	Winsorized	T-1
Non-Economic factors							
Saltwater and Freshwater Universities	Categorization of a mission chief's alma mater institution into saltwater and freshwater schools	2326	Indicator	Lang, Wellner, and Kentikelenis (2024)	No	None	T

B. Distribution of Fiscal Advice by Macroeconomic Conditions

Table AIII.2 cross-classifies the share of IMF fiscal advice recommending tightening by countries' cyclical position and level of fiscal vulnerability. The table shows the proportion of tightening recommendations across three output-gap categories—positive, near balance, and negative—

and three fiscal-risk tiers (low, medium, and high) derived from the Vulnerability Exercise (VE). Overall, the share of tightening advice tends to increase with higher assessed fiscal risk and, to a lesser extent, with stronger cyclical conditions. Patterns are broadly similar across income groups, though the small number of observations in some cells, especially for LICs, calls for caution in interpretation. The table provides a descriptive summary of how the frequency of tightening recommendations varies with both cyclical and fiscal-risk classifications.

Table AIII.2. Distribution of Fiscal Tightening Advice by Macroeconomic and Fiscal Conditions
(Percent of cases recommending tightening)

	Output Gap	**Fiscal Vulnerabilities**		
		Low Risk	Medium Risk	High Risk
All Countries	Positive	71	79	84
	Close to balance	64	84	78
	Negative	60	74	85
AEs	Positive	70	83	100
	Close to balance	58	100	100
	Negative	58	87	100
EMMIEs	Positive	71	88	89
	Close to balance	70	89	60
	Negative	63	76	91
LICs	Positive	71	71	82
	Close to balance	100	71	88
	Negative	83	59	74

Sources: Vulnerability Exercise (VE) assessment; WEO; IEO staff calculations.
Note: Output gap: "Positive" indicates an output gap above 0.5 percent of GDP; "Close to balance" corresponds to a gap between -0.5 percent and +0.5 percent; and "Negative" refers to a gap below -0.5 percent. Fiscal vulnerabilities are classified as Low, Medium, or High according to the IMF's VE for the fiscal sector (IMF, 2021). The VE is a country-specific assessment of near-term macroeconomic risks, estimating the likelihood of various stress events over a 1-to-2-year horizon. These risks are converted into model-based categories: countries above the 80th percentile are classified as High risk, those below the median as Low risk, and those in between as medium risk.

C. Summary Statistics of Baseline Variables

Table AIII.3 reports descriptive statistics for the main explanatory variables used in the baseline regressions—the output gap, the fiscal balance (FB) gap, and the debt-to-GDP ratio—covering 189 countries over 1998–2023.

Table AIII.3. Summary Statistics on Key Explanatory Variables
(In percent)

	1st Quartile	**Mean**	**Median**	**3rd Quartile**
Output gap				
All	-1.07	2.72	0.50	6.09
AEs	-1.88	-0.52	-0.57	0.31
EMMIEs	-1.11	2.09	0.46	4.29
LICs	3.87	7.91	8.35	12.25
FB gap				
All	-4.65	-2.23	-1.39	1.33
AEs	-3.64	-1.17	-1.03	1.62
EMMIEs	-4.81	-2.12	-1.34	1.44
LICs	-6.15	-3.74	-2.14	0.85
Debt-to-GDP				

All	29.86	52.28	46.58	68.03
AEs	36.76	61.28	56.87	83.46
EMMIEs	26.59	48.35	44.01	64.50
LICs	30.36	50.41	42.88	61.81

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: The table reports summary statistics for the key explanatory variables used in the baseline model. The dataset covers 189 countries over the period 1998–23. For countries without WEO-provided output gap estimates, we calculated output gaps using the Hamilton filter.

Output gaps combine IMF WEO estimates, when available, with Hamilton-filter estimates when WEO data are missing. Average output gaps are close to zero in AEs, moderately positive in EMMIEs, and substantially higher in LICs. Part of these differences reflects greater reliance on Hamilton-filter estimates in the latter groups: roughly 35 percent of all observations are Hamilton-based—2 percent for AEs, 30 percent for EMMIEs, and 84 percent for LICs. Because Hamilton-filter gaps tend to be smoother and slightly more positive than WEO estimates, this partly explains the higher means observed for EMMIEs and LICs. Although these methodological differences may introduce small artificial level discrepancies across countries, any resulting bias is likely limited (see Appendix Section V.A for robustness tests on the countercyclicality of IMF advice).

The fiscal balance gap measures the difference between the debt-stabilizing fiscal balance and actual fiscal balance. A negative value indicates that the observed fiscal balance is weaker than the level required to stabilize debt. Given the upward trend in public debt over much of the sample period, it is therefore not surprising that fiscal balance gaps are predominantly negative across income groups. This implies that, on average, fiscal deficits would need to be reduced—or surpluses increased—to prevent further rises in debt ratios. Debt ratios are highest in AEs and somewhat lower in EMMIEs and LICs.

D. Definition and Construction of the Fiscal Balance Gap

The fiscal balance gap (FB gap) measures the distance between a country's actual fiscal balance and the fiscal balance required to stabilize the debt-to-GDP ratio at a given level of debt d^*. Intuitively, it captures the size of fiscal adjustment needed to prevent debt from rising further. A negative gap implies that the actual fiscal balance is weaker than the debt-stabilizing level—meaning deficits would need to be reduced or surpluses increased to stabilize debt dynamics.

We build on the calculations performed by Kose and others (2022). Formally, their fiscal sustainability gap is given by:

$$fbsusgap = b - (\frac{-\gamma}{1+\gamma})d^*$$

where $\boldsymbol{b}$ is the overall fiscal balance, $\boldsymbol{\gamma}$ is the nominal output growth rate (a weighted average of the percent change in GDP expressed in local currency and in U.S. dollars at current exchange rates), and $\boldsymbol{d}^*$ is the target debt ratio defined as the historical median value of a country's peer groups of AEs and EMDEs.

For our purposes, we adapt this formulation to obtain a narrower measure that isolates the fiscal adjustment needed to reach the debt-stabilizing balance, rather than to maintain debt at a specific target level $\boldsymbol{d}^*$. We do so because our baseline specification already includes the debt-to-GDP ratio as a separate explanatory variable capturing cross-country differences in debt levels.

To reconstruct the debt-stabilizing balance, we first rearrange Kose and others (2022)'s formulation as follows:

$$(\frac{-\gamma}{1+\gamma}) = \frac{b - fbsusgap}{d^*}$$

and apply this expression to actual country-year debt data, replacing the target debt ratio d^* with the lagged actual debt ratio d_{t-2}. The debt-stabilizing fiscal balance at time t-1 is then:

$$b_{t-1}^{stab} = \left(\frac{-\gamma}{1+\gamma}\right) d_{t-2}$$

The term b_{t-1}^{stab} assumes debt remains constant between t-2 and t-1 (i.e., $d_{t-1} - d_{t-2} = 0$). Finally, we change the sign convention to define the fiscal balance gap as the difference between the debt-stabilizing balance and the actual balance, such that:

$$FB\ gap_{t-1} = b_{t-1}^{stab} - b_{t-1}$$

This sign convention ensures that higher values of the FB gap correspond to greater fiscal adjustment needs, facilitating interpretation of the regression results—larger positive gaps indicate that stronger fiscal consolidation would be required to stabilize debt dynamics.

APPENDIX IV. BASELINE ROBUSTNESS

A. Robustness to Data Vintage

Table AIV.1 reports the results of a robustness check using fixed-vintage data series to assess whether the use of different WEO vintages materially affects the estimated relationships. In this specification, the output gap and debt-to-GDP ratio are taken from the October 2024 WEO vintage and applied consistently across all years in the sample whenever available. When a historical value is missing from that vintage, the corresponding observation is drawn from the most recent earlier vintage.

The results are consistent with the baseline estimates reported in Table 3. The signs, magnitudes, and significance levels of the coefficients on the output gap, fiscal balance gap, and debt-to-GDP ratio remain broadly unchanged across specifications, confirming that the key determinants of IMF fiscal advice are robust to the use of ex-post rather than contemporaneous data. While later-vintage gaps are somewhat smoother—reflecting the incorporation of information unavailable to staff at the time of surveillance—these differences have only minor effects on the estimated relationships. Overall, the findings indicate that the main results are not sensitive to data vintage or real-time measurement issues.

Table AIV.1. Near-term Stance, Fixed Vintage

Variables	(1)	(2)	(3)	(4)
Output gap	3.705***	5.468***	3.712***	2.170
	(0.835)	(1.516)	(0.840)	(1.709)
FB gap	1.791**	3.347***	1.905***	7.834***
	(0.708)	(1.024)	(0.712)	(1.274)
Debt-to-GDP	1.160***	1.874***	1.185***	2.765***
	(0.176)	(0.368)	(0.178)	(0.452)
τ_1	0.455***	2.483	0.719**	2.660
	(0.101)	(1.398)	(0.283)	(1.452)
τ_2	1.342***	3.574**	1.613***	3.936***
	(0.108)	(1.399)	(0.286)	(1.454)
N	2741	2575	2741	2575
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.127	0.155	0.133	0.242
Pseudo R^2 (Ugba & Gertheiss)	0.283	0.338	0.295	0.493
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

B. Robustness to Alternative LLM Classifications

Table AIV.2 reports a robustness check assessing whether the results depend on the specific LLM classification used to determine the short-term fiscal stance. The baseline specification defines the stance through a majority-vote across three models—OpenAI's o1, Anthropic's Claude 3.7 Sonnet, and DeepSeek's R1. To test sensitivity, columns (2)–(4) replace the majority-vote classification with the output from each individual model in turn, while column (1) reproduces the baseline estimates from Table 2 of the paper for comparison.

The results show a high degree of consistency across all models. The coefficients on the output gap, fiscal balance gap, and debt-to-GDP ratio remain positive, statistically significant, and similar in magnitude to the baseline, confirming that the core relationships are not driven by the specific LLM used. The coefficient on the output gap remains the largest and most significant variable in every case—ranging from 5.33 to 6.61—indicating that stronger cyclical conditions robustly increase the likelihood of tightening advice. Overall, the findings demonstrate that the results are stable and reproducible across alternative LLM classifications, reinforcing confidence in the robustness of the baseline specification.

Table AIV.2. Near-term Stance, Different LLM classifications				
Variables	(1)	(2)	(3)	(4)
Output gap	5.671***	5.333***	5.380***	6.607***
	(0.951)	(0.923)	(0.933)	(1.035)
FB gap	1.752**	1.916***	1.308*	1.794**
	(0.726)	(0.713)	(0.728)	(0.762)
Debt-to-GDP	1.305***	1.355***	1.104***	1.202***
	(0.182)	(0.179)	(0.175)	(0.190)
τ_1	0.362***	0.278***	0.392***	0.573**
	(0.104)	(0.102)	(0.102)	(0.108)
τ_2	1.251***	1.185***	1.340**	1.440***
	(0.110)	(0.108)	(0.109)	(0.115)
N	2720	2722	2673	2709
Country FE	No	No	No	No
Year FE	No	No	No	No
Pseudo R^2 (McFadden)	0.145	0.145	0.140	0.140
Pseudo R^2 (Ugba & Gertheiss)	0.318	0.320	0.309	0.308
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. Column (1) reshows the estimates from Table 2 column (1) of the paper.

