INTERNATIONAL MONETARY FUND

Speaking to the Markets: The Role of IMF Announcements in Investors’ Confidence

Béatrice Sagna and Solo Zerbo

WP/25/258

2025
DEC

WORKING PAPER

 WP/25/258

IMF Working Paper
African Department

Speaking to the Markets: The Role of IMF Announcements in Investors' Confidence

Prepared by Béatrice Sagna and Solo Zerbo*

Authorized for distribution by Olaf Unteroberdoerster
December 2025

ABSTRACT: This paper examines the effect of IMF staff and executive board announcements on sovereign bond spreads across emerging and developing economies during economic uncertainty. We derive testable predictions from a stylized model in which IMF announcements serve as a signal to ambiguity-averse investors, narrowing their posterior beliefs about macroeconomic parameters and lowering the ambiguity premium in bond pricing. To empirically assess this, we train a large language model to identify and categorize economic signals by topic, assessing both the sentiment and the intensity of their coverage. The analysis yields several key insights. First, IMF announcements exhibit strong interdependencies among topics indicating a holistic approach to economic diagnostics. Second, using a local projection approach, we show that increase in the positivity of an IMF press release's tone leads to a significant reduction in sovereign spreads in the short term. This effect is more pronounced for countries with higher spreads, larger program size, and for topics such as debt, FX and rerseves, and fiscal.

JEL Classification Numbers:	D81, D83, E44, G12, G14, G15
Keywords:	Economic Uncertainty, IMF Announcements; Sovereign Spreads; Ambiguity Premium; Large Language Model; Local Projection

Author's E-Mail Address:	BSagna@imf.org; SZerbo@imf.org

* The authors would like to thank Montfort Mlachila, Papa N'Diaye, Anta Ndoye, Olaf Unteroberdoerster, Pilar Garcia Martinez, AFR seminar participants and IMF reviewers for their useful and insightful comments.

Speaking to the Markets : The Role of IMF Announcements in Investors' Confidence

Béatrice Sagna[*] Solo Zerbo[†]

December 12, 2025

Abstract

This paper examines the effect of IMF staff and executive board announcements on sovereign bond spreads across emerging and developing economies during economic uncertainty. We derive testable predictions from a stylized model in which IMF announcements serve as a signal to ambiguity-averse investors, narrowing their posterior beliefs about macroeconomic parameters and lowering the ambiguity premium in bond pricing. To empirically assess this, we train a large language model to identify and categorize economic signals by topic, assessing both the sentiment and the intensity of their coverage. The analysis yields several key insights. First, IMF announcements exhibit strong interdependencies among topics indicating a holistic approach to economic diagnostics. Second, using a local projection approach, we show that increase in the positivity of an IMF press release's tone leads to a significant reduction in sovereign spreads in the short term. This effect is more pronounced for countries with higher spreads, larger program size, and for topics such as debt, FX and reserves, and fiscal.

Keywords: Economic Uncertainty, IMF Announcements; Sovereign Spreads; Ambiguity Premium; Large Language Model; Local Projection.
JEL: D81, D83, E44, G12, G14, G15.

[*]International Monetary Fund. Email: BSagna@imf.org
[†]International Monetary Fund. Email: SZerbo@imf.org

The views expressed in IMF Working Papers are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. The authors would like to thank Montfort Mlachila, Papa N'Diaye, Anta Ndoye, Olaf Unteroberdoerster, Pilar Garcia Martinez, AFR seminar participants and IMF reviewers for their useful and insightful comments.

1 Introduction

Communication is widely acknowledged as a key driver of financial market expectations, influencing asset pricing in a forward-looking manner. The IMF has historically played an active and central role in providing economic diagnostics and policy advice to support the design of public policies in market economies. In times of heightened uncertainty, policymakers often turn to the IMF for guidance in navigating complex policy challenges. Consequently, the IMF's country specific announcements receive significant attention, while contributing to an informational echo chamber that may anchor investors' expectations.[1]

Although substantial research has explored the influence of central bank and media communications on financial markets, the impact of the informational content of IMF statements on countries' economic outlooks remains relatively unexplored. This paper aims to address this gap by investigating how IMF announcements affect investor confidence (size of ambiguity). In particular, our analysis focuses on the effects of statements issued by IMF staff and the executive board on sovereign bond spreads in emerging and developing economies (EMDEs). To the best of our knowledge, this is the first study to systematically document the post-announcement drift in sovereign bond spreads following IMF press releases.

Our paper links theory and data to address the research question. To formalize the role of IMF announcements in shaping market expectations, we begin with a stylized model of signal extraction and bond pricing, where investors update their beliefs about economic fundamentals by using information available to them in the market. We argue that the representative investor ("the market") is averse to ambiguity (*aka* Knightian uncertainty) and evaluates the bond pricing kernel using the worst-case posterior mean of key economic parameters (e.g., fiscal, debt, reserves, FX, economic growth). This behavior gives rise to an ambiguity premium.

Knight (1921) distinguishes between risk and ambiguity. While risk involves known probabilities, ambiguity arises in environment of heightening uncertainty, where assigning probabilities becomes impossible and leaving investors' beliefs shaped by the credibility of the information or data. Consequently, investors consider a range or family of plausible probability distributions associated with macroeconomic parameters. However, when an IMF press release arrives, it provides a tangible signal that shapes market participants' beliefs about these economic parameters. The tangible signal narrows the range of possible posterior means of economic parameters, reducing worst-case discounting. As a result, the pricing kernel becomes less sensitive to ambiguity, which in turn lowers the associated ambiguity premia. We show that even when the ambiguous signal is high, the posterior mean remains stable due to the weight of the tangible signal.

[1] See Review of the IMF's Communications Strategy (2024) for a discussion on the objectives, audiences, and scope of Fund announcements, drawing on its past decade's experience and its perspectives.

Next, we bring the model to the data by testing the implication of the mechanism outlined above. We leverage large language models (LLMs) to systematically extract economic signals embedded in IMF announcements, allowing us to quantify their informational content and assess their impact on investor expectations. We begin by constructing a novel dataset comprising press releases issued by IMF staff and the executive board. These documents, along with their publication dates, were obtained from the IMF's official news website. Our dataset includes 607 press releases spanning from 40 countries from 2020 to 2024. These press releases are typically published during program negotiations, staff visits, program reviews, surveillance missions, and board meetings, and often report on discussions related to macroeconomic developments. They constitute a unique and official source of information on the outcomes of the IMF's economic diagnostics and are regarded by market participants as tangible information.[2]

With the textual data in hand, we extract contextualized sentence representations and classify the content of the announcements into well-defined topics. Specifically, we fine-tune and train a sentence transformer model to categorize the press releases into key economic topics: debt, economic growth, fiscal policy, structural reforms and governance, climate, monetary policy, and foreign exchange and reserves. These classifications are subsequently used to construct novel indices that capture both the intensity of topic coverage and the sentiment expressed.

Having measured the intensity of coverage and the sentiment of topics, we apply a local projections approach (Jordà (2005)) to estimate the impulse responses of sovereign spreads after an IMF announcement over a 30-day horizon. We assess the cumulative impact of the announcements on sovereign spreads, examining both sentiment and topic perspectives while controlling for aggregate uncertainty as proxied by the expected volatility of the market (VIX) and other macroeconomic variables. We find that IMF announcement days are special and have a significant influence on financial markets over an extensive time horizon.

Among various types of IMF announcements, the sharpest decline in sovereign spreads typically occurs after a staff-level agreement is reached on a new program or review. The response from staff-level agreement communication remains significantly negative (around -60 bps) throughout most of the 30-day period, suggesting that the credibility and reassurance provided by IMF involvement have a lasting calming effect on sovereign risk perceptions. The stronger reaction of sovereign bond markets can be attributed to the fact that these announcements often represent exogenous policy intervention shocks. In other words, they convey new and validated information that is not already reflected in the market's pricing and are typically perceived by investors as tangible signals of IMF institutional support, policy commitment, or forthcoming economic reforms.

[2] Notably, communication is especially important in key program countries, where there is a greater demand for transparency.

We also find that bond spreads respond to variations in the press release's tone. Conveying a more positive tone by one standard deviation could reduce the spread by approximately 55 bps in the short term, and the effect persists—though it attenuates slightly—over the 30-day period. We further observe that the decrease in spreads after an IMF press release is stronger for countries with higher spreads and is more pronounced when the press release refers positively to topics such as debt (∼220 bps), fiscal policy (∼110 bps), FX and reserve (∼105 bps). These findings suggest that investors assign different informational value to topics and are more responsive to topics that directly affect a country's repayment capacity—such as debt sustainability, fiscal policy, and foreign exchange reserves.

Moreover, building on prior research, our results indicate that bond market reactions to IMF announcements are notably stronger than those observed for other types of fiscal-related announcements. For instance, David et al. (2022) find that austerity-related announcements in emerging and developing economies (EMDEs) lead to a modest reduction in sovereign spreads—around five basis points over a 30-day horizon—though the effect is economically small and only marginally statistically significant as highlighted by the authors. Conditioning on the announcements from the legislature, they find a decline of about 15 bps. Finally, we show that joint consideration of sentiment and topic intensity yields a further decline in sovereign spreads only for the "debt" topic. In addition, a joint consideration of the sentiment and level of disbursements induces a further decrease in spreads.

Our results are robust to a series of tests, encompassing (i) placebo tests consisting of pseudo press releases days, (ii) central banks' monetary policy announcements, (iii) various definitions of robust standards errors including Newey and West (1987), and (iv) alternative measure of sentiment. Overall, our findings highlight the macro-critical importance of managing expectations in policy development, as sovereign spreads reflect not only government actions but also the perceived credibility of those actions.

2 Literature

This paper contributes to at least three strands of the literature. First, our paper is related to the body of research on factors influencing investors' confidence. Piazzesi et al. (2006) show that under the assumption that agents prefer an early resolution of uncertainty, inflation as bad news for future consumption growth should help generating upward-sloping nominal yield curve. The intuition is that a positive surprise to inflation lowers future consumption growth and at the same time decreases the real payoff of long-term nominal bonds. Bansal and Shaliastovich (2013) show that a long-run risk model with time-varying volatilities of expected consumption growth and inflation can account for bond return predictability. In their model, inflation has time-varying effects on investors' confidence. Zhao (2017) shows that investor confidence about future consumption growth is driven by past consumption growth and inflation. Using the Blue-Chip Financial Forecast (BCFF) dispersion as an empirical measure for the size of ambiguity, Zhao (2020) finds that in the pre-2000 period, the size of ambiguity for long-horizon inflation is bigger than it is for short horizons, and the term structure of ambiguity is reversed afterward. Similarly, Epstein and Schneider (2008) show that investors dislike assets for which information quality is poor, especially when the underlying fundamentals are volatile. These effects induce ambiguity premia and investors overreact to bad news and underreact to good news when the signal is ambiguous. Unlike these studies, we argue that investors' confidence is shaped by expectations about a country's future policy direction, as inferred from IMF announcements.

Second, this paper also contributes to the growing body of literature on economic narration through the application of natural language processing (NLP) methods to measure monetary policy shocks and stances from central bank communication (Hansen et al. (2019), Handlan (2020), Aruoba and Drechsel (2024), Sharpe et al. (2023), Schmeling and Wagner (2025)), media narratives (Goetzmann et al. (2022), Dim et al. (2023)) , financial market sentiment (Soo (2018), Ke et al. (2019)) and uncertainty and geopolitics risk (Baker et al. (2016), Caldara and Iacoviello (2022)). While early research in NLP primarily relied on simple word counts, recent advancements in natural language processing—particularly transformer-based deep learning architectures pretrained on Web-scale text corpora—have significantly enhanced the capabilities of LLMs, shifting from context-free to context-dependent models. Our paper builds on state-of-the-art context-dependent models that enable a more advanced and precise textual approach. This evolution has gained prominence in economic analysis (Minaee et al. (2024), Dell (2025)) and the measure of economic narration (Chen et al. (2024), Leek and Bischl (2025)). As for policy shocks, unlike the central bank communication literature, which generally relies on monetary policy (interest rate) announcement shocks to study the impact on key macrofinancial variables (such as stock prices), we use the IMF press release as our policy intervention shock to assess the effect on sovereign spreads.

