



Independent Evaluation Office  
of the International Monetary Fund

# Reading the Fund: Analyzing IMF Fiscal Advice Using LLMs

Jantsankhorloo Amgalan, Jérémie Cohen-Setton, and Anton Korinek

Preliminary: Do not cite

ASSA | January 3, 2026

## This Paper

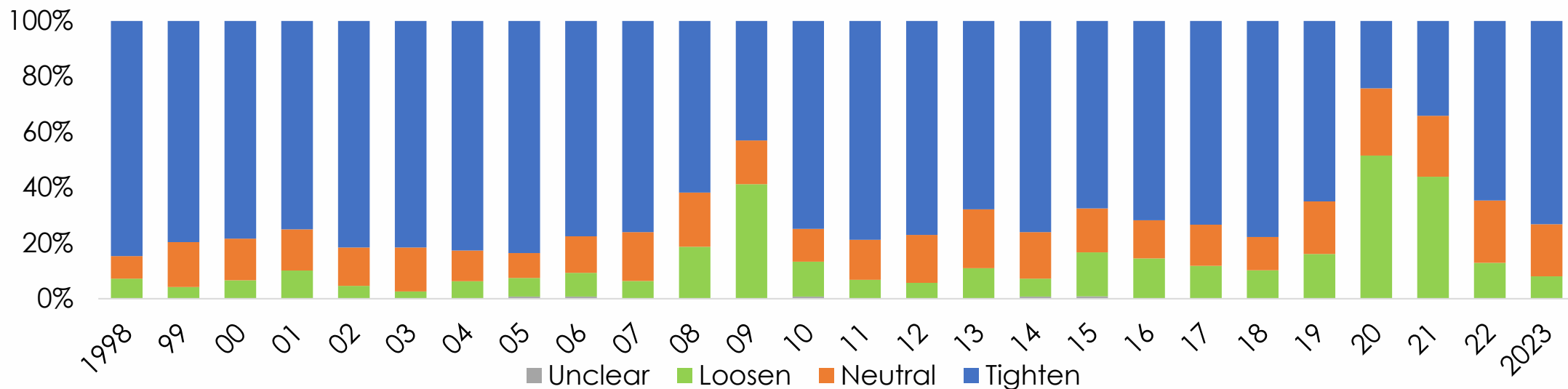
- ▶ **Extract and Classify IMF Staff Advice with LLMs:**
  - More than 3,000 AIV reports from 1998 to 2023.
  - Article IV report is a core document for bilateral surveillance.
- ▶ **Build Macro Dataset:** Compile contemporaneous macroeconomic data.
- ▶ **Identify and Quantify Drivers:** Estimate the influence of key factors on the near-term fiscal stance recommended by staff.

## Main Findings

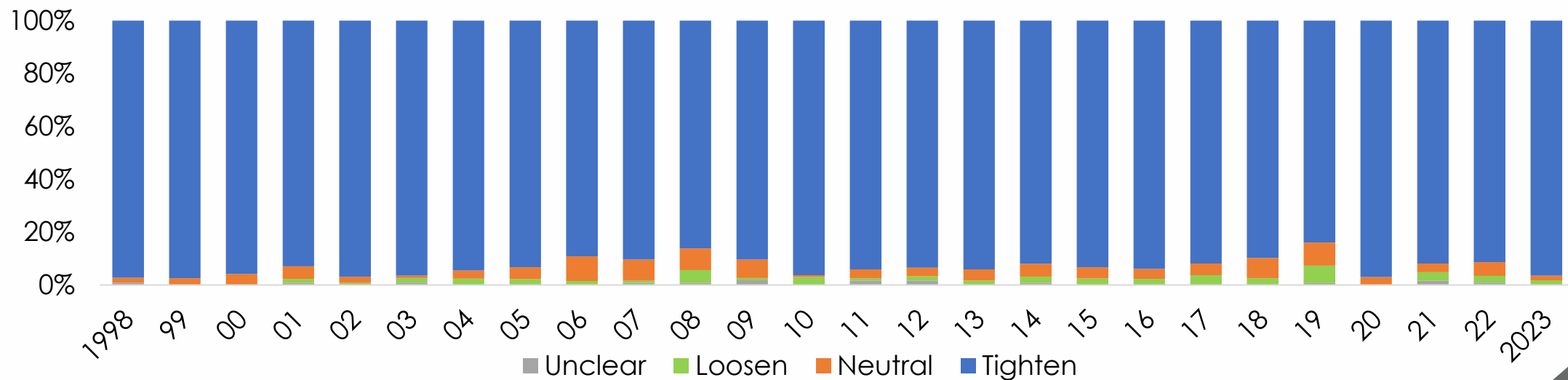
- ▶ **Advice generally framed around three main objectives:**
  - Output Stabilization
  - Fiscal Sustainability
  - Potential Growth & Priority Spending.
- ▶ **Additional factors:**
  - Other macro policies
  - External imbalances
  - Staff background

