IEO

Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/25-01/05

Reading the Fund: A Systematic Analysis of IMF Fiscal Advice Using Large Language Models

Jantsankhorloo Amgalan, Jérémie Cohen-Setton, and Anton Korinek

 BP/25-01/05

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Reading the Fund: A Systematic Analysis of IMF Fiscal Advice Using Large Language Models

Prepared by Jantsankhorloo Amgalan,* Jérémie Cohen-Setton,† and Anton Korinek‡

November 3, 2025

* Research Officer, Independent Evaluation Office of the IMF.

† Senior Economist, Independent Evaluation Office of the IMF.

‡ Consultant, Independent Evaluation Office of the IMF.

Contents Page

Table

Appendixes

ABBREVIATIONS

AE	Advanced Economy
CDS	Credit Default Swap
DSA	Debt Sustainability Analysis
DSF	Debt Sustainability Framework
EMMIE	Emerging Market and Middle-Income Economy
FB	Fiscal Balance
GDP	Gross Domestic Product
GFC	Global Financial Crisis
GFN	Gross Financing Need
HIPC	Heavily Indebted Poor Country
LIC	Low-Income Country
LLM	Large Language Model
WEO	World Economic Outlook
ZLB	Zero Lower Bound

EXECUTIVE SUMMARY

This paper introduces a novel methodology to systematically evaluate the IMF's fiscal policy advice across its entire membership. Using large language models (LLMs), we classify over 3,000 Article IV reports from 1998 to 2023 according to their recommended near-term fiscal stance, medium-term fiscal stance, monetary policy stance, and macroprudential policy stance. We then link near-term fiscal stance recommendations to contemporaneous macroeconomic conditions through econometric analysis. The results show that IMF advice broadly aligns with the dual objectives of fiscal policy emphasized in the literature, namely, output stabilization and fiscal sustainability. Near-term fiscal recommendations tend to be more expansionary in periods of economic slack and more contractionary when fiscal vulnerabilities are elevated.

Our analysis highlights the importance of the macroeconomic policy mix and the broader policy environment in shaping IMF fiscal advice. Over the evaluation period, fiscal and monetary policies were generally treated as complements in IMF surveillance, with fiscal loosening more likely when monetary policy advice was also expansionary. Fiscal recommendations were particularly more accommodative when monetary policy was constrained by the zero lower bound (ZLB), consistent with the view that fiscal policy should play a greater stabilization role when monetary space is limited. In contrast, we find no systematic relationship between fiscal and macroprudential advice, suggesting limited coordination between these two instruments in practice.

IMF fiscal advice has evolved to place greater emphasis on output stabilization since the Global Financial Crisis (GFC), though with variation over time. Fiscal recommendations were clearly countercyclical during major global crises such as the 2008–09 GFC and the 2020–21 pandemic. However, advice became noticeably less countercyclical in Article IV reports published during 2010–12, before turning more countercyclical again in the later post-GFC period. The shift toward greater countercyclicality is particularly evident in advanced economies, where larger fiscal space and greater monetary autonomy have allowed more active use of discretionary fiscal policy.

Despite the increased emphasis on output stabilization, fiscal sustainability concerns remain central to IMF fiscal advice, as reflected in the continued predominance of consolidation recommendations. Consolidation continues to dominate medium-term recommendations, especially for emerging markets and low-income countries. Within fiscal sustainability indicators, the deviation of the fiscal balance from its debt-stabilizing level exerts a stronger influence on IMF advice for advanced economies than the absolute level of public debt. Our analysis shows that, beyond traditional solvency metrics, IMF advice also responds to a broader set of fiscal vulnerability indicators—including liquidity pressures, market access conditions, and structural features of the debt portfolio that may amplify risks during episodes of market stress or external financing shocks. Fiscal institutions also play a crucial role in shaping IMF advice. Stronger fiscal frameworks—as proxied by the presence of fiscal rules and fiscal councils—are associated with more countercyclical recommendations.

In addition to traditional considerations of output stabilization and fiscal sustainability, our analysis finds that external imbalances also influence the fiscal stance recommended by the IMF. Countries with current account deficits are more likely to receive advice to tighten, while those with surpluses are somewhat more likely to be advised to ease. For advanced economies, we find evidence of an asymmetric relationship: external deficits are consistently associated with a higher likelihood of receiving consolidation advice, whereas the effect of external surpluses on the probability of receiving loosening advice is weaker and less systematic.

Finally, our analysis shows that fiscal advice is shaped not only by country-specific conditions but also by the individuals delivering it. In particular, we find that the educational background of IMF mission chiefs matters. Those trained at "saltwater" institutions, typically associated with Keynesian traditions, are more likely to recommend countercyclical fiscal policies compared with their "freshwater" counterparts whose training emphasizes monetarist or neoclassical approaches. This effect remains significant even after controlling for macroeconomic conditions, suggesting that individual-level characteristics can influence the Fund's advice alongside country-specific economic fundamentals.

To facilitate replication and broader application, the paper is accompanied by a practical guide that outlines how to use large language models (LLMs) to analyze textual economic data—from data preparation and prompt design to API-based implementation—enabling other researchers to adapt the methodology to different areas of IMF surveillance and economic policy analysis.