Finally, our work relates to the macro-finance literature that examines the implications of policymakers' announcements for asset prices. A large body of research studies the impact of the Federal Open Market Committee (FOMC) announcements on the cross-section of assets and market variables such as long-term real and nominal interest rates (Hanson and Stein (2015), Herbert et al. (2024)), equity returns (Gorodnichenko and Weber (2016), Narain and Sangani (2023)), various financial assets prices (Gorodnichenko et al. (2023), Curti and Kazinnik (2023)). Leombroni et al. (2021), using high frequency data show that monetary policy communications by the European Central Bank on regular announcement days led to a significant yield spread between peripheral and core countries during the European sovereign debt crisis by increasing credit risk premia. David et al. (2022) build a database to pinpoint the exact announcement dates of fiscal consolidation measures made by the executive (i.e., president or finance minister) or legislature (congress or parliament). Their analysis examines the impact of these announcements on sovereign spreads across a panel of 21 emerging market economies between 2000 and 2018. In contrast to these papers, we study the "IMF information effect". Fratzscher and Reynaud (2011) present the most comparable study to ours, employing a subjective method to classify and analyze the content of Article IV Public Information Notices (PINs) and its impact on financial markets, with a particular emphasis on the influence of political economy factors in IMF surveillance.

The rest of the paper is organized as follows. Section 3 introduces a framework of ambiguity, tangible information and bond pricing. Section 4 introduces our LLM approach and provides an anatomy of IMF communication. Sections 5 and 6 discuss the results of the empirical analysis. Section 7 concludes. The appendix provides additional tables and figures.

3 Theoretical Background: A Stylized Model of Signal Extraction and Bond Pricing

To illustrate how IMF announcements could impact sovereign spreads, we extend the framework of Epstein and Schneider (2008) in two keyways. First, we adapt it to the context of sovereign bond pricing under ambiguity aversion. Second, we introduce two types of signals into the model: ambiguous (opaque) news and tangible (credible) news. The dual signal approach is crucial in environments of uncertainty, where the interpretability and credibility of information significantly influence investor behavior and asset pricing. This setup deviates from Bayesian updating models, in which the quality of information has no impact on prices because the representative investor always knows the precision of signals and update beliefs accordingly.

In our model, the representative investor favors coherence and is ambiguity-averse when assessing the precision of economic signals. Let's denote by θ a macroeconomic parameter that the investor aims to learn about. We assume that he has a unique normal prior, that is, $\theta \sim \mathcal{N}(0,\sigma_\theta^2)$. There can be two types of news available in the market related to θ : news with an ambiguous signal (s) and news with a tangible signal (υ). The signal (s) can be ambiguous for several reasons: It may be vague or imprecise, because it may lack relevant or verifiable information altogether (e.g, unclear perception of fiscal or monetary policies or lack of visibility on debt restructuring). Another source of ambiguity could arise when the signal fails to reinforce prevailing narratives or expectations, meaning that it doesn't create the kind of informational echo chamber that can amplify investor confidence. In the ambiguous signal, the precision σ_s^2 is unknown. Consequently, the representative investor associates with the signal s, which pertains to the parameter θ, a range of precisions ($[\underline{\sigma}_s^2, \bar{\sigma}_s^2]$) rather than a single well-defined precision, although there is a unique prior belief about θ.

$$s = \theta + \epsilon_s, \qquad \epsilon_s \sim \mathcal{N}(0, \sigma_s^2) \qquad and \qquad \sigma_s^2 \in [\underline{\sigma}_s^2, \bar{\sigma}_s^2] \tag{1}$$

However, the precision of the tangible signal σ_v^2 is known ($< \sigma_s^2$). The tangible signal υ is defined as:

$$\upsilon = \theta + \epsilon_\upsilon, \qquad \epsilon_\upsilon \sim \mathcal{N}(0, \sigma_\upsilon^2) \tag{2}$$

As the signal υ is not always available to the representative investor, the signal s introduces uncertainty regarding the parameter θ, making him to distrust any single precision (σ_s^2). Hence, he evaluates outcomes using the worst-case scenario when facing low quality information. When the signal υ becomes available, he updates his belief about θ by forming a weighted average of the dual signals across all likelihoods to obtain a family of posterior means (μ).

$$\mu = \frac{\tau\upsilon + \gamma(\sigma_s^2)s}{\tau + \gamma(\sigma_s^2)}, \qquad \sigma_s^2 \in [\underline{\sigma}_s^2, \bar{\sigma}_s^2] \tag{3}$$

$$with \qquad \tau = \frac{1}{\sigma_\upsilon^2} \qquad and \qquad \gamma(\sigma_s^2) = \frac{1}{\sigma_s^2}.$$

The more credible the tangible signal (i.e., the smaller σ_υ^2), the more weight it receives. When $\tau \to \infty$, the posterior mean converges to υ, meaning the tangible signal dominates. Even if the ambiguous signal is high, the posterior mean remains stable due to the weight from the tangible signal.

For a zero-coupon bond that pays 1 at time T, the price at time t under the worst-case scenario is:

$$P_t = \min_{\sigma_s^2 \in [\underline{\sigma}_s^2, \bar{\sigma}_s^2]} \mathbb{E}[M_{t,T}] \tag{4}$$

Where the pricing kernel (or stochastic discount factor) used to compute the present value of future payoffs can be defined as:

$$M_{t,T} = exp\left(-\int_t^T r(r_0, \mu_{s,\upsilon}(\sigma_s^2, \sigma_\upsilon^2))du\right) \tag{5}$$

For simplicity, we assume that the short-term rate is affine in the perceived signal variance.

$$r(r_0, \mu_{s,\upsilon}(\sigma_s^2, \sigma_\upsilon^2)) = r_0 + \phi\mu_{s,\upsilon}(\sigma_s^2, \sigma_\upsilon^2)$$

ϕ captures the sensitivity of the short-term rate to changes in the posterior mean and r_0 the risk-free rate. We can rewrite the bond pricing as follow:

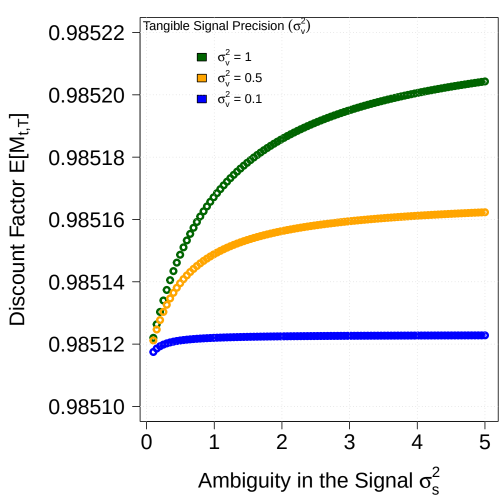
$$P_t = \min_{\sigma_s^2 \in [\underline{\sigma}_s^2, \bar{\sigma}_s^2]} \mathbb{E}[M_{t,T}] = \min_{\sigma_s^2 \in [\underline{\sigma}_s^2, \bar{\sigma}_s^2]} exp\left(-(T-t)r_0 + \frac{(T-t)^2\phi^2}{2} \underbrace{\left[\frac{\sigma_\upsilon^2 \tau + \sigma_s^2 \gamma(\sigma_s^2)}{\tau + \gamma} \right]}_{\text{Ambiguity correction term}} \right)$$

Figure 1 presents a simulation exercise illustrating how bond prices respond under a worst-case scenario to varying levels of ambiguity in the signal, across different levels of tangible signal precision and under plausible parameterization of ϕ and r_0. $\mathbb{E}[M_{(t,T)}]$ is an increasing function over σ_s^2. Higher ambiguity (σ_s^2) leads to a higher discount factor, reflecting a higher ambiguity premium, while more precise tangible signals (lower σ_υ^2) result in lower discount factor, even under high ambiguity.

Whenever a tangible signal is available, it anchors the posterior mean by narrowing the range of possible $\mu_{(s,t)}$ reducing the worst-case discounting. When $\tau \rightarrow \infty$, the representative investor pricing kernel depends only on the signal υ.

$$\lim_{\tau \rightarrow \infty} M_{t,T} = exp[-(T-t)(r_0 + \phi\upsilon)] \tag{7}$$

Figure 1: Discount Factor: Ambiguity Premium vs Tangible Signal



Notes: Simulated relationship between ambiguous signal, tangible signal and bond price with ϕ = 2% and r_0 = 3% and assuming a time horizon $(T-t)$ of 1 year. The x-axis shows the bond price under the wort case scenarios, and the y-axis shows the ambiguity signal variation. Each curve represents a different level of bond price based on the tangible signal value.

In the following section, we bring the model to the data by assuming IMF press releases as a tangible signal. We leverage on LLMs, with their advanced capability to understand context to uncover narratives within IMF press releases on key macroeconomic parameters.

4 Measuring Economic Signals with Large Language Model

This section examines the semantic content of IMF announcements. Our approach involves two main steps: first, we identify a set of key topics relevant to IMF core mandate; second, we apply a sentence embedding technique to capture the underlying semantic meaning of the economic narratives. This approach allows us to quantify the informational content of IMF communications.

4.1 Data

We collect IMF press releases issued by staff and the executive board between 2020 and 2024, focusing on programs, surveillance activities, and board reviews. These documents, along with their publication dates, were sourced from the IMF's official website. We focus on the 2020–2024 period for two key reasons. First, this timeframe is characterized by a series of global shocks that contributed to heightened economic uncertainty—including the COVID-19 pandemic, Russia's invasion of Ukraine, and rising geopolitical tensions. It also witnessed a surge in global inflation, aggressive monetary policy tightening, banking sector stress, climate-related disruptions, and growing risks of geoeconomic fragmentation. Second, this period aligns with the preference shock described in section 3, in which the representative investor's primary concern is uncertainty; resulting in investors' beliefs that are shaped predominantly by the perceived credibility of available information.

Our dataset includes 608 press releases—308 from the Board and 299 from Mission Chiefs[3]—spanning 40 countries in emerging and developing economies. The geographic distribution covers Sub-Saharan Africa (12 countries), South America (14), Europe (3), the Middle East and North Africa (9), and Asia Pacific (2). For program review-related releases, we supplemented the data with disbursement amounts from the IMF's Monitoring of Fund Arrangements (MONA) database.[4] Figure 2 provides the distribution of the press release across sources and document types.

We then match the publication dates of these press releases with information on

[3]The staff prepares country's reports and submit them to the executive board. Based on that, the Board issues its assessment.

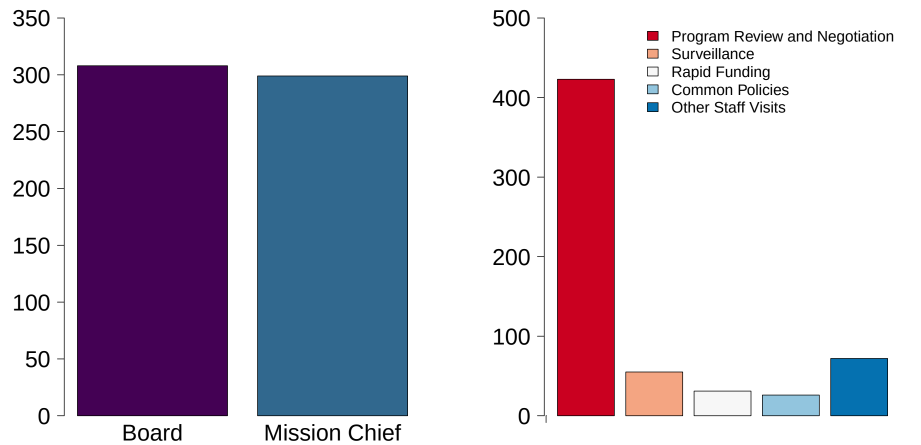
[4]Under an IMF program, several lending instruments are available, each tailored to address different types of balance of payments needs—whether potential, short-term, or medium-term—as well as specific country circumstances.

sovereign bond spreads sourced from Bloomberg. The bond data adheres to several criteria: they must be denominated in USD or euros and feature a fixed rate or step coupon. Additionally, the outstanding amount must be at least USD/Euro 250 million, and the bonds must have either bullet maturities or be amortizing. These bonds must also meet minimum pricing quality standards, have at least one-year remaining maturity for bullet bonds or 18 months for amortizing bonds, and not be covered by external guarantees .