C. Proportional Odds Assumption

A central assumption of the ordered logit model is proportional odds, meaning that explanatory variables have the same effect on the odds of moving up the ordinal scale, regardless of the specific threshold. Table AIV.3 shows the results from the Brant test for proportional odds assumption. The omnibus test is an overall test that checks whether the proportional odds assumption holds for the model as a whole, rather than for each individual variable. Since the p-value is less than 5 percent significance level, the proportional odds assumption does not hold.

Table AIV.3. Results from Brant Test			
Test for	χ^2	**df**	**p-value**
Omnibus	22.729	3	4.6e-05***
Output gap	14.761	1	0.0001***
FB gap	0.487	1	0.485
Debt-to-GDP	2.997	1	0.083*

Source: IEO staff calculations.

To address this, we estimate a partial proportional odds model (Table AIV.4) that relaxes the constraint for variables where the assumption does not hold. The results show that the output gap effect is stronger at the first threshold (moving from Loosen to Neutral/Tighten) than at the second (from Neutral to Tighten), suggesting some variation in its influence across decision points. The FB gap becomes more influential at the second threshold, while debt-to-GDP maintains a consistent and significant effect across both. Despite minor differences, the qualitative results mirror those of the baseline model, confirming that the main conclusions are robust.

Table AIV.4. Baseline Regression, Partial Proportional Odds				
Variables	(1)	(2)	(3)	(4)
Output gap τ_1	8.744***	13.245***	8.738***	9.319***
	(1.384)	(2.113)	(1.386)	(2.322)
Output gap τ_2	5.254***	9.836***	5.263***	6.430***
	(0.950)	(1.799)	(0.956)	(2.002)
FB gap τ_1	1.314	2.482*	1.365	8.058***
	(0.950)	(1.332)	(0.951)	(1.536)
FB gap τ_2	1.853**	3.763***	1.992***	8.220***
	(0.737)	(1.082)	(0.743)	(1.341)
Debt-to-GDP τ_1	1.118***	2.168***	1.134***	2.988***
	(0.237)	(0.422)	(0.238)	(0.504)
Debt-to-GDP τ_2	1.340***	2.347***	1.364***	3.287***
	(0.184)	(0.380)	(0.186)	(0.463)
τ_1	0.358***	2.373	0.624**	2.514
	(0.105)	(1.399)	(0.287)	(1.449)
τ_2	1.291***	3.469**	1.567***	3.873***
	(0.132)	(1.404)	(0.99)	(1.455)
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

D. Alternative Estimators: Linear and Binary Specifications

For additional robustness, we estimate the baseline model using ordinary least squares (OLS) and a binary logit specification. In the OLS regression (Table AIV.5), the ordinal outcome is treated as a continuous variable (Loosen = 1, Neutral = 3, Tighten = 5). The signs and relative magnitudes of coefficients are consistent with the ordered logit results, reaffirming that stronger cyclical conditions and weaker fiscal positions are associated with tighter fiscal advice.

Table AIV.5. Baseline Regression, OLS

Variables	(1)	(2)	(3)	(4)
Output gap	3.346***	4.759***	2.776***	2.582***
	(0.453)	(0.869)	(0.531)	(0.821)
FB gap	0.922**	1.199**	1.857***	2.654***
	(0.381)	(0.551)	(0.550)	(0.613)
Debt-to-GDP	0.627***	0.853***	0.590***	0.885***
	(0.083)	(0.163)	(0.105)	(0.178)
Intercept	3.859***			
	(0.054)			
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Adjusted R^2	0.032	0.175	0.142	0.292

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The near-term fiscal stance is coded as 1 for easing, 3 for neutral, and 5 for tightening.

The binary logit model (Table AIV.6) collapses the outcome into a dichotomous variable (Tighten = 1; Loosen/Neutral = 0). The estimated coefficients and fit statistics remain closely aligned with the ordered logit estimates, confirming that the results are not sensitive to the functional form or to the ordinal coding of the dependent variable.

Table AIV.6. Baseline Regression, Binary Logit

Variables	(1)	(2)	(3)	(4)
Output gap	5.194***	9.082***	4.786***	5.747**
	(0.950)	(1.886)	(0.916)	(2.373)
FB gap	1.786**	3.264***	3.390***	7.634***
	(0.740)	(1.242)	(1.036)	(1.914)
Debt-to-GDP	1.339***	2.295***	1.472***	3.198***
	(0.185)	(0.462)	(0.230)	(0.664)
Intercept	0.358***			
	(0.105)			
N	2720	2481	2720	2481
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Squared correlation	0.032	0.195	0.114	0.294

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The squared correlation is the correlation between the model's fitted probabilities and the actual binary outcomes, serving as a pseudo-R^2 measure. For instance, a value of 0.294 indicates (column 4) that about 29 percent of the variation in observed outcomes is linearly associated with the model's predicted probabilities. In columns (2) and (4), the number of observations is lower than in the baseline ordered logit model because, after converting the three-category outcome (Loosen, Neutral, Tighten) into a binary variable (Tighten = 1, otherwise = 0), some fixed effects were dropped due to containing only one type of outcome (all 0s or all 1s).

APPENDIX V. ROBUSTNESS TO ALTERNATIVE MEASURES OF CYCLICAL CONDITIONS AND FISCAL VULNERABILITY

This section tests whether our main findings—regarding the countercyclicality of IMF fiscal advice and its sensitivity to fiscal risks—remain robust when alternative measures of key explanatory variables are used. We first consider alternative measures of cyclical conditions to test whether our results on the countercyclicality of IMF advice are robust. We then examine alternative measures of fiscal sustainability and vulnerability to assess the robustness of our findings on the sensitivity of IMF advice to fiscal risks. Finally, we examine whether the responsiveness of IMF advice differs systematically across countries facing different degrees of fiscal risk, drawing on the Fund's composite Vulnerability Exercise (VE) assessments.

A. Output Stabilization

(i) Alternative Measures of Cyclical Conditions

To test the robustness of our findings on the countercyclicality of IMF fiscal advice, we consider several alternative indicators of countries' cyclical positions beyond our baseline output gap measure. The baseline combines World Economic Outlook (WEO)—used whenever they were available at the time of each Article IV consultation—with a Hamilton-filter estimates that we compute from the corresponding vintage of the WEO real GDP series when no WEO output gap was reported. Because the real GDP series is available contemporaneously, the Hamilton-filter estimates are constructed using the vintage closest to the publication date of each Article IV report. This approach maximizes sample coverage and gives priority to the Fund's contemporaneous assessments.

That being said, no single measure of the output gap is perfect. To assess the robustness of our results, we test alternative ways of capturing cyclical conditions. Unlike the baseline, which blends WEO and Hamilton-filter gaps for wider coverage, the alternative measures rely on a single method throughout, trading coverage for methodological consistency. The first alternative uses WEO output gaps only, relying exclusively on the contemporaneous staff estimates available at the time of each Article IV report. This approach preserves the judgmental information embedded in Fund assessments but results in a smaller sample, especially for low-income countries (LICs). The second uses the Hamilton-filter output gap. This ensures methodological comparability and remains feasible even in data-poor environments, though it is subject to the usual limitations of statistical filters. The third indicator is the deviation of real growth from its five-year historical average, as in the IMF's Vulnerability Exercise and Kilic Celik and others (2023). This backward-looking measure captures whether current real GDP growth is above or below its recent trend, capturing growth momentum rather than the degree of slack. The fourth indicator is the deviation of actual growth from the five-year-ahead WEO projection, defined as the difference between the current real GDP growth rate and the rate projected five years into the future. Because IMF forecasts generally assume output gaps close within the forecast horizon, this forward-looking measure serves as a practical proxy for potential growth, as also used in Kose and others (2023).

(ii) Main Results

Table AV.1 presents the regression results obtained when replacing the baseline output gap with each of the four alternative cyclical indicators. Across all specifications, the results consistently confirm that IMF fiscal advice is countercyclical.

Table AV.1. Robustness: Alternative Measures of Cyclical Conditions

Variables	(1)	(2)	(3)	(4)
Output gap (WEO only)	11.685***			
	(2.324)			
Output gap (Hamilton filter)		4.882***		
		(0.971)		
Growth deviation from 5-year average			11.442***	
			(2.413)	
Growth deviation from 5-year ahead WEO projection				9.433***
				(1.897)
FB gap	2.821***	1.880**	0.852	1.106
	(0.975)	(0.730)	(0.720)	(0.726)
Debt-to-GDP	1.232***	1.331***	1.114***	1.317***
	(0.210)	(0.179)	(0.179)	(0.182)
τ_1	0.437***	0.145	0.593***	0.475***
	(0.122)	(0.122)	(0.099)	(0.102)
τ_2	1.278***	1.037***	1.484***	1.373***
	(0.129)	(0.126)	(0.107)	(0.109)
N	1767	2734	2732	2606
Country FE	No	No	No	No
Year FE	No	No	No	No
Pseudo R^2 (McFadden)	0.409	0.136	0.142	0.171
Pseudo R^2 (Ugba & Gertheiss)	0.725	0.301	0.312	0.369
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Using the WEO-only output gap (column 1) yields a large and statistically significant positive coefficient (11.69), indicating that as economic slack diminishes—that is, as output rises above potential—the Fund is more likely to recommend fiscal tightening. The Hamilton filter-based output gap (column 2) produces a smaller but still significant positive coefficient (4.88), supporting the same interpretation. The deviation of real GDP growth from its past five-year historical average (column 3) also shows a strong, positive, and statistically significant effect (11.44), and the deviation of real GDP growth from the five-year-ahead WEO projection (column 4) likewise yields a positive and statistically significant effect (9.43).

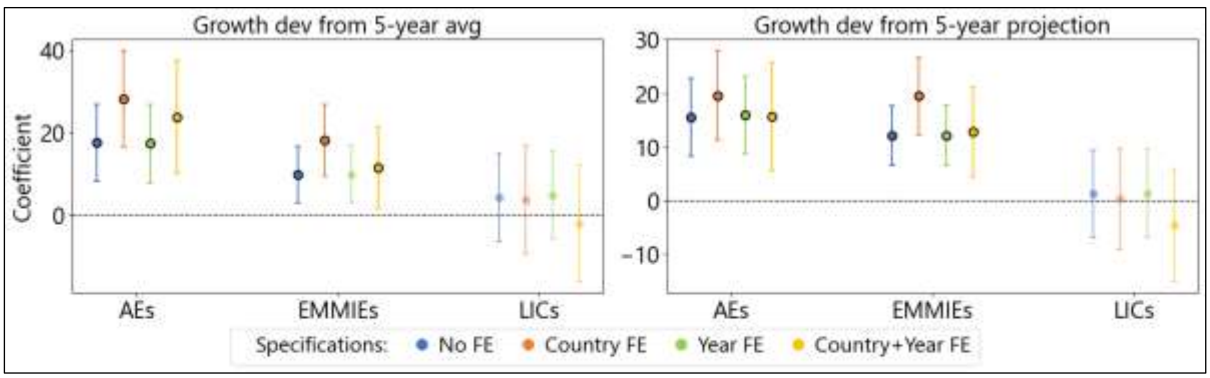
In all specifications, the fiscal balance gap and the debt-to-GDP ratio remain strong and statistically significant predictors of tighter fiscal advice, with coefficients generally ranging from

0.85 to 2.82 and from 0.88 to 1.33, respectively. These findings underscore that IMF fiscal recommendations reflect a combination of macroeconomic stabilization motives and fiscal sustainability considerations.