I. INTRODUCTION[1]

1. **Fiscal policy is one of the core pillars of the IMF's Article IV surveillance.** Through its Article IV consultations, the Fund provides regular advice on how member countries should calibrate their fiscal stance to balance fiscal sustainability with macroeconomic stabilization and the pursuit of longer-term policy priorities. These recommendations influence national policy debates, anchor Fund-supported programs, and shape global perceptions of fiscal prudence and coordination. Understanding the nature and evolution of this advice is therefore essential to assessing the IMF's effectiveness in fulfilling its surveillance mandate.

2. **Yet, no comprehensive dataset exists to document this advice.** The Fund provides regular advice on a wide range of issues—including the stance of macroeconomic policies, the composition and timing of fiscal measures, and structural reforms—but no comprehensive dataset currently documents this advice. As a result, empirical studies of Fund's fiscal advice have relied on limited case studies or small hand-coded samples, leaving open fundamental questions about how fiscal recommendations vary with macroeconomic conditions, differ across country groups, or evolve over time. This paper addresses that gap by constructing the first large-scale, text-based dataset of IMF fiscal policy advice, enabling systematic evaluation of its patterns, determinants, and consistency.[2]

3. **Recent advances in large language models (LLMs) make it possible to study IMF advice at an unprecedented scale.** This background paper, prepared as part of the Independent Evaluation Office's evaluation of the Fund's fiscal policy advice, 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.

4. **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

[1] The authors 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 internal seminars at the IEO and the Peterson Institute for International Economics for their comments and suggestions.

[2] A previous attempt to systematically collect information on IMF bilateral policy advice was made through the Consistent Policy Assessment (CPA) database. The CPA compiles information on the recommended stance of fiscal, monetary, macro-financial, and structural policies through biannual staff surveys. It also collects data on staff assessments of internal and external positions, public and private balance sheets, and policy frameworks. However, the database is not publicly available, covers only the period since 2018, and is only collected on an *ad hoc* basis since 2025 as part of an effort to streamline outputs and business processes.

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 staff 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.

5. **IMF fiscal advice 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)—reflecting 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.

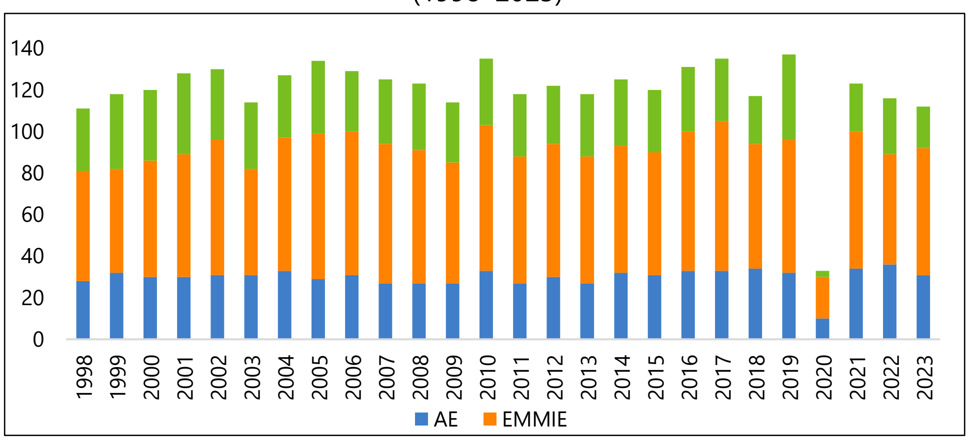
6. **Yet, it would be misleading to portray Fund advice as uniformly contractionary.** For the near term, recommendations are generally countercyclical, more expansionary in periods of economic slack and more contractionary when activity is strong. This countercyclicality has strengthened over time, 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.

7. **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. A methodological appendix accompanies the paper, offering a practical guide for applying LLMs to the analysis of textual economic data—from data preparation and prompt design to API-based implementation—so that the approach can be adapted to other areas of IMF surveillance and economic policy analysis. An Online Appendix reports all underlying regression tables and discusses robustness checks.

II. Methodology and Data Construction

8. **This background paper constructs a novel, text-based dataset of IMF fiscal policy recommendations using LLMs applied to all published Article IV consultation reports from 1998 to 2023.**[3] The corpus 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. 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.

9. **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.**[4] 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

---

[3] 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.

[4] 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.

medium-term fiscal stance, the monetary stance, and the macroprudential stance.[5] Each recommendation is categorized as Tighten, Neutral, Loosen, or Unclear (Box 1). The prompt instructs the models to reason explicitly before assigning a classification and to produce a certainty score from 0 (low confidence) to 100 (high confidence).[6] 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.[7]

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.

[5] 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.

[6] Section II of the Online Appendix analyzes 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. A robustness exercise applies alternative confidence thresholds, reclassifying low-certainty observations as "Unclear." As the threshold rises, the share of reclassified country-year observations increases, with the largest adjustments observed among LICs.

[7] 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.

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

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.

10. **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.

11. **To validate the accuracy of these machine-generated classifications, we conducted a manual review of 105 randomly selected country-year pairs.**[8] 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.[9] Most discrepancies did not reflect clear errors

[8] 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.