Figure 2: Press Release Types

(a) Distribution of Press Releases by Source

(b) Distribution of Press Releases by Document Type



Notes: This figure plots the total number of press releases by type of missions and by managers. The sample comprises of 607 press releases published from 2020-2024 across 40 countries.

After mapping the dates of the press releases with the sovereign spreads data, we are left with 447 press releases for which consecutive daily bond spread information is available for up to 30 days. To enhance our analysis, we supplement the press releases and sovereign spreads data with country-specific macroeconomic variables and global market sentiment indicators to capture a more nuanced understanding of the factors influencing sovereign spreads. The inclusion of global sentiment indicators ensures that we also control for global market factors during IMF communication on a country economic performance, thereby allowing to better gauge the confidence effects of IMF communication on a country's sovereign spreads. We provide in the appendix a detailed descriptive statistic of the data, offering insights into the characteristics and distribution of the variables included in our study.

4.2 Anatomy of IMF Announcements

4.2.1 Semantic Density

The IMF's country specific announcements are organized around key economic topics that are essential for assessing the nation economic outlook. Fiscal policy and debt sustainability, monetary policy, financial stability, structural reforms including governance issues are key to the IMF's mandate to promote macroeconomic and financial stability. These areas underpin the IMF's continued efforts to deliver policy recommendations, provide technical assistance, mitigate economic vulnerabilities, and promote sound fiscal and financial practices.

Given the breadth of macroeconomic issues covered, mapping each IMF press release to a candidate topic represents a significant challenge. To address this, we tokenize each document into individual sentences, enabling a more granular analysis. The sentence-level approach reveals that IMF announcements exhibit high semantic density meaning a sentence can reference multiple and interrelated topics. For example, a single sentence may simultaneously address fiscal and debt-related concerns, underscoring the interconnected nature of economic policymaking. Such communication style ensures that the multifaceted dimensions of economic policy are captured, allowing for a more comprehensive understanding of the issues at hand. By embedding multiple topics within individual sentences, press releases effectively convey the complexity of economic challenges and the necessity for integrated, cross-cutting solutions.

To have a representative set of candidate topics, we selected five single-topic sentences, each corresponding to a core macroeconomic theme: fiscal policy, debt sustainability, economic growth, financial stability, and monetary policy. To account for the semantic density often present in IMF communications, we also included four mixed-topic sentences that reflect the intersection of related policy areas: fiscal and debt, foreign exchange (FX) and monetary policy, governance and structural reforms, and FX and reserve Management. In addition, we incorporate one emerging-topic sentence related to climate change, recognizing its growing relevance in macroeconomic discourse especially over the last years. This selection ensures a comprehensive and nuanced sample of the diverse economic challenges countries may face. Finally, we create a separate category labeled "others". This category helps us distinguish economic near and medium term economic narratives from quantitative indicators related to program performance (quantitative performances, structural benchmarks, and indicative targets), as well as from non-economic content like staff expressions of gratitude to the authorities during missions. This approach ensures that our classification focuses on relevant economic analysis while filtering out IMF program performance as well as unrelated or procedural information. Table 1 below provides an illustrative example of a sentence associated with each candidate topic.

Table 1: Candidate Topics

Candidate Topic	Example of Sentence
Fiscal	In this context and following a request to Congress to make use of the emergency clause included in the Fiscal Responsibility Law, the program envisages a fiscal deficit of the Non-Financial Public Sector of 4 percent of GDP in 2020 and budget reallocations of non-priority current expenditure.
Debt	The authorities have secured debt reprofiling agreements from several large creditors to reduce risks related to debt sustainability.
Economic Growth	GDP growth is expected to remain strong in 2024, driven by dynamism in tourism and related sectors.
Financial Stability	Effective financial sector supervision has contributed to preserving financial stability and improving financial development.
Monetary Policy	The central bank has appropriately lowered the policy rate, and its data-dependent, forward-looking approach should continue to help inflation rise back to target.
Fiscal and Debt	Looking ahead, continued strong commitment to fiscal consolidation over the medium term remains key to reduce debt vulnerabilities.
FX and Reserve	Returning to a market-determined exchange rate and rebuilding FX reserves.
FX and Monetary	The Central Bank will continue to support the ongoing disinflation process and will take any necessary action to ensure that there are no undue pressures on the exchange rate.
Governance and Structural Reforms	Achieving strong and inclusive growth rests on steady progress on structural reforms to support female labor force participation, enhance youth employment and labor market flexibility, promote competition, reduce the costs of doing business, and strengthen governance and transparency
Climate	The central bank's roadmap to integrate climate change considerations into its core activities is commendable.
Others	The IMF mission held constructive discussions with the authorities and reached staff-level agreement on policies needed to complete the fourth review under the PCI.

Notes: This table provides examples of sentences identified to train the LLM.

4.2.2 Using Sentence Transformer Fine-Tuning to Classify IMF Announcements

The foundational statistical technology in artificial intelligence (AI) is the large-scale transformer network, which enables efficient processing and understanding of complex data patterns. To classify our IMF communication corpus, we rely on a sentence transformer fine tuning (SetFit) approach introduced by Tunstall et al. (2022) to generate contextualized embeddings. SetFit represents an innovative approach within the AI sphere and has gained prominence within Natural Language Processing (NLP).

a) **Sentence Transformer**

Sentence transformer is based on a deep learning architecture introduced by Vaswani et al. (2017). It is an encoder-only LLM that focus on understanding and analyzing a given input and producing task-specific outputs, such as labels and classification.[5] Compared to other LLMs, the sentence transformer offers two key advantages. First, like encoder models, sentence transformers are based on a bidirectional architecture, which enables these techniques to interpret context by analyzing both directions within a sentence (Devlin et al. (2019)). This bidirectionality facilitates a comprehensive understanding of sentence meaning, making them particularly well-suited for tasks demanding fine-grained semantic analysis. By comparison, decoder-only LLMs (e.g., GPT, Llama, Claude,...) excel in text generation tasks but are less efficient for direct classification tasks, often relying on additional prompt engineering techniques such as chain-of-thought reasoning (Wei et al. (2022)) to achieve comparable results in limited datasets.

Second, sentence transformers produce fixed-size, semantically rich sentence embeddings, which differ from standard transformer encoders. For example, BERT generates contextualized token-level embeddings that require pooling strategies to aggregate into sentence-level representations. Conversely, sentence transformers are specifically trained to encode complete sentences into dense vector spaces that preserve semantic relationships. This feature is critical for our analysis of IMF communications, as it emphasizes capturing the overarching meaning of entire sentences rather than focusing on token-level details. This capability makes sentence transformers particularly well-suited for our classification task, where understanding the underlying meaning of economic narratives is essential. For example, Table 2 illustrates the superior semantic sensitivity of contextualized embeddings. While context-free models may overestimate similarity due to shared vocabulary, sentence transformers better capture meaningful differences in context and intent.

Table 2: Cosine Similarity Score: Context Free Embedding vs Contextualized Embedding

Sentence	Context Free Score	Sentence Transformer Score
"How are you" "How old are you"	0.9	0.1

Notes: Authors' Illustration. A score close to 1 indicates that the two sentences are highly similar, while a score close to 0 suggests they are dissimilar.

b) **Fine-Tuning: From a Non-Specialized to Specialized LLM on Economic Topics**

[5] There are three type of transformer architecture-based classification: i) The encoder -only Transformer that encodes input sequences for classification tasks; ii) The decoder-only Transformer that generates text iteratively using self-attention mechanisms and (iii) the Encoder-Decoder Transformer that converts input sequence into meaningful output sequences.

We follow a three-step phase to fine-tune the sentence transformer, consisting of pre-training, few-shot learning, and generalization. Each phase plays a distinct role in progressively adapting the model—from general language understanding to a specialized model capable of understanding economic and financial languages for our classification task.

In the pre-training phase, we allow the model to learn millions of parameters by analyzing vast and diverse datasets like Wikipedia, Common Crawl, and WebText. This step enhances the model's ability to understand language syntax and semantics, including word meanings, contextual usage, and overall linguistic structure. We use the pre-trained LLM "all-mpnet-base-v2" architecture, which leverages a Multi-Perspective Network to capture multiple dimensions of a sentence. The "all-mpnet-base-v2" model has approximately 110 million parameters (embeddings layer weights, self-attention weights, ...). While the pre-training endows the model with language understanding capability, it's not inherently expert in financial and economics topics.

To enhance proficiency in financial and economic topics, we expose the "all-mpnet-base-v2" model to a limited amount of labeled data using a few-shot learning technique. The Few-shot learning serves as a crucial step and involves updating the pretrained model's parameters using a new task-specific dataset and a gradient-based optimization. To implement this approach, we construct a labeled dataset comprising 385 sentences derived from the candidate topics. Specifically, we identify 35 representative sentences per topic through expert judgment, ensuring balanced coverage across countries and key macroeconomic themes. This curated dataset provides a diverse and contextually rich foundation for fine-tuning the re-trained model, enabling it to adapt quickly to our classification task with minimal supervision.

We also adopt a holdout validation strategy to improve the performance of the model. We use 85 percent of the labeled dataset as a holdout test (validation) set, which remains completely unseen during model training and hyperparameter tuning. This ensures an unbiased assessment of the model's generalization capability. The remaining 15 percent of the data is used as the training set. This subset is employed for both model training and hyperparameters or learning parameters selection. The hyperparameters guide how the learning happens and are crucial to strike the right balance between memorization and generalization. We focus on three key hyperparameters which are pivotal in shaping how our model learns. These are the learning rate, epochs and batch.

- *The Learning Rate:* controls the magnitude of the model parameters updates and can indirectly affect how well the model generalizes to unseen data. If the learning rate is too large, the model might be overfit to the training data, performing poorly on new examples. A well-chosen learning rate helps the model find a balance between fitting the training data and generalizing to new data.

- *The epochs:* controls the number of times the model sees the training data for gradual learning and refining the weight of the parameters overtime. More epochs allow for deep learning and refinement, but excessive repetition might lead to memorization instead understanding.

- *The batch:* controls for the number of training example proceed together before the model update its parameters. A small batch size gives detailed focus but might be slow while a large batch size speeds things up but might miss key details.

In other words, the epochs can be interpreted as the number of times a student reviews a textbook, while the learning rate is how quickly he absorbs and applies new information during each study session. The batch controls how many pages or chapters the student review at the time. We set a tuple (2.5e-5, 1e-5) learning schedule in which the learning rate start at 2.5e-5 and decrease to 1e-5 during training to help models converge more effectively. For the epoch and the batch hyperparameters, we perform a grid search over combinations of the hyperparameters epochs $\beta_e \in \{4, 15, 20\}$ for gradual learning and refining the weight of the parameters overtime, and batch sizes $\beta_{bs} \in \{4, 16, 32\}$ for the data ingestion strategy. We select the hyperparameters that yield the highest average F1 score.

To further guide the model during the training and enhance its predictive capabilities for the classification task, we employ the cosine similarity loss function commonly utilized for training models that generate sentence embeddings. Its core objective is to minimize the cosine distance between embeddings of similar sentences while maximizing the distance between embeddings of dissimilar ones. The cosine similarity Loss is well-suited for tasks like semantic search and sentences clustering, where understanding nuanced textual similarities is critical. Its robust performance in preserving semantic meaning makes it a widely adopted choice for embedding-based applications. We also apply an alternative function, the Softmax loss, yet the model's performance remains unchanged. Finally, we generalize the model to unseen data. Annex B provides a comprehensive overview of the Sentence Transformer fine-tuning process.