(iii) Results by Income Group

Building on the robustness exercises above, we next examine whether the countercyclical nature of IMF fiscal advice holds across income groups when using the alternative cyclical indicators. Figure AV.1 reports results obtained when the deviation of real growth from its five-year average and the deviation from the five-year-ahead WEO projection are interacted with income-group dummies.

Figure AV.1. Robustness: Income Group Heterogeneity with Alternative Measures of Cyclical Conditions



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: All regressions use contemporaneous data. Vertical lines indicate 95 percent confidence intervals. Detailed estimates corresponding to these results are reported in Table AVI.13. and Table AVI.14. in the Appendix Section VI.

For AEs, the interaction terms are consistently large and statistically significant across specifications, indicating that fiscal advice leans strongly against the cycle regardless of the cyclical measure used. For EMMIEs, countercyclical recommendations emerge in several specifications, particularly when country fixed effects are included, suggesting a more conditional responsiveness to the cycle that may reflect heterogeneity in market access or institutional capacity. For LICs, the coefficients on the cyclical interaction terms are generally positive but not statistically significant, implying weaker or less systematic countercyclical advice. Overall, these results show that while IMF advice is broadly countercyclical across all cyclical indicators, its strength varies with countries' income levels and economic resilience.

B. Fiscal Sustainability

(i) Alternative Measures of Fiscal Sustainability

The previous set of robustness checks examined whether our main findings on the countercyclicality of IMF fiscal advice hold when using alternative measures of cyclical conditions. We now test whether these results remain stable when fiscal sustainability indicators are defined more broadly. Our baseline analysis measures fiscal sustainability with two commonly used indicators—the fiscal balance gap and the debt-to-GDP ratio—which capture only a narrow dimension of fiscal space. As emphasized in IMF (2018), fiscal space is inherently multi-dimensional, reflecting a government's capacity to raise spending or reduce taxes without undermining debt sustainability or market access. We therefore expand the analysis to include additional fiscal vulnerability indicators, drawing on the cross-country fiscal space dataset by Kose and others (2022) supplemented with variables compiled independently. These indicators capture four main dimensions of fiscal risk.

The first dimension corresponds to our baseline solvency metrics, but employs alternative formulations from Kose and others (2022) that assess long-term debt sustainability. Debt related measures include general government gross debt-to-GDP ratio, debt as a share of average tax revenues, and debt relative to a 10-year moving average of GDP, which scale debt either by fiscal capacity or by a smoothed measure of the output base. On the fiscal balance side, we employ sustainability gap indicators that compare the actual or projected primary balance to the level required to stabilize debt under various macroeconomic scenarios (historical, current, stressed, and benign).

The second dimension captures liquidity and market access pressures, which reflect short-term financing risks and investor confidence. This dimension includes several complementary indicators drawn from the Kose and others (2022) dataset and additional sources. Indicators include gross financing needs (as a share of GDP), the effective nominal interest rate on government debt, debt service as a share of general government revenue, and the share of central government debt maturing within 12 months. To capture market sentiment and external financing conditions, we also include foreign-currency long-term sovereign credit ratings (1-21 scale, higher = stronger) and five-year sovereign credit default swap (CDS) spreads (in basis points). Together these proxies describe refinancing pressures and perceived creditworthiness—key determinants of IMF advice when liquidity risks rise.

The third dimension concerns balance sheet vulnerabilities related to debt composition, rollover risk, and external exposure. Indicators include the share of general government debt denominated in foreign currency, the share of debt securities held by nonresidents, and the total share of government debt held by nonresidents, all of which proxy for sensitivity to exchange rate and external funding shocks. We also consider the share of concessional external debt in total public external debt, which mitigates refinancing risk by providing more stable and predictable funding, and the average maturity of sovereign debt, which reflects the rollover profile of public liabilities. Together, these indicators provide a comprehensive view of how the composition and maturity structure of debt affect fiscal vulnerability and the potential for liquidity pressures to translate into solvency risks.

The fourth dimension encompasses external and private sector debt risks, which can generate contingent liabilities for the public sector and heighten macro-financial vulnerabilities. This set includes broad measures of external leverage—such as total external debt as a share of GDP and external debt in foreign currency as a share of total external debt—as well as indicators of private sector exposure, including private external debt and domestic credit to the private sector (both as shares of GDP). To capture near-term refinancing pressures, we also include short-term external debt as a share of total external debt, and several ratios comparing external debt to international reserves, both including and excluding gold holdings. These variables together capture the risk that external or private sector imbalances could spill over to the sovereign balance sheet, prompting the Fund to recommend a more conservative fiscal stance to mitigate systemic vulnerabilities.

The final robustness exercise takes a different approach from the preceding analyses. Rather than substituting or augmenting specific explanatory variables, we directly exploit the composite fiscal vulnerability indicators produced by the IMF's Vulnerability Exercise (VE) to examine how the relationship between macro-fiscal conditions and IMF fiscal advice varies with countries' overall level of fiscal fragility. Conceptually, this exercise is similar to the earlier heterogeneity analysis by income group, but here the sample is divided according to measured fiscal vulnerability rather than income level. The VE provides a model-based, cross-country framework for assessing near-term fiscal crisis risk across the Fund's membership. It produces a fiscal risk index estimating the probability of a fiscal crisis over a one- to two-year horizon, using a machine-learning model trained on historical data for 188 countries. A fiscal crisis is defined broadly to include sovereign defaults or restructurings, recourse to exceptional financing (for example, IMF lending above 100 percent of quota), implicit defaults through arrears or very high inflation, and sharp losses of market confidence leading to a sudden stop or large increases in spreads. The VE assessment combines over 100 indicators spanning fiscal, real, external, and institutional dimensions. Country-year observations are ranked by their estimated fiscal risk and categorized into three tiers: low vulnerability (below the 50th percentile), medium vulnerability (50th–80th percentile), and high vulnerability (above the 80th percentile). We interact these vulnerability tiers with the main explanatory variables from the baseline specification to test whether the responsiveness of IMF advice to cyclical and solvency conditions differs systematically across countries with varying levels of fiscal risk.

(ii) Results: Long-Run Sustainability

Replacing the baseline solvency indicators with Kose and others (2022) measures confirms that higher debt—whether scaled by GDP, trend GDP, or tax revenue—is associated with a greater likelihood of fiscal tightening advice.[8] The coefficients are positive and statistically significant. Sustainability gap measures yield similarly positive results, especially under stressed scenarios, where the fiscal adjustment required to stabilize debt is greatest. This suggests IMF advice is particularly sensitive to sustainability concerns in adverse environments, whereas under benign conditions the response is weaker. Overall, these findings bolster the interpretation that the IMF tailors its fiscal recommendations not only to observable debt levels but also to forward-looking

[8] Detailed estimates corresponding to these results are reported in Table AVI.15. in the Appendix Section VI.

assessments of debt stabilization needs. This responsiveness is especially pronounced when projected debt dynamics appear most precarious.

(iii) Results: Liquidity and Market Access

Gross Financing Needs (GFNs), calculated as the sum of the fiscal deficit and maturing debt over the following 12 months, serve as a key measure of rollover risk. Large GFNs indicate heavier refinancing burdens and greater vulnerability to interest rate or market shocks. The average interest rate on public debt, also from the IMF's Debt Sustainability Analysis (DSA) dataset, reflects the effective cost of borrowing faced by the public sector; higher values may signal investor concerns about liquidity or solvency, particularly in high-debt contexts. The debt service burden as percent of general government revenue, obtained from the IMF's internal WEO database, combines interest payments and principal repayments and captures near-term fiscal pressures associated with debt servicing. The share of central government debt maturing within 12 months as a percent of GDP provides a direct measure of short-term refinancing risk. In addition, sovereign CDS spreads, taken from Bloomberg and J.P. Morgan, serve as a real-time market indicator of perceived sovereign risk and rollover vulnerability, particularly relevant during episodes of financial stress. Finally, we include sovereign credit ratings as a more stable but comprehensive proxy for a country's creditworthiness. Using annual averages of long-term foreign-currency ratings from Moody's, Standard & Poor's, and Fitch (sourced from Bloomberg), we capture a broader set of risk factors, including institutional quality, macroeconomic fundamentals, and fiscal performance. Ratings are averaged across agencies to construct a composite index.

The results from our analysis confirm that market-based indicators of liquidity risk and creditworthiness are strong correlates of IMF fiscal advice.[9] Countries facing tighter liquidity constraints or perceived to be at greater risk of losing market access tend to receive tighter fiscal advice, even after controlling for solvency-related variables. The results also underscore the multidimensional nature of fiscal risk, with both flow and stock vulnerabilities playing an important role in shaping the Fund's recommendations. GFNs are positively and significantly associated with fiscal tightening advice. The estimated coefficient implies that countries facing larger near-term funding pressures tend to be advised to tighten fiscal policy, consistent with the idea that heavy rollover burdens heighten refinancing risk and warrant preemptive adjustment. This finding aligns with IMF guidance that highlights GFN thresholds as early warning indicators in Debt Sustainability Frameworks (DSFs). The effective nominal interest rate on public debt is also positively and significantly related to fiscal tightening advice. A higher average interest rate may signal greater risk premia or refinancing costs and, thus, heighten the urgency for fiscal correction. The large magnitude of the coefficient further suggests that high borrowing costs are viewed by the IMF as a key constraint on fiscal space and a channel through which market signals inform advice. Sovereign credit ratings and CDS spread, the two market perception

[9] Detailed estimates corresponding to these results are reported in Table AVI.16. in the Appendix Section VI.

variables, behave as expected. Lower credit ratings are associated with significantly tighter fiscal advice, and the coefficient is both large and highly significant. Similarly, wider CDS spreads are correlated with tighter advice. These results suggest that the IMF pays close attention to market perceptions of sovereign risk when formulating fiscal recommendations. This is consistent with the Fund's role in helping countries maintain or restore market access and manage vulnerabilities associated with adverse shifts in investor sentiment.

The only unexpected result concerns the debt service and the debt maturing variables, which enter the regression with a negative and significant coefficient. In principle, higher debt service and a higher share of debt maturing over the next 12 months should indicate greater liquidity pressure and hence be associated with tighter advice. One possible explanation for this anomaly is omitted variable bias: countries with higher debt service levels may also have stronger fundamentals or greater debt-carrying capacity (e.g., higher income, broader revenue base, more stable investor base), allowing the IMF to adopt a less conservative stance despite elevated repayment obligations. Alternatively, the Fund may expect that these countries can manage high debt service burdens through other means (e.g., liability management operations or donor support), weakening the link between this variable and fiscal advice.