# Preview of Results

## Near-Term Advice on Fiscal Stance, 1998–2023

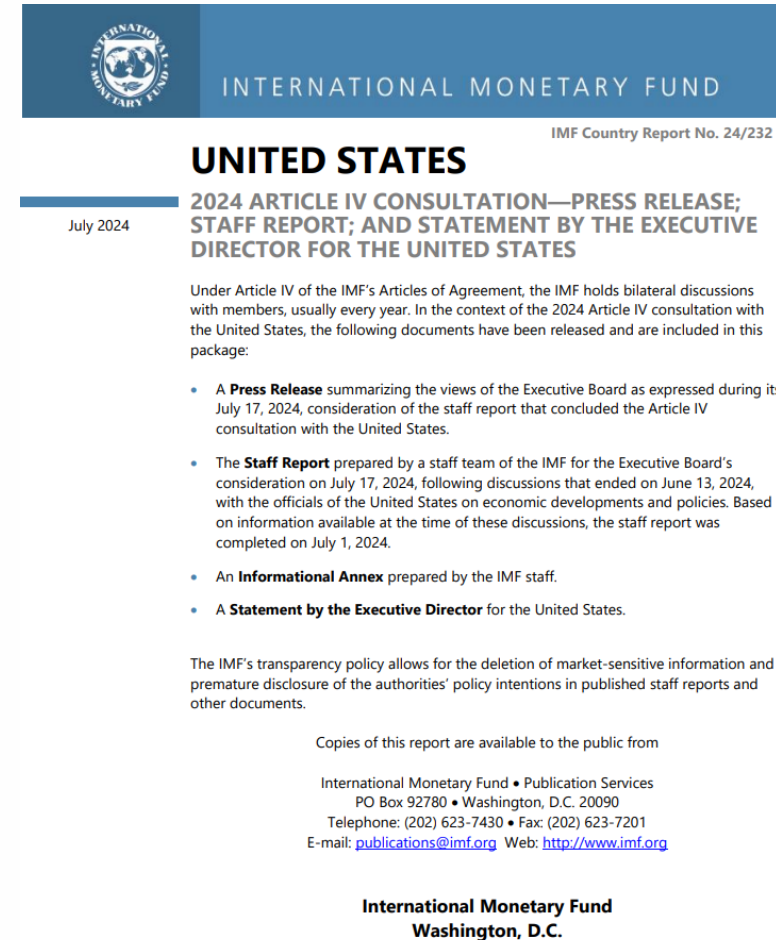


## Medium-Term Advice on Fiscal Stance, 1998–2023



# Fiscal Stance: Concept and Challenges

- **Objective:** Extract the directional change in discretionary fiscal policy.
- **Interpretation:** Terms like *loosening / stimulus* or *tightening / consolidation* are taken as signals of changes in the fiscal stance.
- **Benchmark:** Ideally corresponds to changes in the cyclically adjusted (structural) primary balance.
- **Challenge:** Article IV Staff Reports lack a consistent definition of “fiscal stance,” so usage can be ambiguous.