[9] 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.

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

III. DESCRIPTIVE EVIDENCE FROM THE CLASSIFICATION

12. **This section presents descriptive patterns from the LLM-based classification of IMF policy advice.** It summarizes how recommendations on fiscal, monetary, and macroprudential policies vary across income groups, over time, and across policy horizons, 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

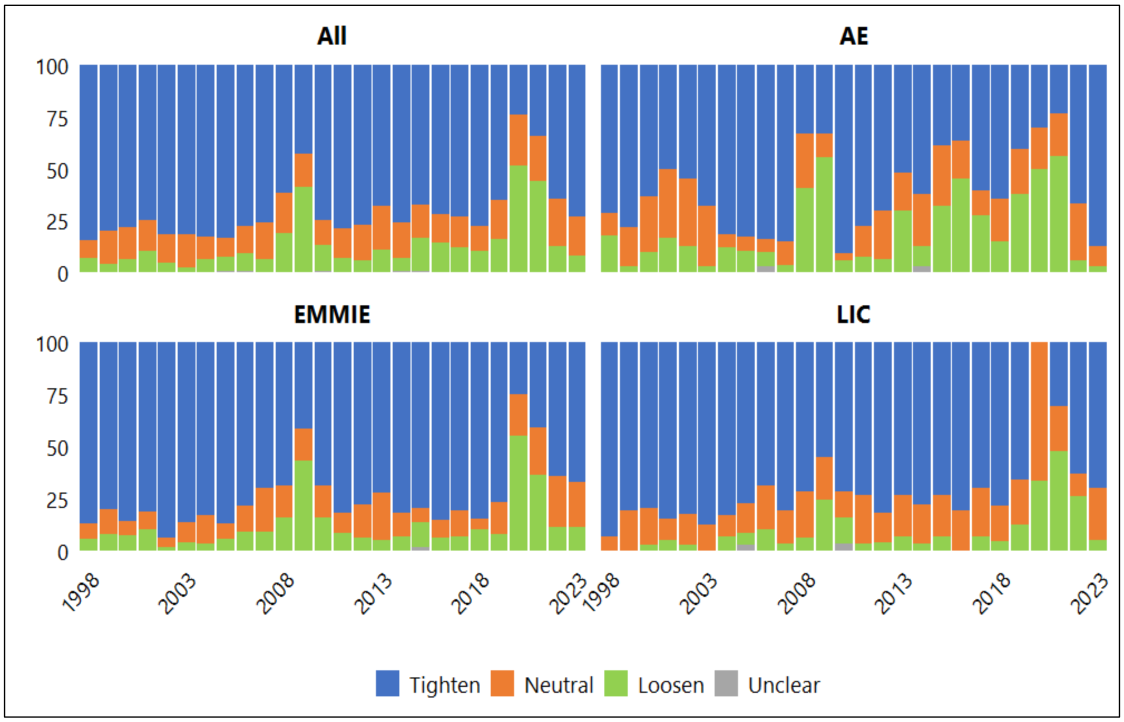
Sources: IMF Article IV Reports; IEO staff calculations.
Note: A substantial share of monetary policy advice is classified as "Unclear," primarily because many countries either participate in a monetary union or operate under fixed exchange rate regimes that constrain independent monetary policymaking. 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

13. **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)



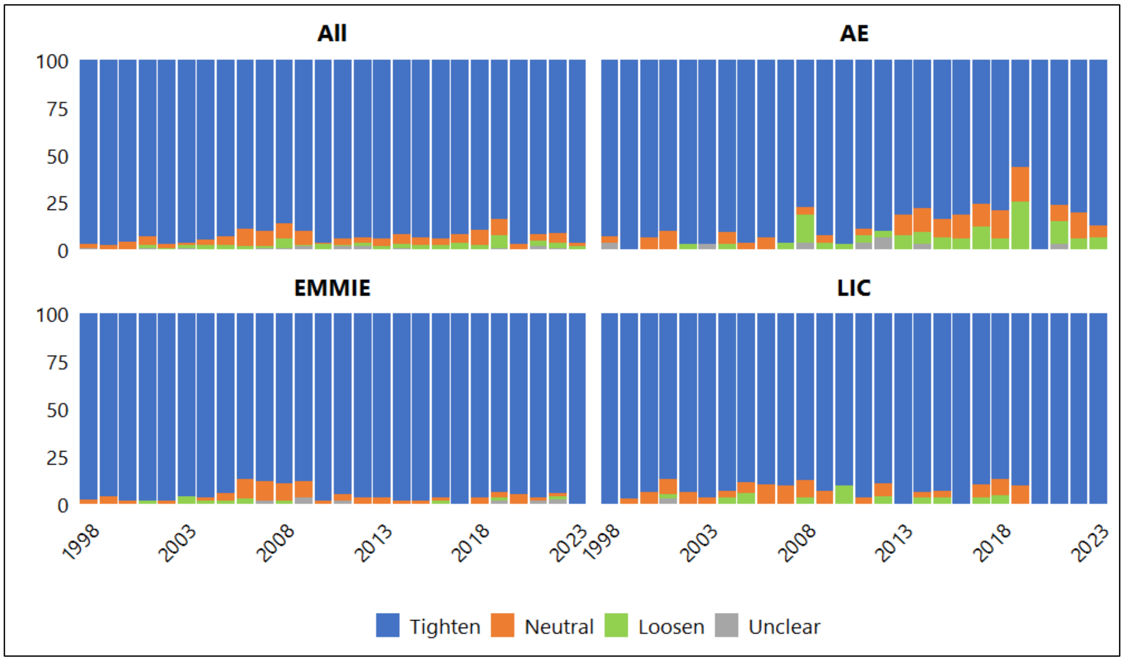
Sources: IMF Article IV Reports; IEO staff calculations.