(iv) Results: Balance Sheet Vulnerabilities

Among these five indicators, only the total share of debt held by nonresidents is significantly associated with tighter IMF fiscal stance advice.[10] This suggests that greater reliance on external creditors raises concerns about rollover risk and market confidence, prompting the IMF to favor a more conservative stance. Other indicators show limited importance. Neither foreign currency-denominated government debt nor the subset of debt securities held by nonresidents show significant effects. This suggests that the IMF may be responding more to aggregate external exposure than to the form or currency denomination of that debt (see next section). Similarly, concessional financing terms and average maturity of sovereign debt do not appear to influence fiscal stance advice in a systematic way. These findings imply that while balance sheet vulnerabilities are conceptually important, the IMF’s fiscal advice appears particularly attuned to indicators that signal potential loss of market access or external investor confidence, rather than those related to rollover structure or concessionality per se.

(v) Results: External and Private Sector Debt

The results offer nuanced insights into how these external and private-sector indicators relate to IMF fiscal advice.[11] Currency mismatch risk emerges as particularly important. External debt denominated in foreign currency is positively and significantly associated with tighter fiscal recommendations, suggesting that high levels of foreign currency debt heighten concerns about

[10] Detailed estimates corresponding to these results are reported in Table AVI.17. in the Appendix Section VI.

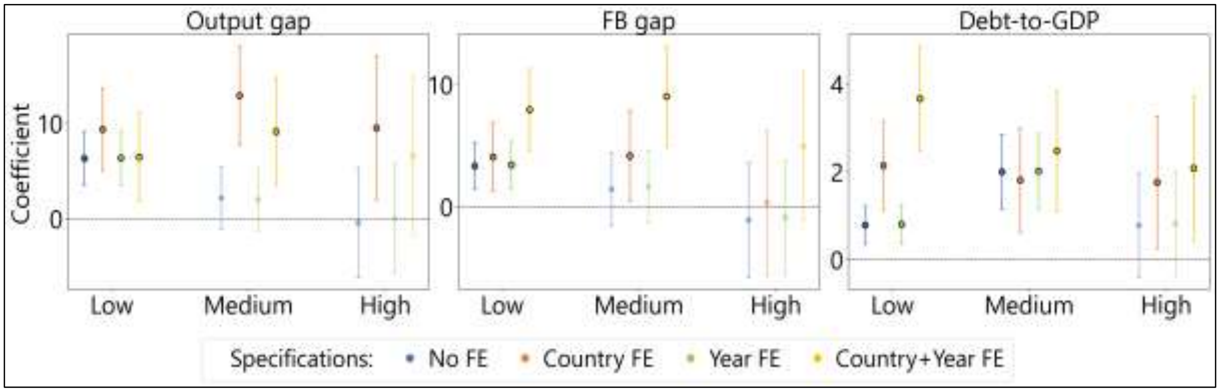
[11] Detailed estimates corresponding to these results are reported in Table AVI.19. in the Appendix Section VI.

balance sheet vulnerabilities. When a depreciation could sharply raise debt-servicing costs, IMF staff appear more likely to favor a more conservative near-term fiscal stance. In contrast, higher levels of total external debt as a share of GDP are associated with looser fiscal advice, a counterintuitive result that may reflect an omitted variable bias, such as differences in institutional quality or market access. Countries with stronger financial credibility may sustain higher external debt without triggering alarm, thereby diluting the role of gross debt as a red flag in Fund advice. Private sector indicators also show an unexpected pattern. Both private external debt and domestic credit to the private sector are negatively and significantly associated with tightening advice, suggesting that deeper or more developed financial systems are interpreted as signs of resilience and capacity to absorb shocks, reducing the perceived need for fiscal tightening. The relationship between short-term external debt and near-term fiscal advice is also inverse to conventional concerns about rollover risk. The share of short-term external debt in total external debt is negatively and significantly associated with tightening advice. One possible explanation is that countries with significant short-term borrowing may also have more active liquidity management frameworks or central bank backstops that mitigate this risk in the eyes of Fund staff. Finally, liquidity buffer indicators, such as external debt-to-reserve ratios, show no significant relationship with IMF's near-term fiscal stance advice. This suggests that these ratios, while commonly used in crisis prediction models, may not strongly influence IMF near-term fiscal recommendations.

(vi) Heterogeneity by Fiscal Vulnerability

The results shown in Figure AI.2 confirm that the IMF's fiscal advice response varies in important ways depending on the degree of underlying fiscal vulnerabilities. Interacting fiscal-vulnerability ratings with key explanatory variables reveals systematic differences in IMF advice. Countries facing higher risks are more likely to receive recommendations for fiscal tightening overall, consistent with the expectation that underlying vulnerabilities shape the Fund's advice even before observable cyclical conditions are taken into account.

Figure AV.2. Fiscal Vulnerability Heterogeneity in Determinants of Recommended Near-Term Fiscal Stance



Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Total effect of the interaction term is reported. All regressions use contemporaneous data. Vertical lines indicate 95 percent confidence intervals. Detailed estimates corresponding to these results are reported in Table AVI.19. in the Appendix Section VI.

Staff recommendations are more clearly countercyclical when fiscal vulnerabilities are lower. For country-years with low fiscal risk, a positive output gap is strongly associated with tighter fiscal advice. This relationship weakens as vulnerability increases. This relationship remains positive but smaller for medium-risk countries and becomes weaker and statistically insignificant when vulnerabilities are high. This pattern suggests that in more fragile contexts, cyclical signals play a smaller role in shaping advice, which likely reflects the overriding need to stabilize debt dynamics or rebuild fiscal buffers.

The relationship between fiscal advice and the fiscal balance gap is strongest when fiscal vulnerabilities are low. For countries with low fiscal vulnerability, a larger required adjustment to stabilize the debt-to-GDP ratio is strongly associated with tighter fiscal advice. In contrast, for countries classified as facing high fiscal vulnerabilities, the coefficients are either negative or statistically insignificant. This likely reflects the reality that highly vulnerable countries rarely have the flexibility to deviate from debt-stabilizing fiscal paths.

Debt levels remain a robust predictor of fiscal tightening advice across most vulnerability categories. For medium-risk countries, the association between debt and tightening recommendations is particularly strong and consistent, with coefficients ranging from 2.0 to 2.5 across specifications. For low-risk cases, debt levels also significantly influence fiscal advice. By contrast, the association between debt and fiscal recommendations weakens considerably for High-vulnerability country-years and becomes statistically insignificant in several specifications. This may reflect an omitted variable problem since countries facing high fiscal vulnerabilities typically do not accumulate very large debt stocks because market access is constrained and borrowing space is limited. As a result, observed debt levels in these settings may understate underlying risks, which reduces the explanatory power of the debt-to-GDP ratio in driving staff advice.

APPENDIX VI. REGRESSION TABLES

Table AVI.1. Income Group Heterogeneity in Determinants of Recommended Near-Term Fiscal Stance

Variables	(1)	(2)	(3)	(4)
Output gap × AEs	13.314***	13.316***	13.525***	11.092***
	(3.798)	(4.024)	(3.876)	(4.211)
Output gap × EMMIEs	6.450***	12.858***	6.474***	8.934***
	(1.561)	(2.717)	(1.565)	(2.947)
Output gap × LICs	1.363	7.421**	1.246	3.030
	(1.865)	(3.234)	(1.876)	(3.563)
FB gap × AEs	4.286***	1.178	4.468***	6.403**
	(1.564)	(2.235)	(1.578)	(2.493)
FB gap × EMMIEs	2.878***	7.617***	2.992***	12.558***
	(1.057)	(1.637)	(1.064)	(1.878)
FB gap × LICs	-1.045	0.338	-0.940	3.230
	(1.582)	(2.038)	(1.585)	(2.191)
Debt-to-GDP × AEs	0.721***	0.617	0.763***	2.692***
	(0.267)	(0.649)	(0.269)	(0.735)
Debt-to-GDP × EMMIEs	2.345***	3.120***	2.364***	4.321***
	(0.327)	(0.667)	(0.328)	(0.793)
Debt-to-GDP × LICs	1.651***	2.681***	1.657***	2.356***
	(0.483)	(0.724)	(0.485)	(0.793)
AEs	-0.022	1.367	-0.055	-0.534
	(0.337)	(1.071)	(0.339)	(1.156)
EMMIEs	-0.133	2.135	-0.149	1.198
	(0.325)	(1.621)	(0.326)	(1.668)
τ_1	0.368	0.035	0.662*	0.978
	(0.285)	(0.792)	(0.396)	(0.928)
τ_2	1.279***	1.141	1.580***	2.273**
	(0.288)	(0.793)	(0.398)	(0.930)
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.165	0.169	0.170	0.254
Pseudo R^2 (Ugba & Gertheiss)	0.357	0.365	0.367	0.512
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Total effect of the interaction term is reported. All regressions use contemporaneous data.

Table AVI.2. Changes in the Countercyclicality of Fiscal Advice Over Time

Variables	(1)	(2)	(3)
Output gap	5.884***	2.675*	3.259*
	(0.965)	(1.533)	(1.790)
FB gap	2.212***	2.556***	1.630**
	(0.725)	(0.731)	(0.753)
Debt-to-GDP	1.440***	1.363***	1.486***
	(0.187)	(0.184)	(0.193)
Year	-0.051***		
	(0.007)		
Post-GFC		-0.768***	
		(0.118)	
Output gap × post-GFC		7.315***	
		(1.191)	
Output gap × 2008–09			7.377***
			(2.267)
Output gap × 2010–12			0.722
			(2.230)
Output gap × 2013–19			7.870***
			(1.889)
Output gap × 2020–23			12.495***
			(2.858)
2008–09			-1.619***
			(0.197)
2010–12			-0.022
			(0.196)
2013–19			-0.705***
			(0.144)
2020–23			-1.546***
			(0.165)
τ_1	104.84***	0.847***	0.871***
	(14.799)	(0.131)	(0.139)
τ_2	105.75***	1.752***	1.820***
	(14.802)	(0.138)	(0.146)
N	2720	2720	2720
Country FE	No	No	No
Year FE	No	No	No
Pseudo R^2 (McFadden)	0.159	0.157	0.185
Pseudo R^2 (Ugba & Gertheiss)	0.346	0.342	0.394
Proportional odds assumption holds	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Total effect of the interaction term is reported. All regressions use contemporaneous data. The output gap estimate in the first row (3.259*) of column (3) refers to the output gap interacted with the year bucket for 1998–2007, the reference year bucket.