4.2.3 Model Evaluation

We evaluate the performance of our model using standard classification metrics: precision, recall, and F1-Score. These metrics provide a comprehensive view of the model's ability to correctly identify relevant instances (precision), capture all relevant instances (recall), and balance the trade-off between the two (F1-score). We also compare the performance of our model against two alternative LLMs namely the knowledge transfer and the knowledge distillation techniques.

In the knowledge transfer approach, we eliminate the need for a human-labeled training dataset by leveraging zero-shot learning. The model draws on its pre-training on large-scale, unlabeled corpora. This enables the model to perform the classification task without any task-specific fine-tuning, relying solely on its generalized linguistic knowledge. In the knowledge distillation approach, we employ a smaller, more efficient model—all-MiniLM-L6-v2 from the Sentence-Transformers library—as the student model. Knowledge is transferred from a larger, more complex teacher model to this student model. This process allows the student to approximate the performance of the teacher while significantly reducing computational overhead, making it suitable for deployment in resource-constrained environments.

Table 3 reports the performance of each model. The validation scores are calculated on the holdout sample of 220 sentences for the Few-shot learning and the knowledge distillation models. The first row shows the performance of the knowledge transfer model. The second-row reports that our baseline Sentence transformer model. Finally, the third row shows the knowledge distillation model performance in classifying our corpus. Across all metrics, the Few-shot Learning model outperforms the Knowledge Distillation and the Knowledge Transfer approach. It achieves an accuracy of 86 percent and F1 score of 0.87 indicating strong capability in correctly suggesting identifying true positives while minimizing classification errors. In contrast, the Knowledge Transfer shows significantly weaker performance across all metrics. To further assess the performance of our model, we present the confusion matrices in Annex A, which provide a detailed view of the model's classification behavior. For example, classes such as "Economic Growth" and "FX and Monetary Policy", indicate high accuracy in predictions, showing these classes are correctly identified more often, while class such as "Financial Stability" has the highest misclassifications, indicating they are more often confused with other topics.

Table 3: Out-of-Sample Performance

Model	Size (# of Parameters)	Precision	Recall	F1-Score
Knowledge Transfer	110 million	0.17	0.10	0.10
Few-shot Learning (Baseline)	110 million	0.88	0.87	0.87
Knowledge Distillation	22 million	0.83	0.84	0.83

Notes: The table reports the out-of-sample performance the different models. The test sample consists of 165 sentences. The validation scores are calculated on the holdout sample of 220 sentences. The precision measures how many of the positive predictions were actually correct while the recall measures how many of the actual positive cases were correctly identified. The F1-score balances both, ensuring that neither precision nor recall is disproportionately low.

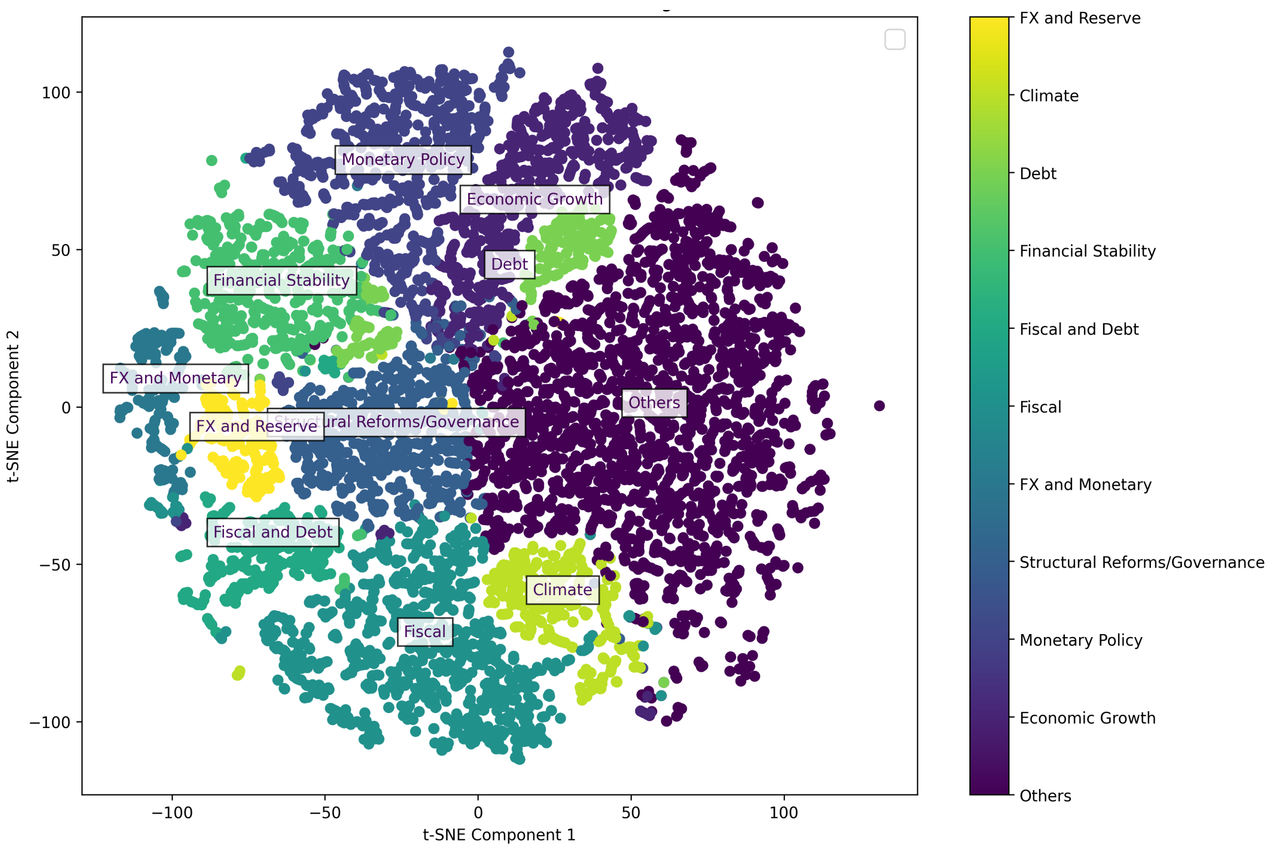
4.2.4 Semantic Gravity: Mapping the Landscape of IMF Announcements

To deepen our understanding of our LLM output, we evaluate the semantic gravity of the corpus, which captures the degree of interconnectedness among topics. This is achieved by applying t-distributed Stochastic Neighbor Embedding (t-SNE), a dimensionality reduction technique that maps the high-dimensional representations produced by the LLM into a two-dimensional space. The visualization facilitates the examination of semantic relationships between topics, revealing how closely related or distinct different topic clusters are in the embedding space of staff and the executive board discourses. Figure 3 paints the t-SNE output. Each point on the graphic represents a sentence embedding, and points are colored according to their topical categories. The x-axis and the y-axis represent the two principal components derived from the t-SNE algorithm that best capture the variance in the data.

Points that are close together within the same color cluster indicate that the sentences share similar content. For example, the cluster for “FX and Reserve” might be tightly grouped, suggesting that sentences within this category have strong thematic coherence. Conversely, if points within a category are more dispersed, it might indicate a broader range of subtopics or less thematic consistency within that category. For example, the “Others” category in the t-SNE visualization is represented by sentences that do not fit into the predefined 10 categories. The dispersion of these points suggests a diverse range of topics and themes that are not strongly related to the main categories listed in the legend. The scattered nature of the “Others” category indicates that these texts cover a wide variety of subjects, potentially including niche topics and miscellaneous discussions.

Finally, the proximity of clusters can reveal how different topics interact or influence each other. For instance, clusters for “Growth” and “Debt” are close to each other, indicating a strong relationship between these topics. In addition, the “Structural reforms and Governance” cluster is in the center and situated near the other clusters. This proximity indicates that discussions around governance and reforms often involve considerations of growth, fiscal, financial and debt, etc.

Figure 3: Discourse Landscape of the IMF Staffs and Executive Board



4.2.5 Measuring Topics' Attention and Sentiment

To measure topic attention, we employ a sentence-based frequency analysis to quantify the intensity with which various topics are discussed within the text. This method involves identifying and counting the number of sentences associated with each predefined topic. The frequency of each topic is then calculated as a proportion of the total number of sentences, providing a normalized measure of how prominently each topic features in the communication. For instance, if a particular topic sees an increase in sentence frequency, it may indicate heightened attention or concern from policymakers. Conversely, a decrease might suggest a waning interest or a shift in focus to other areas.

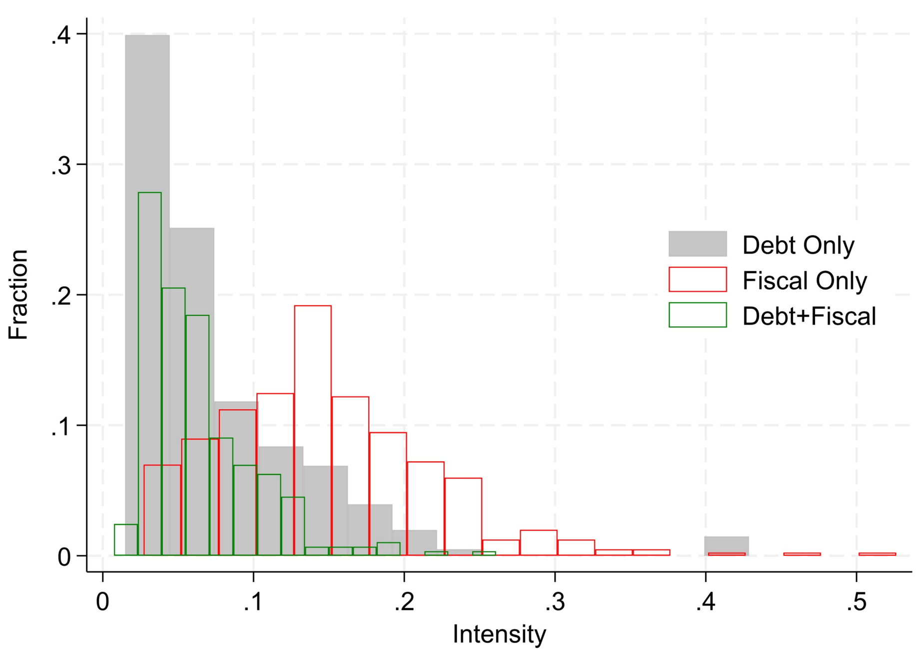
Let's denote I_{it}^{k}, which represents the rate at which sentences pertaining to topic k are mentioned as a share of the total sentences in the press releases for country i at time t, and $i \in J$ indexing all sentences that belong to the press release i. This measure can be interpreted as the average intensity with which the mission chief or the Board pays attention to the economic conditions of country i.

$$I_{it}^{k} = \frac{\sum_{i \in J} \mathbb{I}(Topic\, k\, Sentences_{it})}{\sum_{i \in J} Total\, Sentences_{it}} \tag{8}$$

The proportions $\sum_k I_{it}^k$ add up to 1 , which creates desirable properties for further analysis and cross-countries and time comparison for two main reasons. First, it normalizes the score for the increase in country topical matters over time and accounts for strong heterogeneity between topical issues. Second, it creates a relative measure of topic dominance in each press release i at a date t, which is useful for cross-country issues discussions and temporal comparisons.

For example, "Debt+Fiscal" (green bars) tends to have higher intensity values compared to "Debt Only" (gray) and "Fiscal Only" (red), suggesting that when these topics co-occur, they dominate the content more strongly. "Debt" and "Fiscal" topics only are more evenly spread across lower intensity bins, indicating they are less dominant when discussed in isolation.

Figure 4: Distribution of Fiscal and Debt Topics Coverage Intensity I_{it}^k



Notes: This figure illustrates the intensity distribution of fiscal and debt-related discussions within the IMF press release corpus, based on 12,673 sentences from 2020 to 2024 across 40 countries.