14. **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.[10] 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

[10] 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).

low growth and interest rates, and its calls for higher public investment in countries with ample fiscal space (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

15. **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.

16. **IMF advice is most persistent for medium-term fiscal policy, reflecting its focus on long-term fiscal sustainability.** 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).[11] Medium-term fiscal advice is far more stable,

[11] The numbers are taken from Table AII.2. of the Online Appendix.

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.

IV. DETERMINANTS OF NEAR-TERM FISCAL POLICY ADVICE

17. **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

18. **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.

19. **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).[12] In contrast, models of sovereign default (Eaton and Gersovitz, 1981; Aguiar and Gopinath, 2006; Arellano, 2008) underscored the costs and risks of additional borrowing,

[12] 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.

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.

20. **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.[13] 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.

21. **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 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.[14]

B. Empirical Specification and Data

22. **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

---

[13] 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.

[14] 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 pattern is strongest among AEs, less systematic for LICs, and broadly consistent with the FRF framework. See Table AIII.1. in the Online Appendix for details.

classification of near-term fiscal advice (Loosen, Neutral, or Tighten), which is ordinal by construction. This approach provides a transparent and 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.[15] 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.[16] 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.

23. **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 the influence of outliers. In specifications with country fixed effects, countries with invariant fiscal stance classifications are dropped to avoid collinearity, though they are

[15] Section IV of the Online Appendix provides further details on ordered logit models.

[16] 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.

retained in pooled regressions without fixed effects.[17] 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.[18]

V. Main Results

A. Baseline Results

24. **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.[19] 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 2 reports results from four specifications—without fixed effects, with country effects, with year effects, and with both. Across all models, IMF staff recommendations exhibit clear and statistically significant responsiveness to macro-fiscal conditions.[20]

---

[17] 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.

[18] 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 presented in Section III of the Online Appendix. When no contemporaneous WEO vintage is available, we used the 2024 October WEO to maximize coverage, applying its values retrospectively.

[19] 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.

[20] Section III of the Online 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.

25. **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 2). 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.

Table 2. 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	

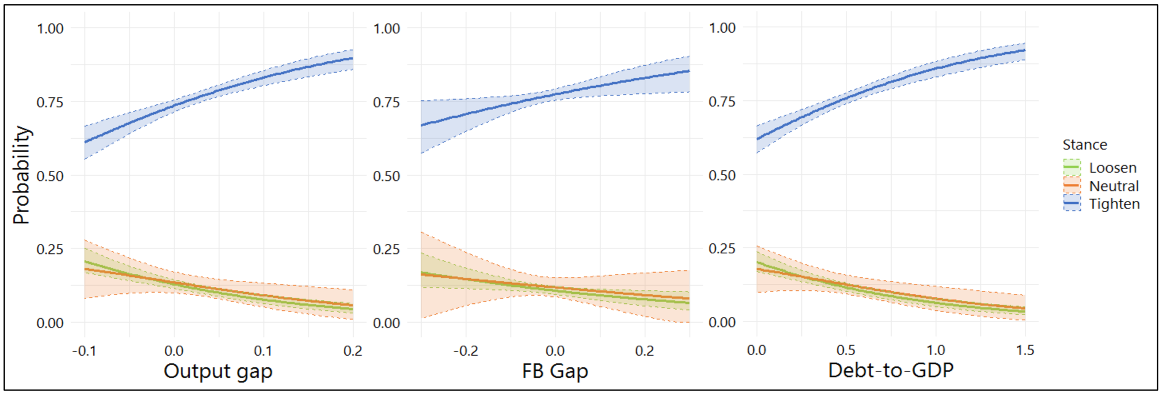
Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. All regressions use contemporaneous data. The average marginal effect (AME) columns report the average expected change in the predicted probability of tightening advice given 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.

26. **The ordered logit framework provides additional insight into how fiscal stance recommendations are structured across outcome categories.** 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.[21]

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 2 column (1). Covariates are presented in decimal.

27. **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 2; 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 (Online Appendix Table AI.1; Figure AI.1). Taken together, these results confirm that IMF staff systematically lean against the cycle when formulating near-term fiscal advice.

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

28. **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 Online Appendix.[22]

B. Heterogeneity by Income Group

29. **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—output gap, fiscal FB gap, and debt-to-GDP ratio—in the ordered logit model.