# Extracting and Classifying Fund Advice from Article IVs

## ► Classification Strategy

- **Task:** Classify IMF advice on policy stance as *loosen*, *neutral*, *tighten*, or *unclear*.
- **Method:**
  - Select **Staff Appraisal** section.
  - Models reason step-by-step: → justification → certainty score → classification.
  - Use multiple LLMs: OpenAI's o1, Claude 3.7 Sonnet, DeepSeek's R1.
  - Majority voting across models for the final classification.
- **Manual validation** of 105 country-year pairs:
  - 86% match rate for near-term and medium-term fiscal stance classifications.

## ► Prompt

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.

...

# Motivating the Empirical Specification

## ► Theoretical Motivation:

- Core trade-off between fiscal sustainability and output stabilization.

(Bianchi, Ottonello, & Presno 2023 and Fournier & Lieberknecht 2023)

## ► Implementation:

- No single variable perfectly captures either dimension.
- Start from a parsimonious specification that proxies both forces.
- Extensions to explore additional drivers and assess robustness.

## Empirical Model: Near-Term Advice

- ▶ We estimate the following near-term fiscal advice reaction function:

$$Y_{i,t}^* = \beta OutputGap_{i,t} + \gamma FBgap_{i,t-1} + \phi Debt_{i,t-1} + \theta' X_{i,t-1} + \varepsilon_{i,t}$$

- ▶ **Outcome variable:**  $Y_{i,t}^*$  is a latent variable capturing the recommended fiscal stance for the near-term, which is mapped onto the observed advice categories: Loosen, Neutral, Tighten.
- ▶ **Fixed effects:**  $\varepsilon$ , the error term, can be decomposed to include country and year fixed effects ( $\varepsilon_{i,t} = \alpha_i + \lambda_t + u_{i,t}$ ).
- ▶ **Timing:** All macro variables are taken from data vintages available at the time of AIV.
- ▶ **FBgap** measures the adjustment in the fiscal balance required to stabilize the debt-to-GDP ratio at its current level.