Next, we extract the sentiment associated with each topic using a domain-specific adaptation of FinBERT tailored to financial and economic texts. This off-the-shell model is trained on a corpus of financial and economic documents, enabling it to better capture the subtleties and specific language used in economic discussions. We extract the sentiment for every sentence. Finally, we aggregate up, the sentiment at the topics level as well as at the press release level. The aggregation at the topic level allows us to track and interpret the overall sentiment associated with specific topics over time and across countries. To formalize this, we define T_{it}^{k}, which represents the tone of topic k at time period t for country i, and apply a logit scaling approach (Lowe et al. (2011)) by taking the logarithmic balance of positive and negative sentences.

$$T_{it}^{k} = log(PositiveSentences_{it}^{k} + 0.5) - log(NegativeSentences_{it}^{k} + 0.5) \tag{9}$$

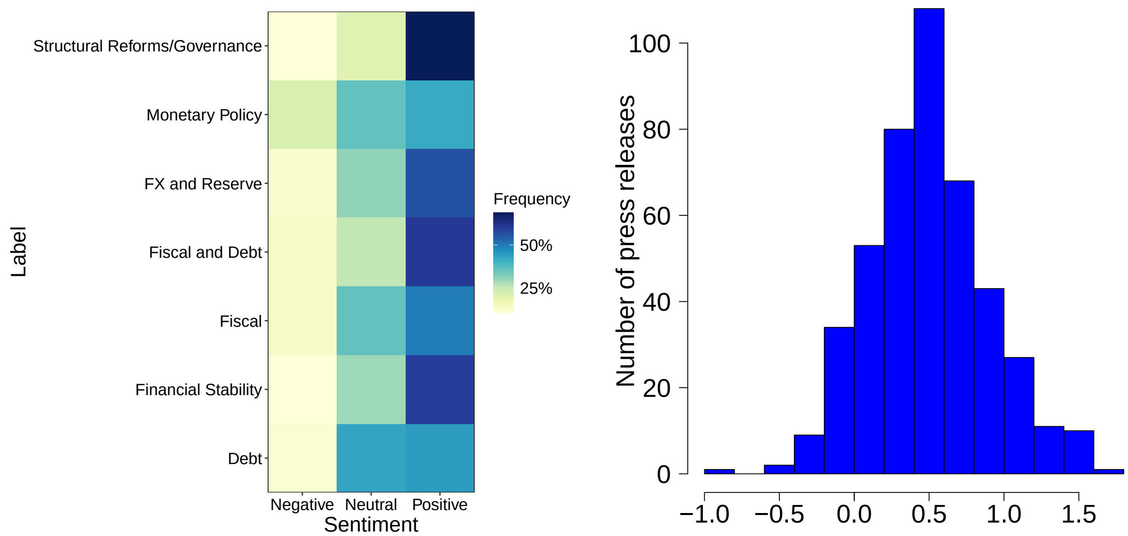
This transformation helps to stabilize the variance and manage cases with zero counts by adding a small constant (0.5). Moreover, The logit scale is unbounded, allowing sentiment scores to increase with the number of sentences. A persistently positive value for a topic might reflect favorable economic conditions or successful policy implementations. Conversely, a negative value may indicate emerging concerns, economic uncertainty, or policy challenges.

In Panel A of Figure 5, we provide a heatmap visualization showing the distribution of sentiment types—Negative, Neutral, and Positive—across our candidate topics. Most categories show a higher proportion of positive sentiment: Monetary Policy discussions tend to be more mixed or cautious, possibly reflecting the complexity and sensitivity of this area. The Debt and Fiscal categories show a more even distribution, indicating nuanced or balanced communication, while structural Reforms/Governance is communicated with the most optimism. Panel B shows the distribution of sentiment scores across our corpus of press releases. Sentiment scores range from -1.0 (very negative) to 1.5 (very positive), with most values clustering around 0.5. The highest bar in the histogram is centered near 0.5, indicating that most press releases have a moderately positive tone. There are fewer press releases with strongly negative or strongly positive sentiment scores. Overall, the data suggests that press releases tend to be written with a generally positive or neutral sentiment.

Figure 5: Sentiment Distribution across Topics and Press Releases

(a) Sentence Sentiment Frequency by Topic

(b) Distribution of Press Release Sentiment



Notes: Distribution of sentences related to Fiscal and Debt issues in the corpus of IMF press release. The sample comprises of 12673 sentences from 2020-2024 across 40 countries.

Table 4 provides examples communication to address economic challenges. Zambia's sentence is direct, labeling its debt as unsustainable, which reflects a negative tone. Egypt's message acknowledges fiscal challenges but avoids strong negative language, showing a more neutral or cautiously positive tone. Ghana's example highlights reform progress and economic recovery, using a clearly positive tone despite underlying complexities.

Table 4: Example Statement/Sentences and their Tones

Country	Topic	Sentence/Paragraph	Sentiment Framing	Tone
Zambia	Debt	PR No.24/476. Zambia's public debt is assessed as sustainable, but the country remains at high risk of overall and external debt distress based on a full post-restructuring macro-framework.	Acknowledged debt distress.	Negative
Egypt	Fiscal + Debt	PR No.24/500. The continued implementation of fiscal consolidation efforts will be necessary to preserve debt sustainability and reduce large interest costs and gross domestic financing requirements.	Acknowledged fiscal strain but avoids declaring debt unsustainable.	Positive
Ghana	Mixed-Topics	PR No.24/447. Ghana's policy and reform efforts under the IMF-supported program have continued to deliver encouraging results. Following acute economic and financial pressures in 2022, the Fund-supported program has provided a credible anchor for the government to adjust macroeconomic policies and launch comprehensive reforms to restore macroeconomic stability and debt sustainability, while laying the foundation for higher and more inclusive growth. These efforts are paying off, with growth recovering rapidly, inflation declining —although at a slower pace—and the fiscal and external positions further improving	Acknowledged past economic pressures but quickly shifts to highlight "encouraging results" and "recovery," without detailing ongoing risks. Emphasized future potential rather than current limitations.	Positive
Panama	Financial Stability	PR No. 20/110. The authorities should focus on macroprudential tools and further upgrading the regulatory toolkit.	.	Neutral

5 Bond Market Reactions to IMF Announcements

This section introduces our empirical strategy as well as the main findings of the analysis. We begin by introducing the local projections methodology used in the strategy. We then discuss potential sources of endogeneity and explain how these are mitigated. Additionally, we highlight our key results, which are supported by a series of robustness checks including alternative sentiment measures and placebo tests.

5.1 Empirical Strategy

For clarity, when a press release is published, we refer to it as policy shock z. We estimate how an exogenous policy news intervention corresponding to an IMF press release publication affects sovereign spreads building on the local projections methodology of Jordà (2005). Let y_t denote the daily change in sovereign spreads on day t representing the outcome variable and exogenous variables including z the shock variable and x_t a set of control variables, encompassing lags of the outcome variable. We characterize how an IMF communication affects the average sovereign spreads relative to a baseline where the IMF does not intervene.

$$\mathcal{R}_{s\rightarrow y}(h,\theta) \equiv E[y_{t+h}|z_t = z_0 + \theta; x_t] - E[y_{t+h}|z_t = z_0; x_t]; \quad with \quad h = 0, 1, ..., H \tag{10}$$

where θ represents the size of the news shocks. In addition, $\mathcal{R}_{s\rightarrow y}$ refers to the extent to which the IMF news intervention impacts the outcome variable y. The local projection of y_{t+h} can be described as followed:

$$y_{i,t+h} = \alpha_i + \gamma_t + \beta_h z_{i,t} + \sigma_h x_{i,t} + \epsilon_{i,t+h}; \quad with \quad h = 0, 1, ..., H \tag{11}$$

We use both country α and time γ fixed effects. This specification allows us to assess whether the informational content of IMF communications leads to differential responses on the cross-section of the sovereign spreads. In our specification, endogeneity may arise from three primary sources: omitted variable bias, measurement error, and reverse causality—all of which we address in our analysis.

First, to mitigate the risk of omitted variable bias, beyond the news shocks specified in Equation (11), we include a comprehensive set of control variables. These include measure of the global volatility Index (VIX) and the US Financial Index Conditions (FCI). We also control for macroeconomic variables that are country-specific including the consumer price index, consumer prices, general government gross debt, general government primary net lending borrowing, and GDP per capita, and include both country and time fixed effects. Additionally, we account for institutional quality by incorporating the six Worldwide Governance Indicators: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. This set of control variables is crucial in enabling us to disentangle the effect of IMF press releases from potentially confounding country- and global-level factors.

Second, we address potential measurement error by utilizing sovereign spreads that are calculated as the par-value weighted average across a country's sovereign bonds with more than one year of remaining maturity, and they reflect a relatively liquid subset of the sovereign bond market. This approach ensures a more accurate and representative measure of market perceptions. Furthermore, our primary sentiment indicator is a domain-specific metric designed for financial and economic contexts. It leverages state-of-the-art natural language processing (NLP) techniques to capture nuanced sentiment in IMF communications. To reinforce the robustness of our findings, we complement this measure with two alternative sentiment indicators, allowing us to validate the consistency of our results across different sentiment specifications.

Third, to address potential reverse causality, we lag all aforementioned yearly macroeconomic variables by one period. These variables typically evolve gradually, reducing the likelihood of abrupt structural changes in the economy. Similarly, we lag the Worldwide Governance Indicators by one year to mitigate contemporaneous feedback effects. Given the persistent nature of financial market volatility, we include up to seven lags of the daily VIX and the quarterly FCI to account for the influence of past volatility on current sovereign spreads. Additionally, to capture potential autocorrelation in sovereign spreads that is not explained by the control variables, we include up to seven lags of the change in spreads.

5.2 Main Findings

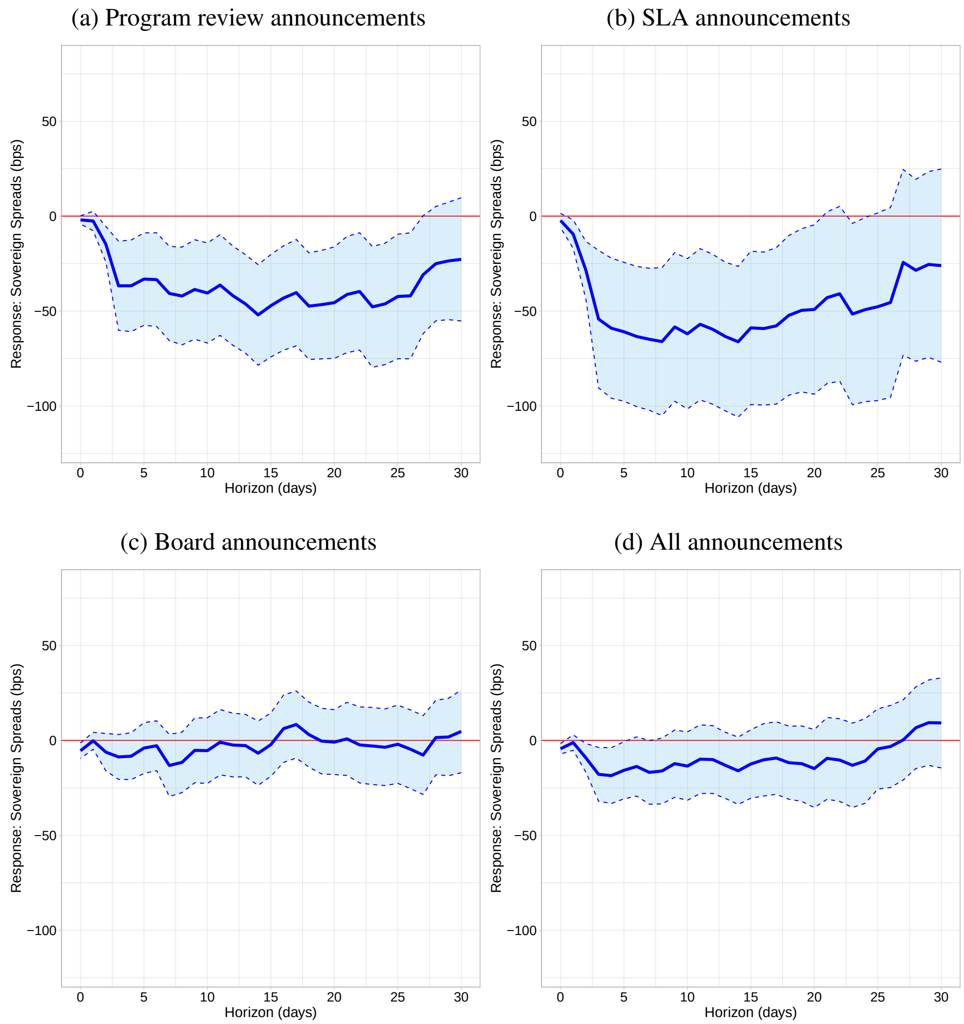
5.2.1 Are IMF Press Release Days Special days?