Table AVI.3. Output Gap and Monetary Stance at ZLB

Variables	(1)	(2)	(3)	(4)
Output gap	3.444**	8.169***	3.538**	7.073***
	(1.370)	(2.335)	(1.378)	(2.497)
Output gap × ZLB	5.784***	12.160***	5.506***	8.581***
	(1.603)	(2.566)	(1.621)	(2.754)
Monetary policy stance	0.364***	0.351***	0.363***	0.401***
	(0.056)	(0.065)	(0.057)	(0.071)
Monetary policy stance × ZLB	0.535***	0.680***	0.539***	0.596***
	(0.071)	(0.086)	(0.072)	(0.094)
ZLB	-1.192***	-1.746***	-1.193***	-3.756***
	(0.309)	(0.367)	(0.315)	(0.687)
FB gap	4.047***	6.696***	4.128***	10.360***
	(0.861)	(1.347)	(0.869)	(1.614)
Debt-to-GDP	1.254***	3.024***	1.275***	3.999***
	(0.210)	(0.480)	(0.212)	(0.559)
τ_1	-0.634***	1.286	-0.077	0.281
	(0.229)	(1.482)	(0.408)	(1.525)
τ_2	0.362	2.578*	0.928**	1.710
	(0.231)	(1.483)	(0.410)	(1.526)
N	2081	1946	2081	1946
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.359	0.235	0.363	0.287
Pseudo R^2 (Ugba & Gertheiss)	0.663	0.481	0.669	0.563
Proportional odds assumption holds	No	No	Yes	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The years 2009–15 and 2020–21 are coded as ZLB years, based on U.S. monetary policy conditions during which the federal funds rate was at or near the effective lower bound. This classification is used as a proxy for global monetary constraints given the central role of U.S. interest rates in shaping global financial conditions. The monetary stance is coded as 1 for easing, 3 for neutral, and 5 for tightening.

Table AVI.4. Macroprudential Stance				
Variables	(1)	(2)	(3)	(4)
Output gap	5.700***	10.351***	5.717***	6.895***
	(0.951)	(1.798)	(0.956)	(1.999)
FB gap	1.735**	3.462***	1.855**	8.134***
	(0.726)	(1.065)	(0.730)	(1.316)
Debt-to-GDP	1.310***	2.270***	1.332***	3.209***
	(0.182)	(0.376)	(0.183)	(0.459)
Macroprudential policy stance	0.074	0.066	0.068	0.084
	(0.048)	(0.058)	(0.049)	(0.063)
τ_1	0.061	2.151	0.342	2.249
	(0.222)	(1.415)	(0.351)	(1.469)
τ_2	0.950***	3.247**	1.238***	3.530**
	(0.224)	(1.416)	(0.352)	(1.471)
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.145	0.163	0.151	0.249
Pseudo R^2 (Ugba & Gertheiss)	0.320	0.353	0.330	0.504
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The stance for macroprudential measures is coded as 1 for easing, 3 for neutral/unclear, and 5 for tightening.

Table AVI.5. External Imbalances								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Output gap	4.821***	10.289***	4.842***	6.819***	4.634***	10.277***	4.643***	6.576***
	(0.990)	(1.854)	(0.994)	(2.058)	(1.000)	(1.858)	(1.004)	(2.062)
FB gap	-0.644	1.632	-0.568	6.382***	-0.399	1.650	-0.291	6.779***
	(0.807)	(0.153)	(0.814)	(1.431)	(0.827)	(1.173)	(0.834)	(1.451)
Debt-to-GDP	1.072***	2.271***	1.091***	3.127***	1.078***	2.273***	1.098***	3.174***
	(0.183)	(0.385)	(0.185)	(0.469)	(0.184)	(0.387)	(0.185)	(0.472)
CA deficit as % of GDP	5.635***	5.825***	5.594***	5.604***				
	(0.668)	(1.233)	(0.672)	(1.348)				
CA surplus only as % of GDP					-4.393***	-5.696***	-4.205***	-2.960
					(1.153)	(2.001)	(1.166)	(2.171)
CA deficit only as % of GDP					6.963***	5.941***	7.058***	8.072***
					(1.247)	(1.920)	(1.250)	(2.156)
τ_1	0.366***	2.089	0.657**	2.221	0.298**	2.082	0.588**	2.068
	(0.106)	(1.403)	(0.293)	(1.462)	(0.118)	(1.406)	(0.297)	(1.465)
τ_2	1.275***	3.193**	1.573***	3.512**	1.207***	3.187**	1.504***	3.360**
	(0.112)	(1.404)	(0.296)	(1.463)	(0.124)	(1.407)	(0.300)	(1.466)
N	2686	2524	2686	2524	2686	2524	2686	2524
Country FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.174	0.171	0.179	0.255	0.174	0.171	0.179	0.256
Pseudo R^2 (Ugba & Gertheiss)	0.374	0.368	0.383	0.514	0.375	0.368	0.384	0.515
Proportional odds assumption holds	No	No	No	No	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.6. External Imbalances by Income Group				
Variables	(1)	(2)	(3)	(4)
Output gap × AEs	10.637***	12.772***	10.688***	10.264**
	(3.936)	(4.201)	(4.015)	(4.313)
Output gap × EMMIEs	6.726***	12.880***	6.738***	8.783***
	(1.612)	(2.805)	(1.613)	(3.048)
Output gap × LICs	1.427	7.695**	1.329	2.949
	(1.906)	(3.340)	(1.916)	(3.665)
FB gap × AEs	-1.080	-0.112	-1.116	5.220**
	(1.817)	(2.316)	(1.841)	(2.605)
FB gap × EMMIEs	0.803	5.200***	0.901	10.660***
	(0.194)	(1.807)	(1.204)	(2.039)
FB gap × LICs	-1.247	-0.043	-1.089	1.798
	(1.688)	(2.175)	(1.694)	(2.353)
Debt-to-GDP × AEs	1.039***	0.974	1.074***	2.810***
	(0.291)	(0.676)	(0.293)	(0.759)
Debt-to-GDP × EMMIEs	1.916***	2.860***	1.937***	2.810***
	(0.341)	(0.675)	(0.343)	(0.759)
Debt-to-GDP × LICs	1.665***	2.724***	1.681***	2.224***
	(0.504)	(0.751)	(0.506)	(0.826)
CA deficit as % of GDP × AEs	11.450***	8.387***	11.492***	6.541**
	(1.587)	(2.894)	(1.602)	(3.042)
CA deficit as % of GDP × EMMIEs	4.098***	6.334***	4.037***	5.168***
	(1.001)	(1.775)	(1.005)	(1.923)
CA deficit as % of GDP × LICs	0.222	1.640	0.121	4.104*
	(1.456)	(2.265)	(1.457)	(2.485)
AEs	-0.104	1.340	-0.137	-0.507
	(0.347)	(1.081)	(0.349)	(1.162)
EMMIEs	-0.050	1.875	-0.137	0.987
	(0.329)	(1.629)	(0.349)	(1.672)
τ_1	0.345	0.027	0.676*	0.993
	(0.288)	(0.800)	(0.404)	(0.927)
τ_2	1.286***	1.143	1.624***	2.296**
	(0.291)	(0.802)	(0.406)	(0.929)
N	2686	2524	2686	2524
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.195	0.177	0.199	0.260
Pseudo R^2 (Ugba & Gertheiss)	0.412	0.380	0.420	0.521
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Total effect of the interaction term is reported. All regressions use contemporaneous data.

Table AVI.7. External Imbalances: Asymmetry				
Variables	(1)	(2)	(3)	(4)
Output gap × AEs	10.603***	12.715***	10.707***	10.035**
	(4.012)	(4.339)	(4.104)	(4.485)
Output gap × EMMIEs	5.919***	13.014***	5.930***	8.823***
	(1.645)	(2.817)	(1.646)	(3.067)
Output gap × LICs	1.313	7.556**	1.205	2.606
	(1.911)	(3.358)	(1.920)	(3.689)
FB gap × AEs	-0.635	0.826	-0.708	6.415**
	(1.867)	(2.381)	(1.892)	(2.670)
FB gap × EMMIEs	1.593	4.935***	1.745	10.572***
	(1.248)	(1.856)	(1.257)	(2.101)
FB gap × LICs	-0.952	0.224	-0.784	2.247
	(1.720)	(2.191)	(1.726)	(2.356)
Debt-to-GDP × AEs	1.072***	1.085	1.116***	3.066***
	(0.292)	(0.683)	(0.295)	(0.769)
Debt-to-GDP × EMMIEs	1.950***	2.820***	1.971***	3.951***
	(0.344)	(0.676)	(0.346)	(0.806)
Debt-to-GDP × LICs	1.702***	2.750***	1.722***	2.191***
	(0.507)	(0.756)	(0.510)	(0.831)
CA surplus only as % of GDP × AEs	-9.471***	-1.772	-9.538***	2.962
	(2.210)	(4.070)	(2.228)	(4.373)
CA surplus only as % of GDP × EMMIEs	-1.018	-7.686***	-0.812	-5.329*
	(1.769)	(2.836)	(1.777)	(3.044)
CA surplus only as % of GDP × LICs	6.012	3.116	6.870	4.044
	(4.395)	(5.544)	(4.453)	(5.591)
CA deficit only as % of GDP × AEs	16.370***	19.310***	16.349***	21.067***
	(4.252)	(6.010)	(4.242)	(6.186)
CA deficit only as % of GDP × EMMIEs	7.047***	5.011*	7.091***	5.114*
	(1.792)	(2.782)	(1.796)	(3.040)
CA deficit only as % of GDP × LICs	2.496	3.630	2.708	8.179**
	(2.064)	(3.084)	(2.090)	(3.499)
AEs	-0.057	1.210	-0.066	-0.762
	(0.384)	(1.097)	(0.386)	(1.173)
EMMIEs	-0.019	2.095	-0.017	1.216
	(0.357)	(1.639)	(0.360)	(1.686)
τ_1	0.148	-0.112	0.474	0.714
	(0.310)	(0.813)	(0.418)	(0.940)
τ_2	1.090***	1.005	1.424***	2.025**
	(0.313)	(0.814)	(0.420)	(0.942)
N	2686	2524	2686	2524
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.197	0.179	0.202	0.264
Pseudo R^2 (Ugba & Gertheiss)	0.416	0.384	0.425	0.528
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Total effect of the interaction term is reported. All regressions use contemporaneous data.