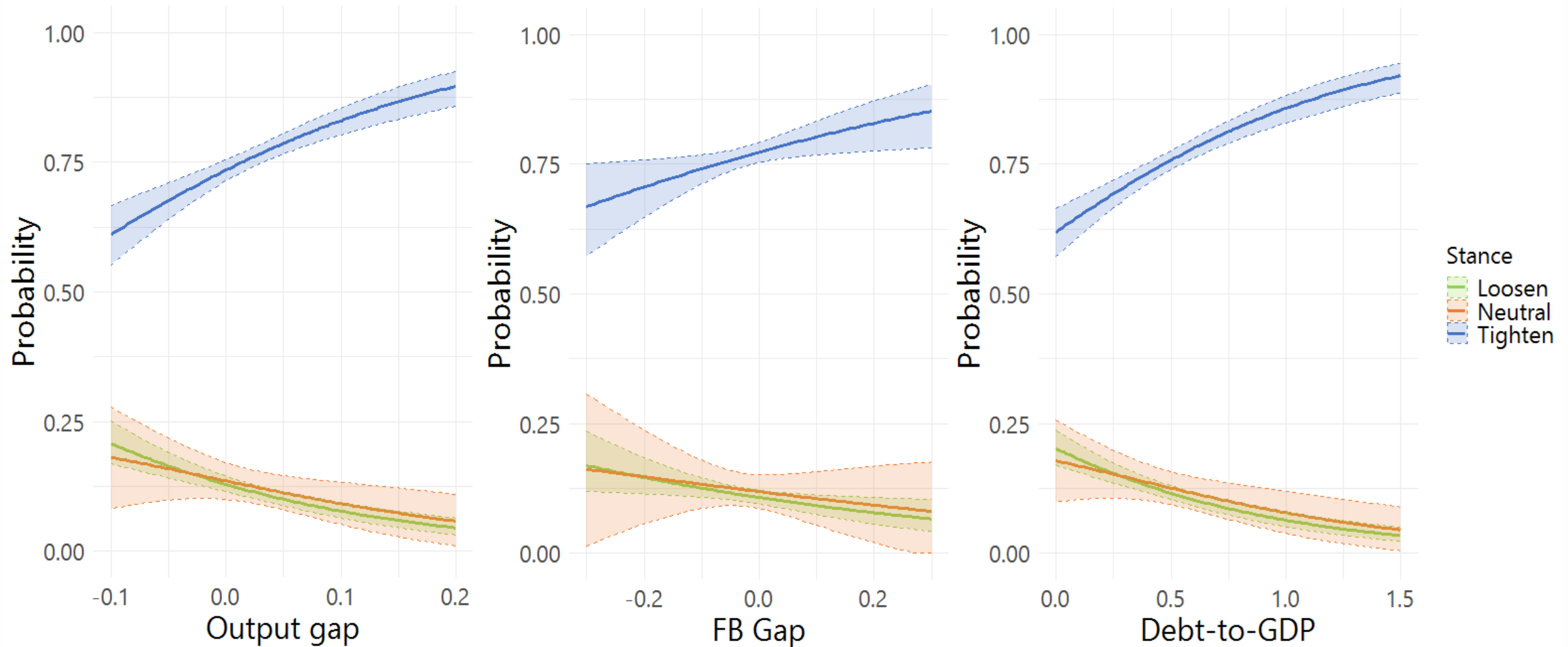


# Baseline Results

Variables	(1)		(2)		(3)		(4)	
	AME	$\beta(se_\beta)$	AME	$\beta(se_\beta)$	AME	$\beta(se_\beta)$	AME	$\beta(se_\beta)$
<b>Output gap</b>	1.022	5.671*** (0.951)	1.610	10.448*** (1.797)	1.015	5.689*** (0.956)	0.941	7.001*** (1.996)
<b>FB gap</b>	0.315	1.752** (0.726)	0.536	3.482*** (1.066)	0.334	1.873** (0.730)	1.096	8.149*** (1.316)
<b>Debt-to-GDP</b>	0.235	1.305*** (0.182)	0.351	2.279*** (0.377)	0.237	1.328*** (0.183)	0.433	3.220*** (0.459)
$\tau_1$		0.362*** (0.104)		2.399 (1.400)		0.626** (0.287)		2.548 (1.451)
$\tau_2$		1.251*** (0.110)		3.494** (1.401)		1.521*** (0.289)		3.828*** (1.453)
<b>N</b>	2720		2556		2720		2556	
<b>Country FE</b>	No		Yes		No		Yes	
<b>Year FE</b>	No		No		Yes		Yes	