A natural starting point for our analysis is to examine whether sovereign bond spreads exhibit different behavior on IMF press release days compared to all other days. Figure 19 presents the impulse response functions derived from Equation (11). We estimate this specification separately for each horizon h = [0,30] and plot the estimated coefficients, $\{\beta_h\}_{h=1}^{30}$, which capture the cumulative change in sovereign spreads following an IMF press release. The blue solid line represents the estimated mean point-wise response of daily sovereign spreads to IMF announcements, and the shaded areas correspond to the point-wise 90 percent confidence intervals.

Our findings indicate a strong downward drift in sovereign spreads in the days following the IMF program review announcements, whether issued by mission chiefs or the board (down to 50 bps), and this response remains significant for nearly 30 days after the announcement (Panel A). Among the program review announcements, the sharpest decline in sovereign spreads occurs immediately after a staff-level agreement issued by mission chiefs (day 0 to ∼day 5), indicating that markets react swiftly and strongly to this specific IMF announcement. The response remains significantly negative (around ∼60 bps) throughout most of the 30-day period, suggesting that the credibility and reassurance provided by IMF involvement have a lasting calming effect on sovereign risk perceptions (Panel B). After approximately 25 days, the response begins trending upward, gradually reverting to zero, which implies that initial optimism wanes slightly over time or that other market factors begin to take over. In parallel, sovereign spreads do not narrow following press releases issued exclusively by the Board, indicating that since markets have already priced the information at the stage of the SLA, they do not reprice the information (Panel C). Finally, we consolidate the announcements by considering jointly program reviews, surveillance, and staff visits (Panel D). The global response of sovereign spreads is not significant, indicating that beyond surveillance, program review announcements have a substantial impact on shaping sovereign spreads in EMDEs (35 bps on average).[6] Indeed, program review-related press releases constitute a particular form of announcement from the Fund that markets scrutinize, and this announcement helps ease the pressure on sovereign spreads.

[6]This result echoes the existing literature, which documents that IMF programs matter, as evidenced by, for instance, greater access to international capital markets (Lisi (2022), Kogan et al. (2024)).

Figure 6: The effects of IMF Press Release Days on Sovereign Spreads



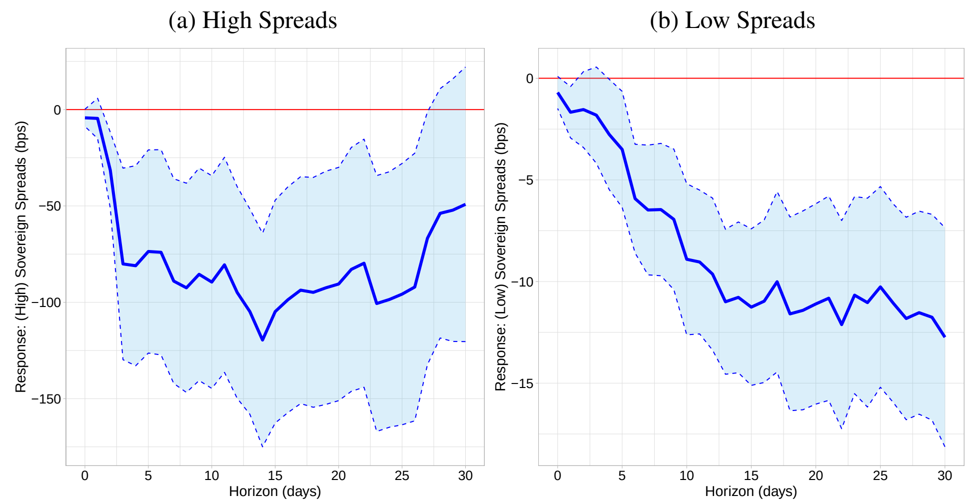
Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI and the yearly macroeconomic variable (growth, inflation and debt) as additional controls. The shock corresponds to the press release publication is equal to 1 where there is an announcement and zero otherwise. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

Our results on the announcement drift of the IMF communication relate to the work of Lucca and Moench (2015), which documents a 49-bp increase in the S&P500 in the 24 hours before scheduled FOMC announcements, and are consistent with existing studies related to the effect of policymakers' announcements on sovereign spreads. For example, David et al. (2022) find a decline in sovereign spreads following news of fiscal consolidation approved by Congress or Parliament, in periods of high sovereign spreads, or countries with an IMF program. This decline reaches around 15 bps within a 30-day window after the austerity announcement by the Congress.

To further validate our findings, Section 6 presents a series of robustness checks through a placebo test, designed to assess whether the observed effects of IMF press releases on sovereign spreads are effectively driven by the timing of the announcements, rather than by external factors or coincidental trends. In this test, we simulate alternative scenarios by assuming that each press release was published one day or one week earlier or later than its actual release date. The results reveal distinct market reactions in these hypothetical cases, reinforcing the conclusion that the timing of IMF announcements plays a critical role in shaping investor sentiment, as reflected in the decline of sovereign spreads.

We then explore the non-linear effects of IMF press releases in the case of an IMF policy intervention news, acknowledging potential heterogeneity across countries and over time. To do so, we differentiate between market conditions by classifying sovereign spreads into high and low categories based on the median value of the spread distribution.

Figure 7: Non-linearities Effect of Spreads



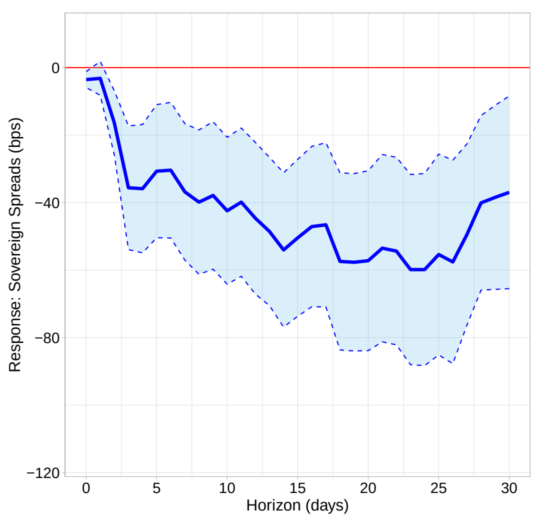
Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock corresponds to the press release publication is equal to 1 where there is an announcement and zero otherwise. We define two sub-samples: (1) High spreads are above the median spreads' distribution, (2) whereas low spreads are below the median distribution. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

As illustrated in Figure 7 above, countries with higher sovereign spreads experience a more pronounced decline—approximately 100 basis points—following an IMF press release (Panel A). In contrast, countries with lower spreads exhibit a more moderate response, with spreads decreasing by around 12 basis points (Panel B). These findings suggest that the effectiveness of IMF announcements is amplified in more vulnerable market conditions. In addition, the heterogeneity in the sustainability of the effects could reflect differences in market confidence, and underlying fundamentals.[7]

5.2.2 Sentiment and Sovereign Spreads

Since our analysis has previously established that the timing of the IMF announcements matters, we now examine the informational content of these press releases to determine if they convey meaningful information to markets. To achieve this objective, we first investigate the effect of sentiment, derived from the textual analysis of IMF press releases, to assess whether the tone conveyed influences sovereign bond spreads. The sentiment shock is defined as the level of tone associated with each press release. As shown in Figure 8, sovereign spreads tend to decline following press releases with a positive tone. Specifically, a positive sentiment is associated with a short-term reduction in sovereign spreads of approximately 55 bps. This effect is persistent, although it moderates slightly over the 30-day horizon. These findings suggest that the tone of IMF announcements plays a significant role in shaping investor perceptions and influencing market outcomes.

Figure 8: The effect Press Releases' tone on sovereign spreads



Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock is defined as the interaction term between the announcement dummy and the sentiment index in level for a given press release. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

Although analyzing sentiment at the press release level reveals general trends, it does

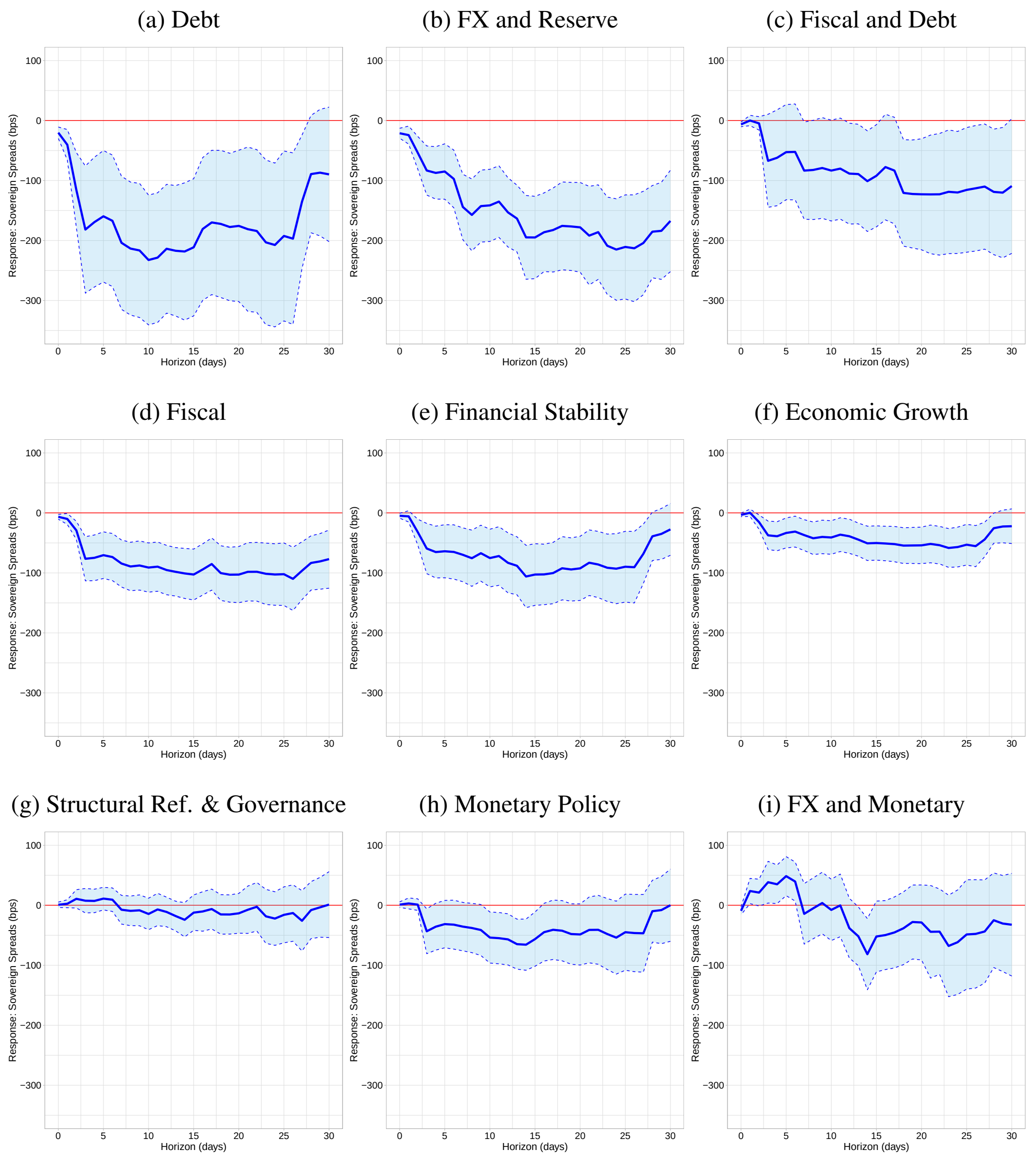
[7] We have also conducted an additional robustness test to distinguish between countries that experienced debt restructuring during the analyzed period and those that did not, to account for the fact that high volatility and the fat-tailed distribution of sovereign spreads can induce heterogeneous impacts when the IMF press release is issued. Such analysis does not change qualitatively our results.

not identify which specific topics within the releases are driving the observed decline in sovereign spreads. To address this, we extend our analysis by examining the impact of sentiment at the topic level. Figure 9 presents the response of sovereign spreads to sentiment associated with different topics.