30. **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

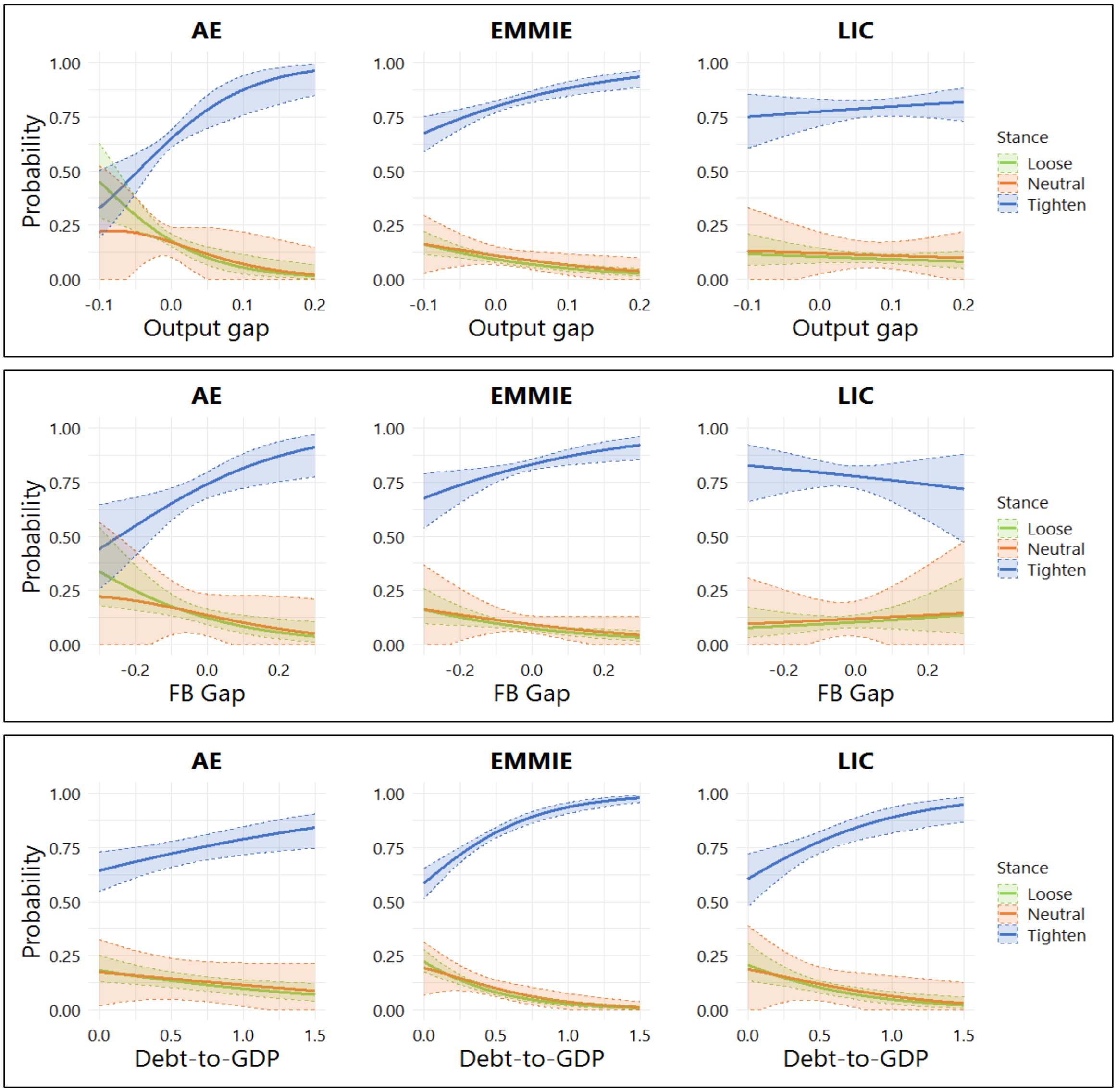
---

[22] Section I.B. of the Online 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.

countercyclical policies.[23] 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)



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 AV.1. column (1) of the Online Appendix. Covariates are presented in decimal.

[23] 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.

31. **The influence of fiscal vulnerability indicators varies across income groups.** In AEs, both indicators matter, but the FB gap plays the more prominent role in shaping sustainability-related recommendations. The positive and significant FB gap coefficients across most specifications indicate that IMF staff place greater weight on the recommended fiscal stance than the debt-to-GDP ratio itself, as shown by the flatter slope. For EMMIEs, both the debt ratio and the FB gap exert strong and significant effects, reflecting greater sensitivity to debt-market conditions and risks to investor confidence. In LICs, by contrast, only the debt ratio consistently predicts tighter advice, while the FB gap has limited explanatory power—likely because these countries face tighter financing constraints and smaller observed variation in fiscal adjustment needs.