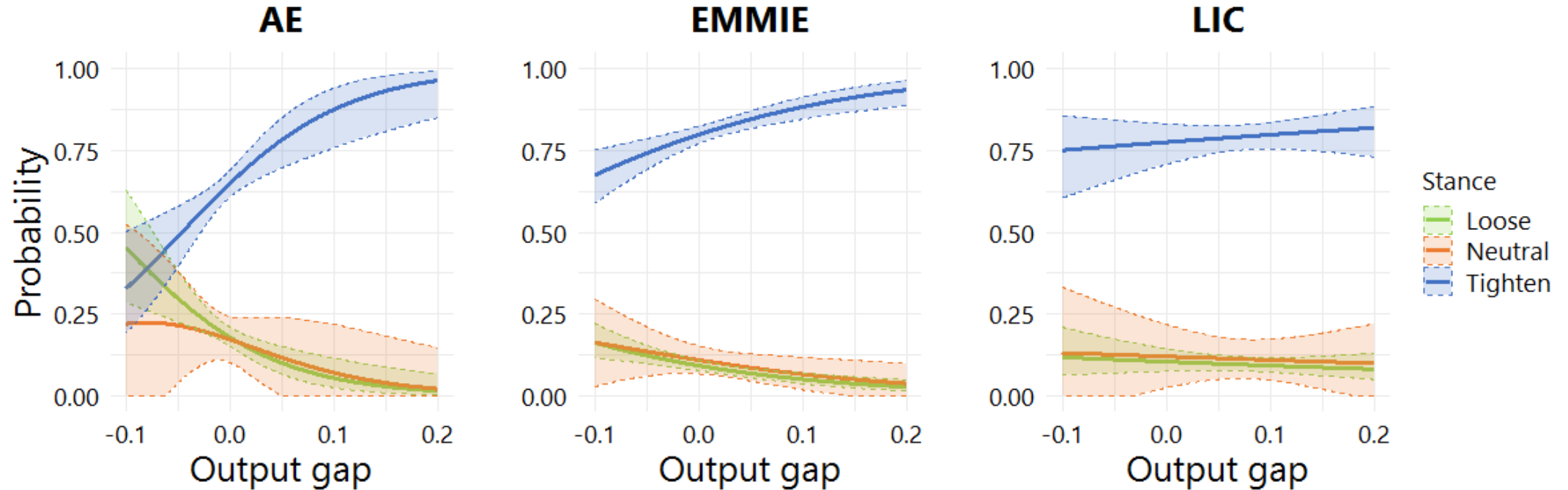
- The three variables are statistically significant.
- We estimate the average marginal effect (AME) toward tightening advice.

## Baseline Results: Probability Plots



- ▶ **AME** (ppts): Output gap (1.02), FB gap (0.31), Debt-to-GDP (0.23).