Our findings suggest that there is a heterogeneity in market sensitivity due to the non-linear aggregation of sentiment. The overall tone of a press release is likely a weighted average of topic-specific tones. If only a small portion of the release is devoted to debt—and the rest is neutral or mixed—the aggregate sentiment signal is weaker, leading to a smaller average effect. The results reveal heterogeneous effects across topics. The most pronounced decline—approximately 200 bps—occurs when the press release discusses “Debt” in a positive tone. This is followed by topics such as “FX and Reserve”, “Fiscal and Debt”, “Fiscal”, and “Financial Stability”, each associated with a decline of around 100 basis points.[8] Other topics, including “Economic Growth”, “Structural Reforms and Governance”, and “Monetary Policy”, also lead to statistically significant, though more moderate, reductions in sovereign spreads—typically less than 100 bps. These findings underscore the importance of topic-specific sentiment in influencing market reactions to IMF announcements. The results suggest that investors are more sensitive to topics that directly affect a country’s repayment capacity, such as debt sustainability, fiscal policy, and foreign exchange reserves. These topics carry higher informational value, are more closely tied to default risk, and trigger stronger belief updating. Taken together, our results highlight that topic-specific sentiment plays a pivotal role in shaping investors’ beliefs about key economic fundamentals.

[8] David et al. (2022) find that the impact of fiscal announcements on spreads is larger in economies with high-perceived sovereign risk. In periods of high perceived sovereign risk, defined as those at or above the 75th percentile of the empirical distribution, spreads decline significantly after the announcements by around 100 bps within a 12-month window. However, the announcement is also typically followed by a large and protracted output loss (of around 4 percent).

Figure 9: The effect of Topic Tone on Sovereign Spreads



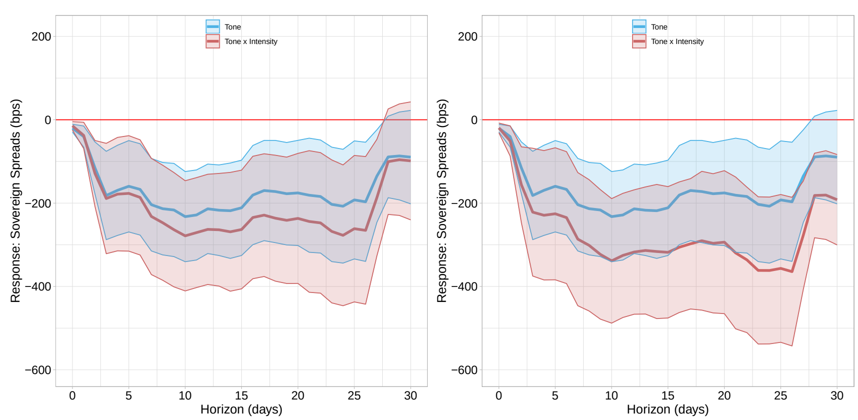
Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock is defined as the interaction term between the announcement dummy and the sentiment index in level for a given topic. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

5.2.3 Does the Intensity of Coverage Matter for Sovereign Spreads?

So far, our findings underscore the importance of sentiment in IMF announcements as a pivotal dimension of textual analysis, with a key highlight: sentiment related to the debt topic shows the most substantial impact, leading to the most considerable reductions in sovereign spreads. Building on this, we now investigate whether, beyond the tone itself, the intensity of coverage—that is, how prominently or frequently the debt topic is discussed in a press release—also plays a meaningful role in explaining the cross-sectional variation in sovereign spreads. To capture the joint influence of tone and topical salience, we construct an interaction term between the sentiment score and the intensity of topic coverage. Specifically, we define a binary indicator (dummy variable) for high-intensity coverage of the debt topic, which takes the value of 1 when the frequency or prominence of debt-related content exceeds a defined threshold, and zero otherwise. This allows us to test whether the effect of sentiment on sovereign spreads is amplified when the debt topic is not only discussed positively but also emphasized more heavily in the communication. High intensity of coverage is defined as an intensity greater than the 75th and 85th percentiles of the intensity's distribution for each topic. Figure 10 presents the results of our analysis, incorporating the interaction between sentiment and the intensity of topic coverage.

Figure 10: Intensity of Debt Topic Coverage and Sovereign spreads

(a) Debt (P75) (b) Debt (P85)



Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The intensity's shock is equal to 1 when the topic's coverage is greater than respectively the 75th (left-) and 85th (right-hand side chart) percentile of the intensity's distribution for the 'Debt' topic. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

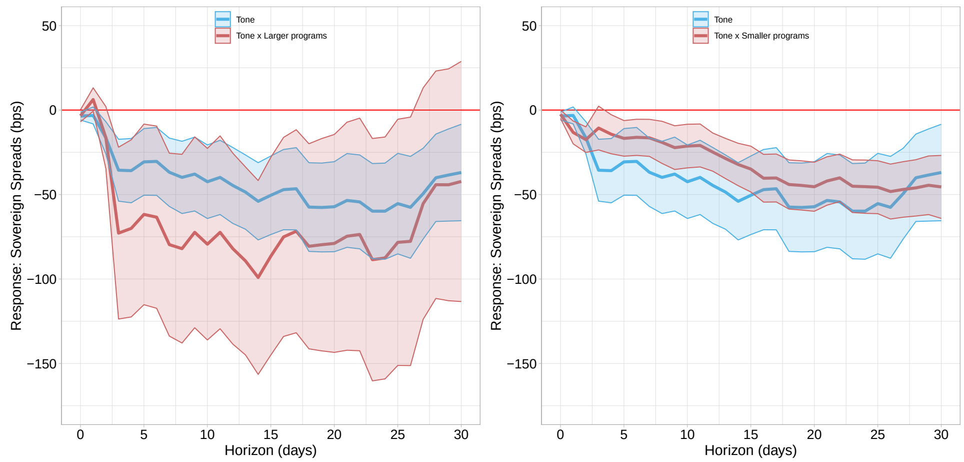
The results show that the joint effect results in a further decline in sovereign spreads. Specifically, when the sentiment is positive and the debt topic is discussed with high intensity, the reduction in spreads is more pronounced compared to scenarios where only sentiment is considered. A higher intensity of coverage from other topics, such as "Economic Growth" or "Financial Stability", does not induce such an additional decline in sovereign spreads. This finding illustrates that the coverage of "Debt" in the IMF press releases is a topic that markets particularly scrutinize and internalize.

5.2.4 Amount of Programs' Size and Sovereign Spreads

We finally explore the IMF program size to assess if it plays a multiplicative role in shaping sovereign spreads following the publication of an IMF press release. In particular, we examine whether countries with larger programs benefit from a larger decrease in their sovereign spreads. Programs' size is defined as the access in percent of GDP. We compute the median access and create a dummy variable for larger programs such that larger programs are above the median, and we attribute a dummy of 1, zero otherwise. Conversely, we define a dummy variable for smaller programs, which equals 1 when programs are below the median distribution and zero otherwise. On the one hand, Panel A of Figure 11 shows that larger programs induce a higher decline in sovereign spreads, an additional decline that reaches down to 30 bps. On the other hand, Panel B indicates that smaller programs do not exhibit this effect. In a related study, Chahine et al. (2025) examine the impact of IMF programs' size on borrowing costs and find that the approval of the program leads to a 72-bp decrease in borrowing costs and that when program size rises by one percent of GDP, borrowing costs decline by 23 bps.

Figure 11: The effect of IMF programs' size on sovereign spreads

(a) Larger programs (b) Smaller programs



Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX, and the quarterly FCI, as well as 1 lag for the six-yearly WBGI, as additional controls. The shock is defined as the triple interaction term between the announcement dummy, the program size dummy, and the sentiment index in level for a given press release. Larger programs are defined as those with a distribution higher than the median access in percent of GDP. Smaller programs are smaller than the median distribution. We define a categorial variable of each of these two states of nature: a large program's dummy equal to 1 for larger access in percent of GDP and zero otherwise; similarly, a small program's dummy equal to 1 for smaller access in percent of GDP and zero otherwise. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

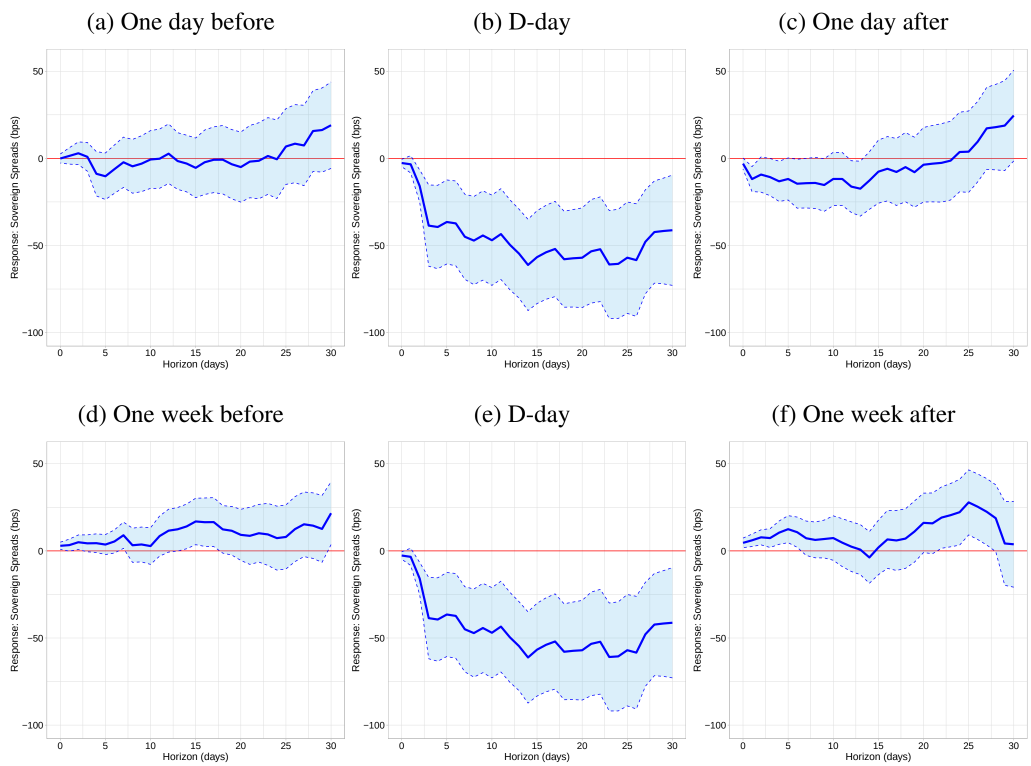
6 Additional Robustness Tests

To assess the sensitivity of our findings to additional assumptions and measurements, we perform a series of robustness analyses.

6.1 Placebo Test

We conduct a placebo test to assess the robustness of our findings by assuming that the IMF press release is published at different points in time relative to its actual release. Specifically, we consider scenarios where the publication occurs either one day before or after the actual date, as well as one week earlier or later. By analyzing the impact of these hypothetical shifts, we can determine whether the timing of the press release plays a critical role in influencing sovereign spreads. Figure 12 illustrates the outcomes of this analysis. The findings from the placebo test reveal different market reactions, primarily muted effects on sovereign spreads when the release day is shifted. This result reinforces the credibility of our findings, suggesting that the timing of the press release is indeed relevant. Markets react promptly to IMF announcements, and the post-drift announcement starts indeed on the day of the IMF press release (D-day), illustrating the timely impact of the IMF in shaping investor confidence and sovereign spreads.