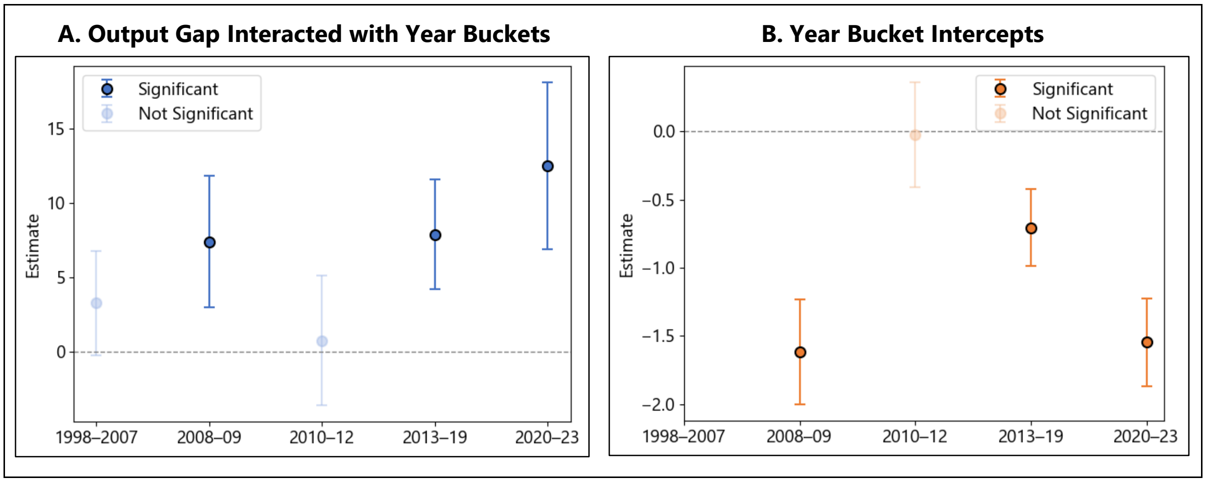
C. Heterogeneity Over Time

32. **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.

33. **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.[24]

Figure 6. Evolution of the Countercyclicality of IMF Fiscal Advice



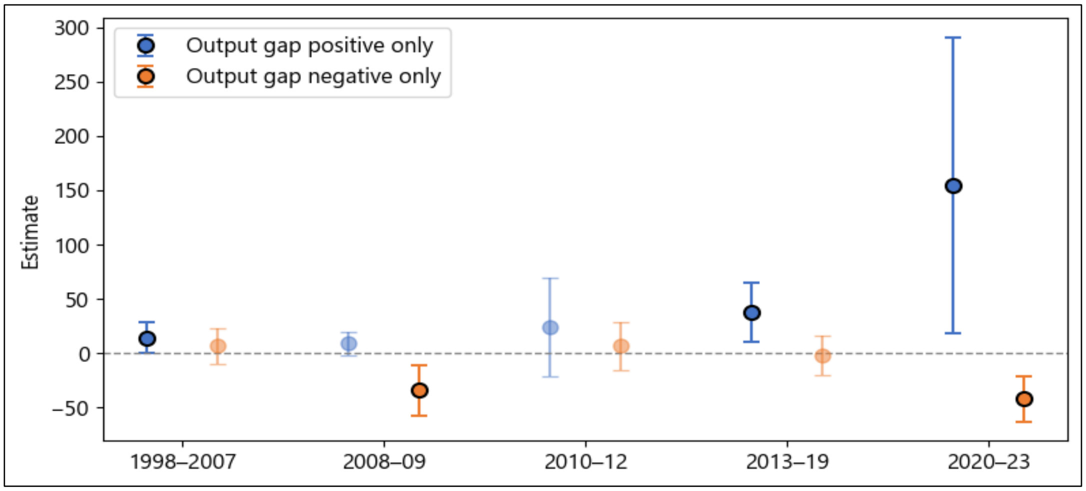
Sources: WEO; Kose and others (2022); IEO staff calculations.
Note: The figure is based on the estimates in Table AV.2, column (3) of the Online Appendix. It shows coefficient estimates from an ordered logit model where the output gap is interacted with period dummies. Each point represents the estimated 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.

34. **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.[25] 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.

[24] Detailed estimates corresponding to these results are reported in Table AV.2. of the Online Appendix.

[25] 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.

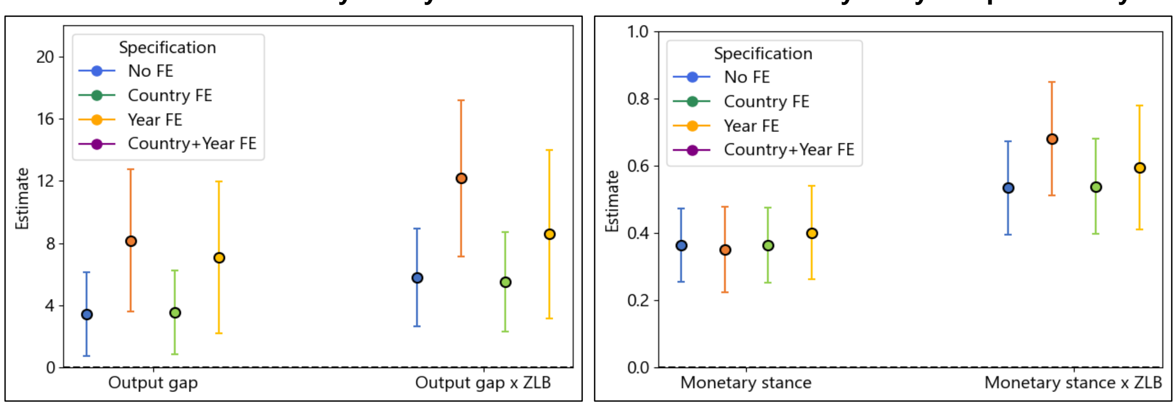
D. Macroeconomic Policy Mix

35. **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).