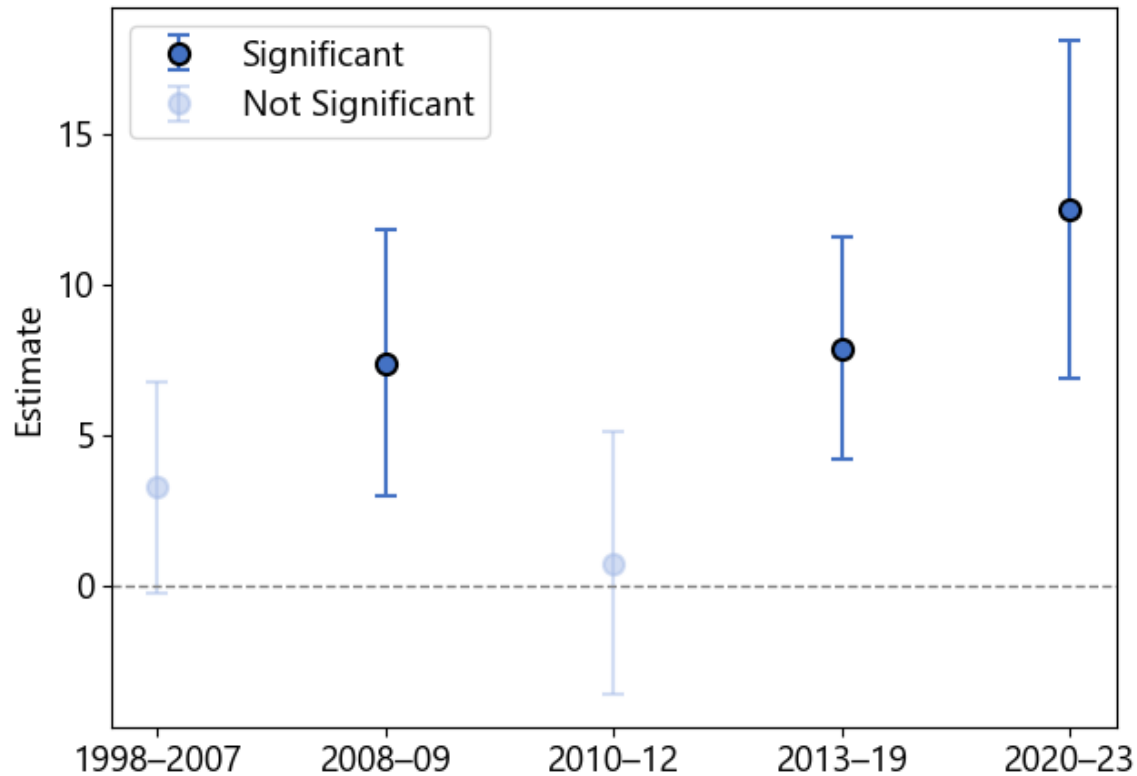
## Output Stabilization by Income Group



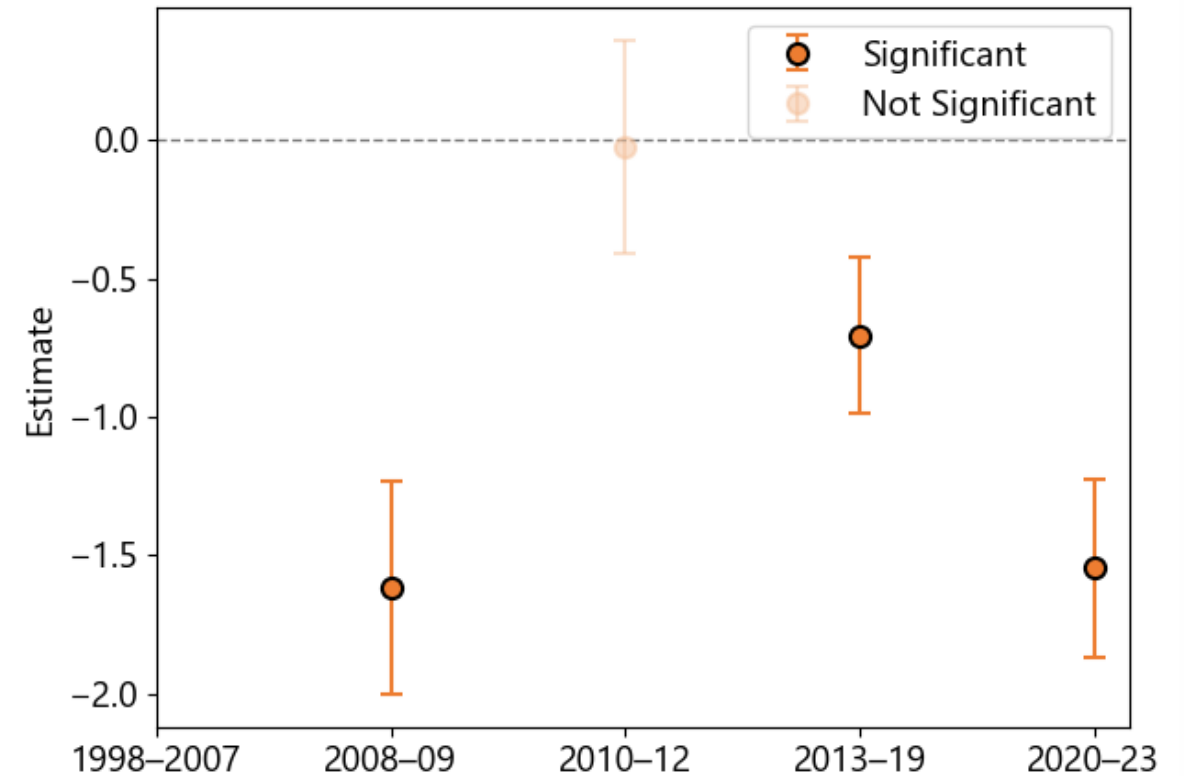
- IMF fiscal advice is **not procyclical**.
- Advice is **more countercyclical for AEs than for EMs**, but we obtain positive coefficients for both.
- For **LICs**, it is at best mildly countercyclical and most likely **acyclical**.
- ATE (ppts): AEs (2.3), EMs (1.1), LICs (0.2).