Table AVI.8. Countercyclicality by Exchange Rate Regime

Variables	(1)	(2)	(3)	(4)
Output gap × Flexible	11.933***	14.534***	12.257***	8.159**
	(2.357)	(2.478)	(2.376)	(3.566)
Output gap × Intermediate	4.014***	10.325***	3.940**	6.517**
	(1.521)	(2.761)	(1.541)	(2.754)
Output gap × Fixed	0.470	10.053***	0.318	6.535*
	(1.866)	(3.338)	(1.882)	(3.622)
Flexible	-1.082***	-1.356***	-1.109***	-0.573
	(0.149)	(0.339)	(0.150)	(0.373)
Intermediate	-0.501***	-0.754*	-0.499***	-0.484
	(0.167)	(0.386)	(0.168)	(0.411)
FB gap	2.523***	4.353***	2.667***	8.645***
	(0.824)	(1.185)	(0.830)	(1.476)
Debt-to-GDP	1.551***	2.476***	1.582***	3.311***
	(0.202)	(0.433)	(0.204)	(0.520)
τ_1	0.832***	3.418	1.081***	2.065
	(0.152)	(1.444)	(0.325)	(1.496)
τ_2	1.739***	4.531***	1.998***	3.370**
	(0.158)	(1.446)	(0.328)	(1.498)
N	2325	2182	2325	2182
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.261	0.176	0.267	0.262
Pseudo R^2 (Ugba & Gertheiss)	0.523	0.378	0.533	0.525
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IMF AREAER Database; IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.9. Countercyclicality by Exchange Rate Regime and Income Group

Variables	(1)	(2)	(3)	(4)
Output gap × AE × Flexible	23.464***	28.772***	24.190***	22.820***
	(6.094)	(7.213)	(6.227)	(7.602)
Output gap × EMMIE × Flexible	9.879**	7.383	10.366**	-1.022
	(4.012)	(5.380)	(4.036)	(5.661)
Output gap × LIC × Flexible	7.761*	9.391	7.989*	4.016
	(4.333)	(6.496)	(4.359)	(6.997)
Output gap × AE × Intermediate	11.196**	15.775**	11.464**	12.183*
	(5.361)	(7.203)	(5.588)	(7.107)
Output gap × EMMIE × Intermediate	5.592*	9.058**	5.612*	6.226
	(2.867)	(4.602)	(2.867)	(4.904)
Output gap × LIC × Intermediate	1.466	9.170**	1.128	3.223
	(2.404)	(4.123)	(2.445)	(4.509)
Output gap × AE × Fixed	6.008	7.921	5.799	4.998
	(6.924)	(7.712)	(6.957)	(7.820)
Output gap × EMMIE × Fixed	1.867	13.056***	1.835	7.259
	(2.689)	(4.514)	(2.708)	(4.722)
Output gap × LIC × Fixed	0.407	7.262	0.519	9.560
	(4.485)	(6.655)	(4.577)	(7.849)
FB gap	3.008***	4.467***	3.158***	8.656***
	(0.843)	(1.211)	(0.849)	(1.502)
Debt-to-GDP	1.777***	2.529***	1.816***	3.478***
	(0.211)	(0.437)	(0.213)	(0.530)
AE	0.180	0.663	0.323	0.033
	(0.488)	(2.079)	(0.497)	(2.204)
EMMIE	0.293	-0.895	0.323	-1.962
	(0.443)	(2.747)	(0.451)	(2.806)
Flexible	-0.730	-1.357	-0.713	-0.124
	(0.568)	(1.824)	(0.579)	(1.976)
Intermediate	-0.135	-0.916	-0.076	0.274
	(0.485)	(1.738)	(0.494)	(1.878)
τ_1	0.543	1.370	0.812	0.076
	(0.420)	(1.910)	(0.515)	(2.057)
τ_2	1.464***	2.495	1.743***	1.404
	(0.422)	(1.911)	(0.517)	(2.058)
N	2325	2182	2325	2182
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.270	0.181	0.276	0.270
Pseudo R^2 (Ugba & Gertheiss)	0.538	0.388	0.547	0.537
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IMF AREAER Database; IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.10. Fiscal Institutions				
Variables	(1)	(2)	(3)	(4)
Output gap	4.330***	10.590***	4.344***	7.064***
	(0.985)	(1.847)	(0.990)	(2.006)
FB gap	2.184***	3.944***	2.303***	7.761***
	(0.743)	(1.086)	(0.749)	(1.315)
Debt-to-GDP	1.500***	3.051***	1.518***	3.427***
	(0.189)	(0.414)	(0.190)	(0.469)
Fiscal Rules Only	-0.294**	-0.757***	-0.295**	-0.336
	(0.116)	(0.265)	(0.117)	(0.296)
Fiscal Councils Only	-0.477*	-0.389	-0.506*	0.209
	(0.270)	(0.475)	(0.270)	(0.525)
Both Fiscal Rules and Councils	-0.891***	-2.225***	-0.891***	-1.216***
	(0.150)	(0.332)	(0.151)	(0.409)
τ_1	0.580***	3.924	0.851***	3.526
	(0.121)	(1.427)	(0.295)	(1.504)
τ_2	1.481***	5.051***	1.758***	4.815***
	(0.127)	(1.429)	(0.298)	(1.506)
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.154	0.181	0.160	0.253
Pseudo R^2 (Ugba & Gertheiss)	0.337	0.386	0.347	0.511
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); Alonso and others (2025a and 2025b); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The fiscal institutions variables are constructed based on the presence of fiscal rules and fiscal councils, based on the IMF Fiscal Rules and Fiscal Councils Datasets (Alonso and others, 2025a; 2025b). For each country-year observation, a dummy variable for fiscal rules and another for fiscal councils are coded as 1 if present and 0 otherwise. These are then combined into a four-category indicator: None (no rule or council), Rule only (rule without council), Council only (council without rule), and Both (presence of both). "None" serves as the reference category in the regressions, meaning that the coefficients for the other categories represent their effects relative to countries with neither a rule nor a council.

Table AVI.11. Fiscal Institutions Interaction				
Variables	(1)	(2)	(3)	(4)
Output gap × No Fiscal Institution	2.570**	8.737***	2.480**	5.817**
	(1.255)	(2.281)	(1.257)	(2.487)
Output gap × Fiscal Institution	7.646***	12.481***	7.820***	8.214***
	(1.597)	(2.505)	(1.618)	(2.652)
FB gap × No Fiscal Institution	2.548**	4.756***	2.691***	8.150***
	(1.000)	(1.437)	(1.004)	(1.624)
FB gap × Fiscal Institution	1.619	2.616	1.701	8.038***
	(1.098)	(1.603)	(1.106)	(1.847)
Debt-to-GDP × No Fiscal Institution	1.987***	3.066***	2.010***	3.086***
	(0.333)	(0.554)	(0.334)	(0.610)
Debt-to-GDP × Fiscal Institution	1.170***	1.824***	1.196***	3.354***
	(0.226)	(0.502)	(0.228)	(0.591)
Fiscal Institution	-0.227	-0.477	-0.239	-0.327
	(0.214)	(0.372)	(0.215)	(0.402)
τ_1	0.468***	2.912	0.776**	2.796
	(0.166)	(1.422)	(0.318)	(1.484)
τ_2	1.365***	4.016***	1.680***	4.077***
	(0.171)	(1.423)	(0.321)	(1.486)
N	2720	2556	2720	2556
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.153	0.168	0.159	0.249
Pseudo R^2 (Ugba & Gertheiss)	0.335	0.364	0.346	0.504
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); Alonso and others (2025a and 2025b); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The fiscal institutions variables are constructed based on the presence of fiscal rules and fiscal councils, based on the IMF Fiscal Rules and Fiscal Councils Datasets (Alonso and others, 2025a; 2025b). For each country-year observation, a dummy variable takes the value 1 if there is a presence of any fiscal rule or council, and 0 if neither exists.

Table AVI.12. Non-Economic Factors: Saltwater versus Freshwater Schools

Variables	(1)	(2)	(3)	(4)
Output gap	3.438***	6.875***	3.382**	4.624*
	(1.225)	(2.418)	(1.238)	(2.586)
FB gap	4.552***	8.859***	4.788***	9.644***
	(0.891)	(1.534)	(0.908)	(1.763)
Debt-to-GDP	1.349***	3.603***	1.375***	2.859***
	(0.222)	(0.578)	(0.224)	(0.614)
Output gap × Saltwater universities	10.296***	12.861***	10.455***	11.132***
	(3.276)	(4.148)	(3.489)	(4.156)
Output gap × Freshwater universities	7.072**	7.546*	7.342**	6.395
	(3.276)	(4.379)	(3.259)	(4.689)
Saltwater universities	-0.131	0.290	-0.123	0.198
	(0.162)	(0.200)	(0.164)	(0.206)
Freshwater universities	-0.356**	-0.237	-0.360**	-0.355
	(0.179)	(0.224)	(0.182)	(0.234)
τ_1	0.702***	1.737	1.187***	1.156
	(0.131)	(1.436)	(0.380)	(1.459)
τ_2	1.654***	2.963**	2.151***	2.515
	(0.141)	(1.438)	(0.385)	(1.461)
N	1995	1873	1995	1873
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.167	0.214	0.176	0.267
Pseudo R^2 (Ugba & Gertheiss)	0.361	0.446	0.378	0.533
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); Lang, Wellner, and Kentikelenis (2024); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Total effect of the interaction term is reported. All regressions use contemporaneous data. The result is based on data up to 2016, as information on mission chiefs is only available through that year.

Table AVI.13. Robustness: Income Group Heterogeneity with Alternative Measures of Cyclical Conditions

Variables	(1)	(2)	(3)	(4)
5-year average growth deviation × AEs	17.462***	28.066***	17.286***	23.644***
	(4.809)	(5.969)	(4.870)	(7.024)
5-year average growth deviation × EMMIEs	9.811***	18.125***	9.807	11.427**
	(3.511)	(4.424)	(3.536)	(5.034)
5-year average growth deviation × LICs	4.217	3.645	4.864	-2.052
	(5.422)	(6.727)	(5.443)	(7.173)
FB gap × AEs	1.554	-3.322	1.819	3.237
	(1.500)	(2.257)	(1.511)	(2.513)
FB gap × EMMIEs	2.361**	5.323	2.460**	11.520***
	(1.040)	(1.527)	(1.044)	(1.783)
FB gap × LICs	-1.151	-0.796	-1.057	3.579*
	(1.610)	(2.003)	(1.613)	(2.163)
Debt-to-GDP × AEs	0.496*	-0.040	0.527*	2.266***
	(0.267)	(0.689)	(0.269)	(0.773)
Debt-to-GDP × EMMIEs	2.101***	2.143***	2.119***	3.713***
	(0.320)	(0.631)	(0.321)	(0.776)
Debt-to-GDP × LICs	1.648***	2.296***	1.636***	2.427***
	(0.490)	(0.707)	(0.491)	(0.792)
AEs	-0.064	0.556	-0.089	-0.897
	(0.294)	(1.021)	(0.296)	(1.091)
EMMIEs	-0.002	1.578	-0.014	1.076
	(0.278)	(1.576)	(0.279)	(1.619)
τ_1	0.467**	0.871	0.718**	1.246
	(0.232)	(0.719)	(0.360)	(0.834)
τ_2	1.382***	1.989***	1.642***	2.544***
	(0.236)	(0.721)	(0.363)	(0.836)
N	2732	2568	2732	2568
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.163	0.170	0.168	0.253
Pseudo R^2 (Ugba & Gertheiss)	0.354	0.367	0.364	0.511
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.14. Robustness: Income Group Heterogeneity with Alternative Measures of Cyclical Conditions