36. **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.[26] 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 aproach 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.[27]

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 AV.3. of the Online Appendix.

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

[27] 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 AV.3. of the Online Appendix.

37. **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.[28]

E. External Imbalances

38. **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.

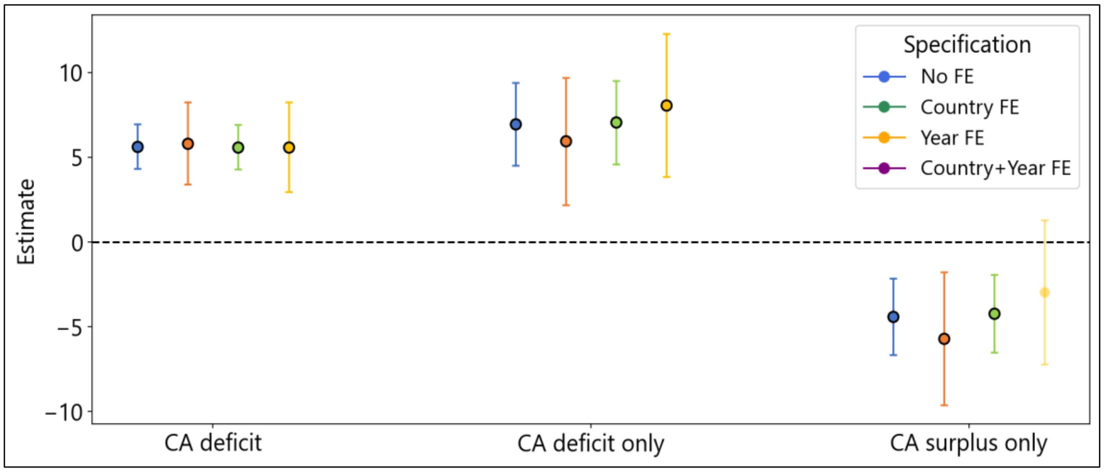
39. **Fiscal advice displays a modest asymmetry in its response to external imbalances.** When distinguishing between current account deficit and surplus cases,[29] the results show that fiscal tightening recommendations are positively associated with current account deficits across all specifications (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.

---

[28] Detailed estimates corresponding to these results are reported in Table AV.4. of the Online Appendix.

[29] 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 AV.5. of the Online Appendix. Brighter colors denote statistically significant estimates; lighter shades indicate insignificant results.

40. **The asymmetry becomes more pronounced when the analysis is disaggregated by income group.**[30] 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

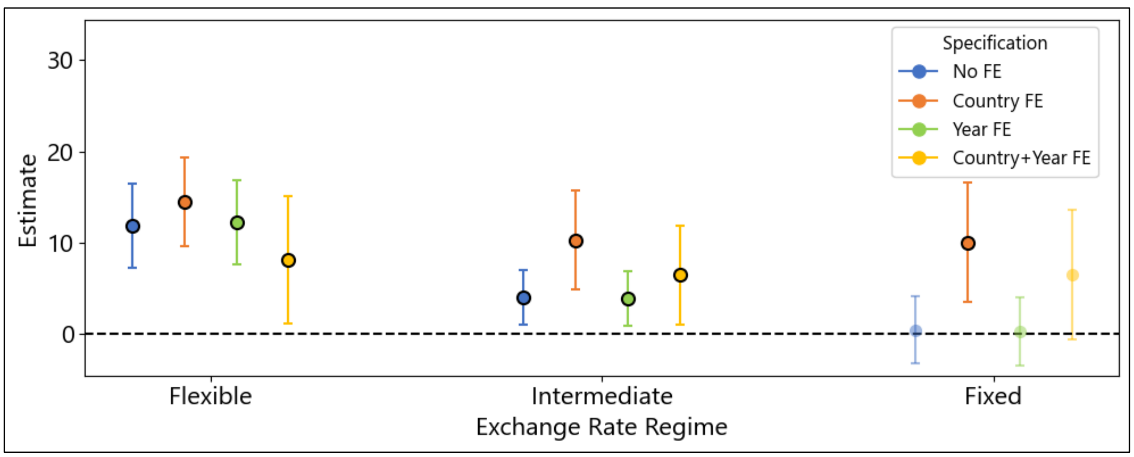
41. **The relationship between exchange rate regimes and the countercyclicality of IMF fiscal advice can be assessed through the lens of the Mundell–Fleming model.** In the textbook framework, fiscal policy plays a stronger stabilization role under fixed exchange rates, where monetary policy is constrained 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

[30] Detailed estimates corresponding to these results are reported in Tables AV.6. and AV.7. of the Online Appendix.

test this prediction, we interact the output gap with indicators for flexible, intermediate, and fixed 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).[31]

42. **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.[32] 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 AV.8. of the Online Appendix. Brighter colors indicate statistically significant estimates, while lighter shades represent insignificant results.

[31] 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.

[32] Detailed estimates corresponding to these results are reported in Table AV.8. and AV.9. of the Online Appendix.