# Output Stabilization by Time Periods

## Output gap



## Year bucket constants

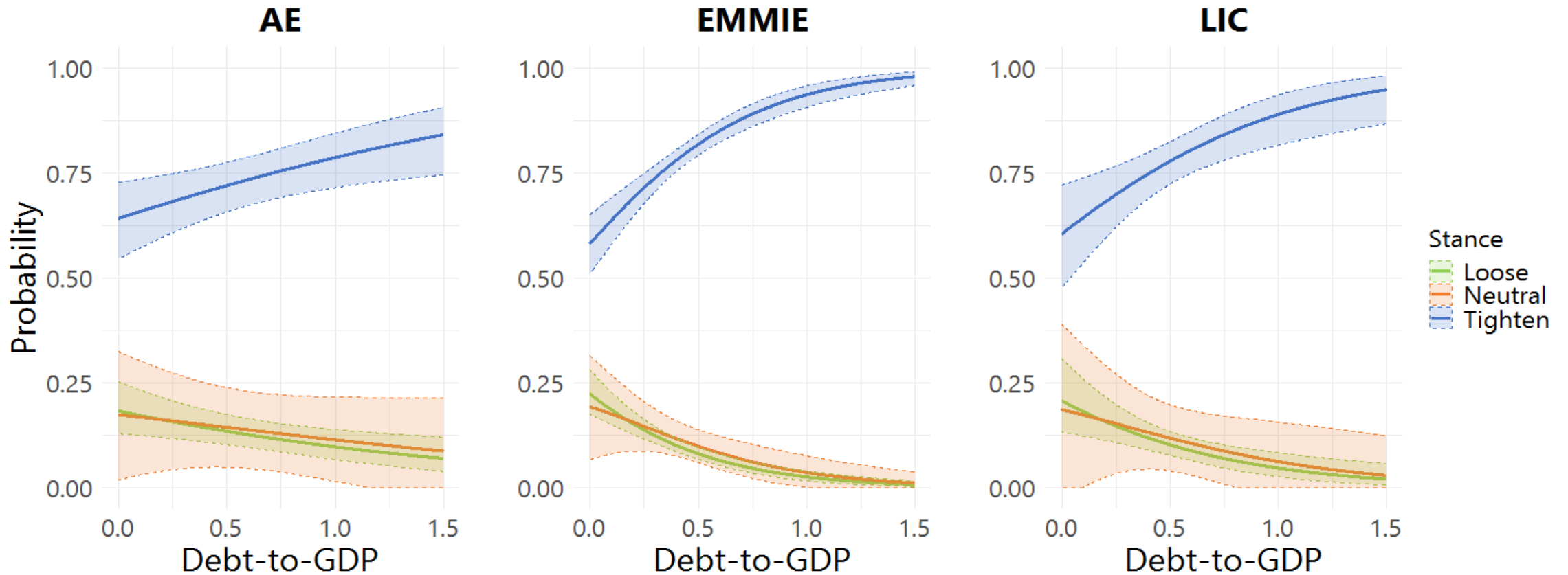


- **Over time**, advice has become more countercyclical.