Variables	(1)	(2)	(3)	(4)
Growth deviation from 5-year ahead WEO projection × AEs	15.531***	19.565***	15.958***	15.642***
	(3.697)	(4.269)	(3.729)	(5.141)
Growth deviation from 5-year ahead WEO projection × EMMIEs	12.136***	19.457***	12.107***	12.744***
	(2.835)	(3.656)	(2.878)	(4.258)
Growth deviation from 5-year ahead WEO projection × LICs	1.245	0.235	1.320	-4.572
	(4.165)	(4.786)	(4.211)	(5.378)
FB gap × AEs	3.248**	-0.492	3.536**	5.261**
	(1.487)	(2.152)	(1.499)	(2.445)
FB gap × EMMIEs	2.214**	6.577***	2.348**	12.883***
	(1.069)	(1.611)	(1.078)	(1.880)
FB gap × LICs	-1.109	-0.754	-0.976	3.258
	(1.606)	(2.023)	(1.612)	(2.201)
Debt-to-GDP × AEs	0.621**	0.361	0.659**	2.579***
	(0.267)	(0.662)	(0.270)	(0.756)
Debt-to-GDP × EMMIEs	2.491***	2.049***	2.519***	3.581***
	(0.331)	(0.639)	(0.333)	(0.773)
Debt-to-GDP × LICs	1.632***	2.277***	1.633***	2.361***
	(0.483)	(0.727)	(0.483)	(0.816)
AEs	-0.129	0.569	-0.154	-0.976
	(0.2933)	(1.022)	(0.296)	(1.089)
EMMIEs	-0.222	1.577	-0.234	1.044
	(0.279)	(1.583)	(0.281)	(1.619)
τ_1	0.478**	0.857	0.770**	1.396*
	(0.231)	(0.720)	(0.360)	(0.835)
τ_2	1.407***	1.985***	1.709***	2.721***
	(0.235)	(0.722)	(0.363)	(0.838)
N	2606	2454	2606	2454
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.195	0.170	0.202	0.257
Pseudo R^2 (Ugba & Gertheiss)	0.413	0.366	0.425	0.517
Proportional odds assumption holds	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.15. Robustness: Alternative Measures of Solvency

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Output gap	4.558***	5.516***	5.851***	7.150***	7.060***	8.683***	7.295***	7.051***
	(0.950)	(0.955)	(0.951)	(1.242)	(1.242)	(1.579)	(1.307)	(1.243)
FB gap	1.930***	2.654***						
	(0.740)	(0.745)						
Debt-to-GDP			1.088***	1.421***	1.416***	1.469***	1.357***	1.407***
			(0.181)	(0.222)	(0.223)	(0.241)	(0.227)	(0.230)
General government gross debt as % of average tax revenues	0.074***							
	(0.019)							
General government gross debt as % of 10-year moving average GDP		0.952***						
		(0.130)						
Sustainability gap, fiscal balance as % of GDP			1.796**					
			(0.862)					
Sustainability gap, primary balance, country-specific conditions, % of GDP				4.515***				
				(1.699)				
Sustainability gap, primary balance, historical conditions, % of GDP					4.538***			
					(1.674)			
Sustainability gap, primary balance, current conditions, % of GDP						3.762***		
						(1.352)		
Sustainability gap, primary balance, stressed conditions, % of GDP							8.238***	
							(1.533)	
Sustainability gap, primary balance, benign conditions, % of GDP								0.420
								(1.430)
τ_1	0.778***	0.359***	0.455***	0.144	0.158	0.050	-0.156	0.169
	(0.080)	(0.101)	(0.108)	(0.125)	(0.126)	(0.138)	(0.139)	(0.149)
τ_2	1.666***	1.260***	1.344***	1.061***	1.082***	0.970***	0.769***	1.081***
	(0.090)	(0.107)	(0.115)	(0.131)	(0.132)	(0.143)	(0.143)	(0.154)
N	2691	2713	2765	1861	1860	1580	1838	1837
Country FE	No	No	No	No	No	No	No	No
Year FE	No	No	No	No	No	No	No	No
Pseudo R^2 (McFadden)	0.137	0.143	0.132	0.379	0.381	0.455	0.393	0.383
Pseudo R^2 (Ugba & Gertheiss)	0.303	0.315	0.294	0.688	0.691	0.774	0.705	0.694
Proportional odds assumption holds	No	No	No	No	No	No	Yes	Yes

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.16. Robustness: Indicators of Liquidity and Market Access Measures

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Output gap	5.935***	7.054***	2.527**	10.506***	5.624***	13.028***
	(1.053)	(1.047)	(1.068)	(2.128)	(1.370)	(3.015)
FB gap	-0.173	1.317*	1.791**	1.177	2.494***	7.863***
	(0.972)	(0.793)	(0.879)	(1.299)	(0.921)	(1.918)
Debt-to-GDP	0.843***	1.554***	1.928***	1.791***	1.279***	1.868***
	(0.241)	(0.203)	(0.254)	(0.345)	(0.210)	(0.322)
Gross financing needs	3.354***					
	(0.864)					
Effective nominal interest rate		9.884***				
		(2.310)				
Debt service as % of general government revenue			-0.334**			
			(0.138)			
Central government debt maturing in 12 months or less as % of GDP				-2.643*		
				(1.535)		
Foreign currency long-term sovereign debt ratings (index from 1 to 21 [best])					-0.089***	
					(0.012)	
5-year sovereign CDS spreads (basis points)						41.265***
						(7.353)
τ_1	0.240**	-0.289*	0.490***	-0.134	1.547***	-0.725***
	(0.118)	(0.158)	(0.142)	(0.177)	(0.209)	(0.208)
τ_2	1.109***	0.582***	1.494***	0.744***	2.420***	0.212***
	(0.124)	(0.160)	(0.152)	(0.182)	(0.215)	(0.209)
N	2211	2284	2009	931	2011	867
Country FE	No	No	No	No	No	No
Year FE	No	No	No	No	No	No
Pseudo R^2 (McFadden)	0.301	0.274	0.424	0.668	0.343	0.694
Pseudo R^2 (Ugba & Gertheiss)	0.585	0.544	0.741	0.933	0.647	0.945
Proportional odds assumption holds	No	No	No	No	Yes	No

Sources: WEO; Kose and others (2022); DSA/DSF database; IEO staff calculations.

Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.17. Robustness: Indicators of Balance Sheet Vulnerabilities

	(1)	(2)	(3)	(4)	(5)
Output gap	9.342***	15.506***	9.442***	4.798***	11.777***
	(3.429)	(4.463)	(2.710)	(1.268)	(3.086)
FB gap	2.385	3.849**	1.522	1.250	0.641
	(2.558)	(1.675)	(1.596)	(1.188)	(1.436)
Debt-to-GDP	2.388***	0.835***	1.821***	2.464***	1.177***
	(0.452)	(0.284)	(0.306)	(0.326)	(0.324)
General government debt in foreign currency as % of total	0.583				
	(0.411)				
Debt securities held by nonresidents as % of total		0.795			
		(2.698)			
General government debt held by nonresidents as % of total			1.390***		
			(0.399)		
Concessional external debt stocks as % of external public debt				-0.246	
				(0.276)	
Sovereign debt average maturity in years at T					-0.044
					(0.032)
τ_1	-0.609**	0.039	-0.843***	-0.173	0.280
	(0.289)	(0.197)	(0.242)	(0.210)	(0.305)
τ_2	0.386	0.907***	0.053	0.842***	1.209***
	(0.293)	(0.203)	(0.242)	(0.216)	(0.310)
N	543	668	941	1453	696
Country FE	No	No	No	No	No
Year FE	No	No	No	No	No
Pseudo R^2 (McFadden)	0.811	0.742	0.648	0.572	0.737
Pseudo R^2 (Ugba & Gertheiss)	0.983	0.963	0.923	0.874	0.962
Proportional odds assumption holds	Yes	No	No	Yes	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.18. Robustness: Indicators of External and Private Sector Debt Vulnerabilities

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Output gap	5.876***	15.394***	5.795***	2.486**	4.173***	6.822***	7.176***	7.085***
	(1.068)	(4.542)	(1.179)	(1.061)	(1.061)	(1.120)	(1.147)	(1.146)
FB gap	1.402	5.612**	2.020**	2.330***	1.068	2.135**	2.040**	2.034**
	(0.898)	(2.708)	(0.971)	(0.786)	(0.913)	(0.915)	(0.927)	(0.922)
Debt-to-GDP	1.643***	2.218***	1.471***	1.599***	1.527***	1.369***	1.362***	1.383***
	(0.220)	(0.530)	(0.227)	(0.195)	(0.211)	(0.216)	(0.218)	(0.219)
Total external debt stocks as % of GDP	-0.187***							
	(0.046)							
External debt in foreign currency as % of total		1.673***						
		(0.427)						
Private external debt stocks as % of GDP			-0.160***					
			(0.043)					
Domestic credit to private sector as % of GDP				-0.755***				
				(0.083)				
Short-term external debt stocks as % of total					-1.610***			
					(0.344)			
Short-term external debt stocks as % of reserves						-0.001		
						(0.004)		
Total external debt stocks as % of reserves							0.0006	
							(0.001)	
Total external debt stocks as % of reserves excl. gold								-0.0001
								(0.0007)
τ_1	0.258**	-1.777***	0.251**	0.831***	0.548***	0.223*	0.203	0.208*
	(0.121)	(0.487)	(0.127)	(0.123)	(0.142)	(0.124)	(0.124)	(0.124)
τ_2	1.153***	-0.897*	1.127***	1.760***	1.449***	1.106***	1.083***	1.087***
	(0.127)	(0.481)	(0.133)	(0.130)	(0.148)	(0.129)	(0.130)	(0.130)
N	2216	465	1852	2567	2233	2020	2014	2015
Country FE	No	No	No	No	No	No	No	No
Year FE	No	No	No	No	No	No	No	No
Pseudo R^2 (McFadden)	0.290	0.818	0.380	0.203	0.380	0.342	0.344	0.343
Pseudo R^2 (Ugba & Gertheiss)	0.568	0.984	0.690	0.427	0.691	0.641	0.644	0.643
Proportional odds assumption holds	No	Yes	No	No	No	No	No	No

Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

Table AVI.19. Interaction with Vulnerability Exercise Assessment for Fiscal at T

Variables	(1)	(2)	(3)	(4)
Output gap × Low	6.337***	9.361***	6.355***	6.483***
	(1.433)	(2.212)	(1.440)	(2.391)
Output gap × Medium	2.204	12.899***	2.052	9.129***
	(1.666)	(2.631)	(1.684)	(2.901)
Output gap × High	-0.418	9.513**	0.045	6.594
	(2.938)	(3.853)	(2.938)	(4.220)
FB gap × Low	3.339***	4.084***	3.423***	7.926***
	(0.973)	(1.438)	(0.983)	(1.688)
FB gap × Medium	1.439	4.167**	1.642	9.031***
	(1.508)	(1.907)	(1.512)	(2.080)
FB gap × High	-1.097	0.324	-0.871	4.927
	(2.409)	(3.012)	(2.381)	(3.096)
Debt-to-GDP × Low	0.790***	2.139***	0.807***	3.667***
	(0.222)	(0.514)	(0.224)	(0.608)
Debt-to-GDP × Medium	1.998***	1.808***	2.018***	2.477***
	(0.436)	(0.605)	(0.437)	(0.702)
Debt-to-GDP × High	0.777	1.764**	0.826	2.081**
	(0.608)	(0.772)	(0.614)	(0.838)
Medium	0.064	0.588	0.072	1.315***
	(0.276)	(0.403)	(0.278)	(0.444)
High	1.087**	1.080**	1.045**	1.480**
	(0.432)	(0.545)	(0.436)	(0.596)
τ_1	0.388***	2.305	0.616**	2.215
	(0.124)	(1.411)	(0.297)	(1.472)
τ_2	1.297***	3.410**	1.532***	3.510**
	(0.129)	(1.412)	(0.299)	(1.473)
N	2707	2543	2707	2543
Country FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Pseudo R^2 (McFadden)	0.164	0.168	0.169	0.256
Pseudo R^2 (Ugba & Gertheiss)	0.355	0.463	0.364	0.515
Proportional odds assumption holds	No	No	Yes	No

Sources: WEO; Kose and others (2022); Vulnerability Exercise (VE) assessment; IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data.