G. Program vs. Surveillance

43. **IMF fiscal advice is markedly less responsive to cyclical conditions under IMF-supported programs than during regular surveillance.**[33] The output gap displays a strong and highly significant positive association with fiscal tightening recommendations in surveillance cases, consistent with countercyclical advice—staff tend to recommend tightening when output is above potential and easing when there is slack. Under program arrangements, however, the interaction term between the output gap and program participation is small and statistically insignificant, indicating that cyclical conditions play a much smaller role in shaping fiscal recommendations. This reduced sensitivity likely reflects the context in which programs are adopted: countries seeking IMF support typically face acute macroeconomic stress and financing pressures, leaving limited scope for discretionary countercyclical policy. Program-related fiscal targets are often guided more by financing envelopes and the need to restore market confidence than by cyclical stabilization objectives.

44. **Differences in the responsiveness of fiscal advice to sustainability indicators tell a similar story.** During surveillance, both the FB gap and the debt ratio are strong and statistically significant predictors of tightening advice. Once a program is in place, however, the influence of the FB variable weakens noticeably. This attenuation likely reflects that many program countries have already initiated fiscal adjustment before entering a program—often under market or creditor pressure—leaving less variation in fiscal stance to explain. It may also stem from the fact that, as noted by Cohen-Setton and Montiel (2024), fiscal consolidation is a standard element of IMF-supported programs even when fiscal imbalances were not the primary source of the crisis.

H. Fiscal Institutions

45. **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.

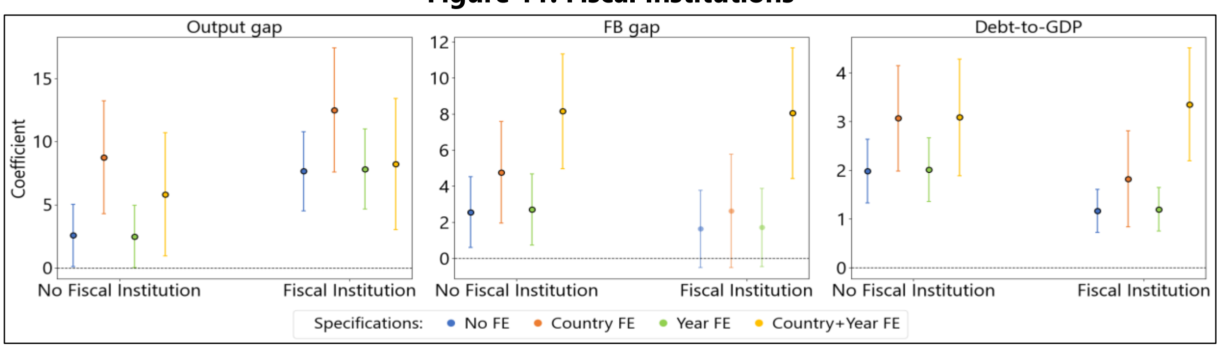
46. **To explore this question, we extend the baseline framework to incorporate the role of fiscal institutions.** Specifically, we introduce a dummy variable that equals one when a fiscal rule or an independent fiscal council is in place, using data from the IMF Fiscal Rules and Fiscal

[33] Program participation is identified through a country-year dummy constructed from the IMF's MONA database. In our sample, program years account for about 21 percent of total observations. Detailed estimates are reported in Tables AV.10. and AV.11. of the Online Appendix.

Councils Datasets (Alonso and others, 2025a; 2025b).[34] 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.

47. **The results show that countries with fiscal institutions receive more countercyclical and less consolidation-oriented IMF advice.** As illustrated in Figure 11,[35] 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 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 AV.15. of the Online Appendix. Brighter colors indicate statistically significant estimates, while lighter shades represent insignificant results.

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

[35] Detailed estimates corresponding to these results are reported in Tables AV.12. and AV.13. of the Online Appendix.

I. Non-Economic Factors

48. **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).

49. **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.

50. **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.

51. **The results indicate that staff educational background meaningfully influences the formulation of fiscal advice.**[36] 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

52. **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 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.

53. **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. 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.

---

[36] Detailed estimates corresponding to these results are reported in Table AV.14. of the Online Appendix.

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.

54. **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 CLASSIFICATION

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.

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

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. 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.

Box AI.2. Ireland 2019 (concluded)

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.

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

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

Table AIII.1. Data Description (concluded)

Indicator	Description	N	Unit	Source	Vintage	Transformation	Time index
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

ONLINE APPENDIX

An Online Appendix accompanies the paper. It presents robustness analyses using alternative indicators of macroeconomic stabilization needs and fiscal sustainability. It also provides full documentation of the LLM-based classification, data construction, and baseline results obtained when using the classification from a single LLM model (rather than the one obtained from majority-voting) and from alternative WEO vintages. The Appendix also details the ordered logit estimation framework, including diagnostic tests, and reports the complete set of regression tables referenced in the main text. The Online Appendix is available upon request from the corresponding author at jamgalan@imf.org.

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).

————, 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).

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.

Liu, Dungang, and Heping Zhang, 2017, "Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach," Journal of the American Statistical Association, 113(522), 845-854.

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.