## Output Stabilization: Other Dimensions

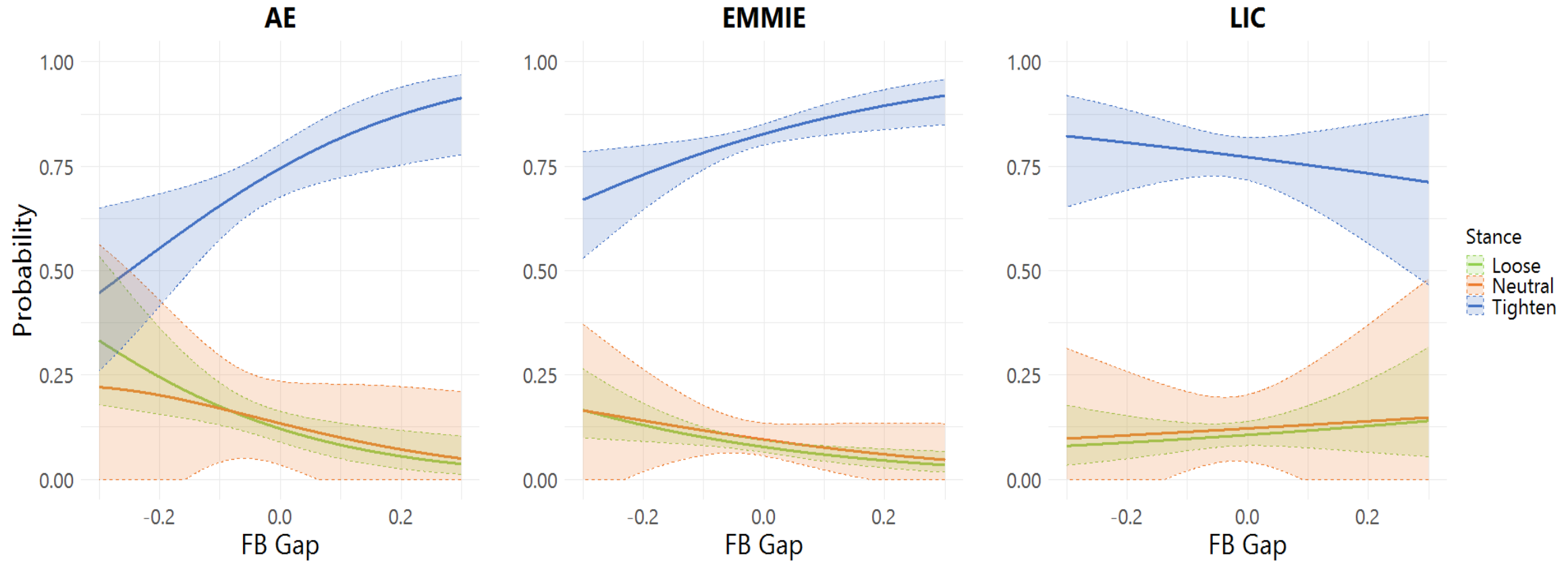
- ▶ Advice is more countercyclical for countries with lower **fiscal risks**.
- ▶ Countries with stronger **fiscal institutions** receive more countercyclical and less consolidation-oriented advice.
- ▶ Fiscal and monetary **policies** act as complements (not substitutes); no systematic link with macroprudential advice.
- ▶ Asymmetrical response to **external imbalances** (stronger response to deficits than surpluses).
- ▶ **Mission Chiefs** from “saltwater” schools give more countercyclical advice (Lang, Wellner & Kentikelenis, 2024).

# Fiscal Sustainability: Probability Plots of Debt-to-GDP by Income Group



- ▶ Nearly flat for AEs but steeper for EMs and LICs.
- ▶ **AME** (ppts): AEs (0.12), EMs (0.41), LICs (0.29).

# Fiscal Sustainability: Probability Plots of Fiscal Balance Gap by Income Group



- ▶ The fiscal balance gap is a more important explanatory variable for AEs.
- ▶ **AME** (ppts): AEs (0.75), EMs (0.50), LICs (-0.18).

# Conclusion

## Near-Term Fiscal Advice

- ▶ IMF advice is systematic rather than ad hoc, reflecting a coherent analytical framework.
- ▶ Recommendations explicitly balance output stabilization and fiscal sustainability, with the relative weight evolving over the cycle.
- ▶ Advice has become increasingly countercyclical over time, correcting earlier procyclical leanings.
- ▶ Countercyclicity is strongest when monetary policy is constrained (zero lower bound), fiscal institutions are stronger, and staff graduated from salt-water universities.

## Medium-Term Fiscal Advice

- ▶ Persistent emphasis on consolidation to anchor debt sustainability.
- ▶ Became more accommodative during secular stagnation in advanced economies.



**THANK YOU!**  
*Visit us at [ieo.imf.org](http://ieo.imf.org)*