REFERENCES

Adler, Gustavo and others, 2024, “An Updated Action-based Dataset of Fiscal Consolidation,” IMF Working Paper No. 2024/210, (Washington: International Monetary Fund).

Aguiar, Mark, and Gita Gopinath. 2006. “Defaultable Debt, Interest Rates and the Current Account.” Journal of International Economics, 69 (1): 64–83.

Alesina, Alberto, and others, 2018, “Is It the ‘How or the When’ That Matters in Fiscal Adjustment?” IMF Economic Review 66: 144–88, (Washington: International Monetary Fund).

Alonso, Virginia, and others, 2025a, "Fiscal Rules Dataset: 1985-2024", (Washington: International Monetary Fund).

————, 2025b, "Fiscal Councils Dataset: The 2024 Update", (Washington: International Monetary Fund).

Arellano, C., 2008, “Default Risk and Income Fluctuations in Emerging Economies,” American Economic Review, 98(3), 690-712.

Bal Gunduz, Yasemin, and Reginald Darius, 2021, “Exchange Rate Adjustment and Growth in IMF-Supported Programs”, IEO Background Paper No. BP/21-01/05 (Washington: International Monetary Fund).

Bankowsky, Krzysztof, and Marien Ferdinandusse, 2017, “Euro Area Fiscal Stance,” ECB Occasional Paper No. 182, European Central Bank.

Barro, Robert. J., 1979, “On the Determination of the Public Debt,” Journal of Political Economy, 87(5), 940-971.

Baum Anja, and others, 2017, "Can They Do It All? Fiscal Space in Low-Income Countries," IMF Working Papers No. 2017/110, (Washington: International Monetary Fund).

Bianchi, Javier, Pablo Ottonello, and Ignacio Presno, 2023, “Fiscal Stimulus under Sovereign Risk,” Journal of Political Economy, 131(9), 2328-2369.

Bohn, Henning, 1998, “The Behavior of U.S. Public Debt and Deficits,” Quarterly Journal of Economics, August, Vol. 113, No. 3, pp. 949–963.

Carnot, Nicolas, 2014, “Evaluating Fiscal Policy – A Rule of Thumb,” European Commission, Economic Papers No. 526.

Carriere-Swallow, Yan, Antonio David, and Daniel Leigh, 2021, "Macroeconomic Effects of Fiscal Consolidation in Emerging Economies: New Narrative Evidence from Latin America and the Caribbean," Journal of Money, Credit and Banking.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo, 2011, "When Is the Government Spending Multiplier Large?" Journal of Political Economy, 119(1), 78-121.

Cohen-Setton, Jeremie and Peter Montiel, 2024, "The IMF's Exceptional Access Policy: Program Seign and Outcomes in IMF-Supported Programs with Exceptional Access," IEO Background Paper No. BP/24-02/02 (Washington: International Monetary Fund).

————, 2025, "IMF Fiscal Advice to Advanced Economics," IEO Background Paper No. BP/25-01/01 (Washington: International Monetary Fund).

Cohen-Setton, Jeremie, Peter Montiel, and Edda Zoli, 2025, "IMF Tools and Analysis Underpinning Fiscal Policy Advice, 2008-2023," Chapter 1 in "IMF Advice on Fiscal Policy—Selected Issues," IEO Background Paper No. BP/25-01/04 (Washington: International Monetary Fund).

Cole, H. L. and T. J. Kehoe, 2000, "Self-Fulfilling Debt Crises," Review of Economic Studies, Oxford University Press, vol. 67(1), pages 91-116.

Copelovitch, Mark S., 2010, "Master or Servant? Common Agency and the Political Economy of IMF Lending." International Studies Quarterly, 54(1), 49–77.

Chwieroth, Jeffrey M., 2010, "Capital Ideas: The IMF and the Rise of Financial Liberalization." Princeton: Princeton University Press.

Clarida, Richard, Jordi Gali, and Mark Gertler, 1999, "The Science of Monetary Policy: A New Keynesian Perspective," Journal of Economic Literature, 37(4), 1661-1707.

Devries, Pete, Jaime Guajardo, Daniel Leigh, and Andrea Pescatori, 2011, "A New Action-Based Dataset of Fiscal Consolidation in OECD Countries," IMF Working Paper No. 11/128, (Washington: International Monetary Fund).

Dreher, Axel, Valentin Lang, B. Peter Rosendorff, and James R. Vreeland, 2022, "Bilateral or Multilateral? International Financial Flows and the Dirty-Work Hypothesis." The Journal of Politics, 84(4), 1932–1946.

Eaton, Jonathan, and Mark Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical Analysis." Rev. Econ. Studies 48 (2): 289–309.

Eggertsson, Gauti B, 2011, "What Fiscal Policy Is Effective at Zero Interest Rates?" NBER Macroeconomics Ann. 25:59–112.

Eichengreen, Barry, and Poonam Gupta, 2025, "IMF's Advice on Fiscal Rules and Institutions," Chapter 2 in "IMF Advice on Fiscal Policy—Selected Issues," IEO Background Paper No. BP/25-01/04 (Washington: International Monetary Fund).

European Central Bank, 2016, ECB Economic Bulletin, Issue 4.

European Commission, 2015, "Report on Public Finances in EMU 2015."

————, 2020, "European Semester Thematic Factsheet: Sustainability of Public Finances."

Eyraud, Luc, and others, 2018, "How to Calibrate Fiscal Rules: A Primer," IMF How-To Notes No. 2018/002, (Washington: International Monetary Fund).

Farhi, Emmanual, and Ivan Werning, 2016, "Fiscal Multipliers: Liquidity Traps and Currency Unions," In John B. Taylor and Harald Uhlig (eds), Handbook of Macroeconomics, 2417-2492, Volume 2, Elsevier.

Fournier, Jean-Marc, and Philipp Lieberknecht, 2020, "A Model-based Fiscal Taylor Rule and a Toolkit to Assess the Fiscal Space," IMF Working Paper No. 2020/033, (Washington: International Monetary Fund).

Gali, Jordi, 2008, "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework."

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy, 2019, "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," Econometrica, 87(4), 1307-1340.

Hansen, Anne Lundgaard, and Sophia Kazinnik, 2024, "Can ChatGPT Decipher Fedspeak?" SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.4399406.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, 2019, "Firm -Level Political Risk: Measurement and Effect," The Quarterly Journal of Economics, 134(4), 2135-2022.

Heinzel, Mirko, 2022, "International Bureaucrats and Organizational Performance: Country-Specific Knowledge and Sectoral Knowledge in World Bank Projects." International Studies Quarterly, 66(2).

Independent Evaluation Office of the International Monetary Fund (IEO), 2023, The IMF's Emergency Response to the COVID-19 Pandemic, (Washington: International Monetary Fund).

International Monetary Fund, 1995, "Guidelines for Fiscal Adjustment," Pamphlet Series No. 49 (Washington).

———, 2013, “The Interaction of Monetary and Macroprudential Policies,” IMF Policy Paper (Washington).

———, 2018, “Assessing Fiscal Space—An Update and Stocktaking,” April (Washington).

———, 2022, “Guidance Note for Surveillance Under Article IV Consultations,” IMF Policy Paper No. 2022/029 (Washington).

Jalles, Joao T., Youssouf Kiendrebeogo, Raphael Lam, and Roberto Piazza, 2024, “Revisiting the Countercyclicality of Fiscal Policy,” Empirical Economics 67, 877–914.

Jeanne, Olivier, and Anton Korinek, 2013, “Macroprudential Regulation versus Mopping Up After the Crash,” NBER Working Paper No. 18675.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, 2024, “ChatGPT and Corporate Policies,” National Bureau of Economic Research, Working Paper 32161.

Kalai, Adam Tauman, Ofir Nachum, Santoshi S. Vempala, and Edwin Zhang, 2025, “Why Language Models Hallucinate,” https://doi.org/10.48550/arXiv.2509.04664.

Kalyani, Aakash, and others, 2025, “The Diffusion of New Technologies,” The Quarterly Journal of Economics, 140(2), 1299-1365.

Kaminsky, Graciela L., Carmen M. Reinhart, and Carlos A. Végh, 2004, “When It Rains, It Pours: Procyclical Macropolicies and Capital Flows.” NBER Macroeconomics Ann. 19:11–53.

Kanda, Daniel S., 2011, “Modeling Optimal Fiscal Consolidation Paths in a Selection of European Countries,” IMF Working Paper No. 11/164 (Washington: International Monetary Fund).

Kilic Celik, Sinem, Ayhan M. Kose, Franziska Ohnsorge, and F. Ulrich Ruch, 2023, “Potential Growth: A Global Database,” Policy Research Working Paper 10354, (Washington: World Bank).

Korinek, Anton, 2023, "Generative AI for Economic Research: Use Cases and Implications for Economists," Journal of Economic Literature, 51(4), 1281-1317.

———, 2024, “LLMs Level Up—Better, Faster, Cheaper,” Update to Section 3 of "Generative AI for Economic Research: Use Cases and Implications for Economists," Journal of Economic Literature.

———, 2025, “AI Agents for Economic Research,” Update of "Generative AI for Economic Research: Use Cases and Implications for Economists," Journal of Economic Literature.

Korinek, Anton, and Alp Simsek, 2016, “Liquidity Trap and Excessive Leverage,” American Economic Review, 106(3), 699-738.

Kose, Ayhan M., Segio Kurlat, Franziska Ohnsorge, and Naotaka Sugawara, 2022, "A cross-country database of fiscal space," Journal of International Money and Finance, 128(November): 102682.

Krugman, Paul, 2009, "How Did Economists Get It So Wrong?" The New York Times. September 2.

Lang, Valentin, Lukas Wellner, and Alexandros Kentikelenis, 2024, "Biased bureaucrats and the policies of international organizations." American Journal of Political Science 1–19.

Saxegaard, Elif C. Arbatli, and others, 2012, "A toolkit for Assessing Fiscal Vulnerabilities and Risks in Advanced Economics," IMF Working Paper No. 2012/011 (Washington: International Monetary Fund).

Shapiro, Adam Hale, and Daniel J. Wilson, 2022, "Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis," The Review of Economics Studies, 89(5), 2768-2805.

Stone, Randall W., 2011, "Controlling Institutions: International Organizations and the Global Economy." Cambridge: Cambridge University Press.

Ugba, Ejike R., and Jan Gertheiss, 2023, "A modification of McFadden's R2 for binary and ordinal response models," Communications for Statistical Applications and Methods, 30(1), 49–63.

Weaver, Catherine, Mirko Heinzel, Samantha Jorgensen, and Joseph Flores, 2022, "Bureaucratic Representation in the IMF and the World Bank," Global Perspectives, 3(1), 39684.

Woodford, Michael, 2003, "Interest and Prices: Foundations of a Theory of Monetary Policy," Princeton, NJ: Princeton University Press.

————, 2011, "Simple Analytics of the Government Expenditure Multiplier," American Economics Journal: Macroeconomics, 3(1), 1-35.