Figure 12: The effect of IMF press releases on sovereign spreads for Placebo days

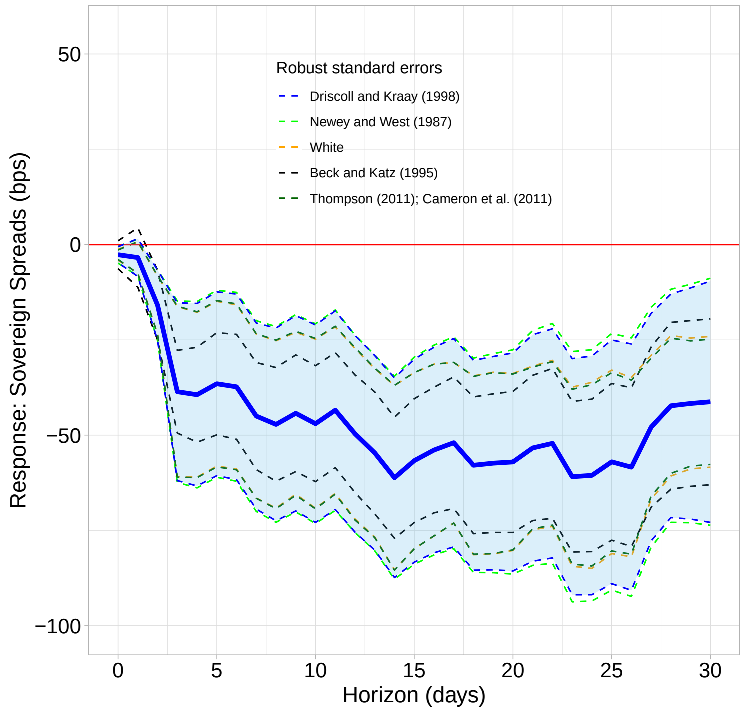


Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock corresponds to the press release publication is equal to 1 where there is an announcement and zero otherwise. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

6.2 Robust Standard Errors

We re-run specification 11 while applying various robust standard errors. Figure 13 displays the confidence bands for these different robust standard errors. The confidence intervals vary across standard error types. Beck and Katz (1995), which provides unconditional robust covariance matrix estimators, exhibits the narrower confidence bands. On the other hand, Newey and West (1987), which returns nonparametric robust covariance matrix estimators that account for serial correlation, has the largest and most robust confidence bands. These latter intervals are very close to the Driscoll and Kraay (1998) interval bands used in the main findings section, which provide nonparametric covariance matrix estimators robust to spatial and temporal dependence. Overall, we find that the results remain quantitatively consistent regardless of the definition of standard errors considered.

Figure 13: The effect of IMF press releases on sovereign spreads with various robust standard errors

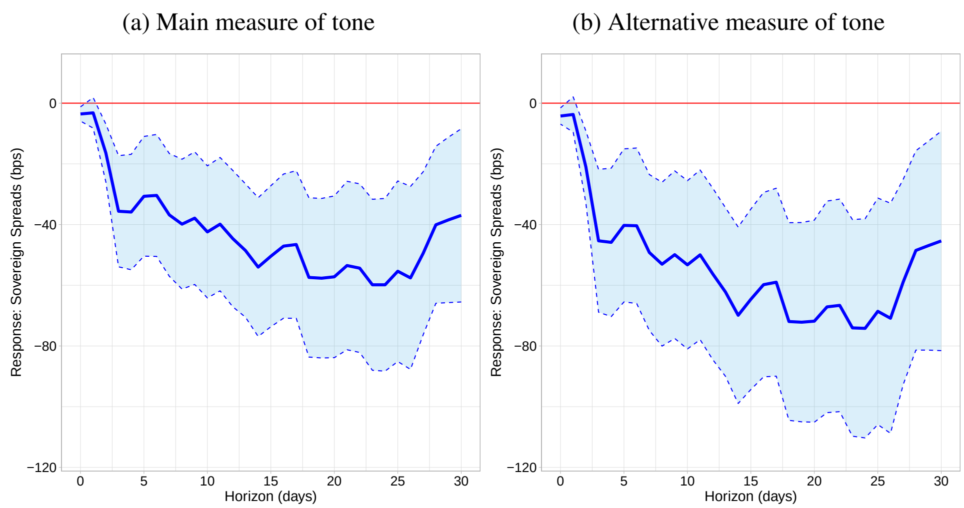


Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock is defined as the interaction term between the announcement dummy and the sentiment index in level for a given press release. Dashed lines represent point-wise 68 percent significance bands. The sample ranges from 2020-03-17 to 2024-11-19.

6.3 Alternative Measure of Tone

To demonstrate the robustness of our results to alternative measures of tone, we conduct a similar exercise as in Section 5. We calculate the measure of tone differently, using the difference in counts between positive and negative sentences, normalized by the total number of positive and negative sentences in the press release. Figure 14 shows that the impact of sentiment on sovereign spreads remains statistically significant, regardless of the measure employed, thus further strengthening our main findings.

Figure 14: Alternative measure of sentiment and Sovereign Spreads



Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI as additional controls. The shock is defined as the interaction term between the announcement dummy and the sentiment index in level for a given press release. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

7 Conclusion

Communication is a key tool for sharing economic policy. This study provides novel empirical evidence on the impact of IMF press releases-specifically those from Mission Chiefs and the Executive Board—on sovereign bond spreads. To characterize ideas, we provide a theoretical framework grounded in ambiguity aversion, where we demonstrate that tangible information on macroeconomic parameters can contribute to reducing uncertainty among investors. By narrowing the range of plausible macroeconomic scenarios, IMF announcements lower the ambiguity premium embedded in sovereign bond pricing. We then leverage LLMs to extract both the thematic intensity and sentiment of IMF press releases.

Our findings reveal that these announcements are not only rich in content and interconnected in scope but also carry significant signaling power for financial markets. In particular, positive sentiment in IMF press releases, particularly when focused on critical macroeconomic topics such as debt sustainability, fiscal policy, and foreign exchange reserves, is associated with a sizable reduction in sovereign spreads. This effect is especially pronounced in countries with higher initial spreads, suggesting that markets are particularly responsive to reassuring signals in less robust macroeconomic contexts. We also document that countries with larger IMF program size experience an additional decline in their sovereign spreads following the publication of the press release. Our findings underscore the macro-criticality of expectations' management in policy development, as sovereign spreads reflect both government actions and their perceived credibility.[9]

[9] See Making Public Debt Public—Ongoing Initiatives and Reform Options

References

Aruoba, S. B. and T. Drechsel (2024). Identifying Monetary Policy Shocks: A Natural Language Approach. Technical report, National Bureau of Economic Research.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Bansal, R. and I. Shaliastovich (2013). A Long-run Risks Explanation of Predictability Puzzles in Bond and Currency Markets. *The Review of Financial Studies 26*(1), 1–33.

Beck, N. and J. N. Katz (1995). What to do (and not to do) with Time-series Cross-section Data. *American Political Science Review 89*(3), 634–647.

Caldara, D. and M. Iacoviello (2022). Measuring Geopolitical Risk. *American Economic Review 112*(4), 1194–1225.

Chahine, S., U. Panizza, and G. Suedekum (2025). IMF Programs and Borrowing Costs Does Size Matter? *European Economic Review*, 105070.

Chen, Y., B. T. Kelly, and D. Xiu (2024). Expected Returns and Large Language Models. *Available at SSRN 4416687*.

Curti, F. and S. Kazinnik (2023). Let's face it: Quantifying the Impact of Nonverbal Communication in FOMC Press Conferences. *Journal of Monetary Economics 139*, 110–126.

David, A. C., J. Guajardo, and J. F. Yepez (2022). The Rewards of Fiscal Consolidations: Sovereign Spreads and Confidence Effects. *Journal of International Money and Finance 123*, 102602.

Dell, M. (2025). Deep Learning for Economists. *Journal of Economic Literature 63*(1), 5–58.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186.

Dim, C., F. Sangiorgi, and G. Vilkov (2023). Media Narratives and Price Informativeness. *SSRN Electronic Journal*.

Driscoll, J. C. and A. C. Kraay (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics 80*(4), 549–560.

Epstein, L. G. and M. Schneider (2008). Ambiguity, Information Quality, and Asset Pricing. *The Journal of Finance 63*(1), 197–228.

Fratzscher, M. and J. Reynaud (2011). IMF Surveillance and Financial Markets—A Political Economy Analysis. *European Journal of Political Economy 27*(3), 405–422.

Gao, T., X. Yao, and D. Chen (2021). Simcse: Simple Contrastive Learning of Sentence Embeddings. *arXiv preprint arXiv:2104.08821*.

Goetzmann, W. N., D. Kim, and R. J. Shiller (2022). Crash Narratives. Technical report, National Bureau of Economic Research.

Gorodnichenko, Y., T. Pham, and O. Talavera (2023). The Voice of Monetary Policy. *American Economic Review 113*(2), 548–584.

Gorodnichenko, Y. and M. Weber (2016). Are Sticky Prices Costly? Evidence from the Stock Market. *American Economic Review 106*(01), 165–199.

Handlan, A. (2020). Text Shocks and Monetary Surprises: Text Analysis of FOMC Statements with Machine Learning. *Published Manuscript*.

Hansen, S., M. McMahon, and M. Tong (2019). The Long-run Information Effect of Central Bank Communication. *Journal of Monetary Economics 108*, 185–202.

Hanson, S. G. and J. C. Stein (2015). Monetary Policy and Long-term Real Rates. *Journal of Financial Economics 115*(3), 429–448.

Herbert, S., K. Istrefi, and B. Sagna (2024). Coordinating the Message: Media Coverage of Fed News and Market Reactions. Working paper 983, Banque de France.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review 95*(1), 161–182.

Ke, Z. T., B. T. Kelly, and D. Xiu (2019). Predicting Returns with Text Data. Technical report, National Bureau of Economic Research.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*, Volume 31. Houghton Mifflin.

Kogan, J., R. Kazandjian, S. Luo, M. Mbohou, and H. Miao (2024). The Role of IMF Arrangements in Restoring Access to International Capital Markets.

Leek, L. and S. Bischl (2025). How Central Bank Independence Shapes Monetary Policy Communication: A Large Language Model application. *European Journal of Political Economy 87*, 102668.

Leombroni, M., A. Vedolin, G. Venter, and P. Whelan (2021). Central Bank Communication and the Yield Curve. *Journal of Financial Economics 141*(3), 860–880.

Lisi, G. (2022). IMF's Precautionary Lending Instruments: Have They Worked?

Lowe, W., K. Benoit, S. Mikhaylov, and M. Laver (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative studies quarterly 36*(1), 123–155.

Lucca, D. O. and E. Moench (2015). The Pre-FOMC Announcement Drift. *The Journal of finance 70*(1), 329–371.

Minaee, S., T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, and J. Gao (2024). Large Language Models: A survey, 2024. *arXiv preprint arXiv:2402.06196*.

Narain, N. and K. Sangani (2023). The Market Impact of Fed Communications: The Role of the Press Conference. *Available at SSRN 4354333*.

Newey, W. K. and K. D. West (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, 777–787.

Piazzesi, M., M. Schneider, P. Benigno, and J. Y. Campbell (2006). Equilibrium Yield Curves [with comments and discussion]. *NBER macroeconomics Annual 21*, 389–472.

Schmeling, M. and C. Wagner (2025). Does Central Bank Tone Move Asset Prices? *Journal of Financial and Quantitative Analysis 60*(1), 36–67.

Sharpe, S. A., N. R. Sinha, and C. A. Hollrah (2023). The Power of Narrative Sentiment in Economic Forecasts. *International Journal of Forecasting 39*(3), 1097–1121.

Soo, C. K. (2018). Quantifying Sentiment with News Media across Local Housing Markets. *The Review of Financial Studies 31*(10), 3689–3719.

Tunstall, L., N. Reimers, U. E. S. Jo, L. Bates, D. Korat, M. Wasserblat, and O. Pereg (2022). Efficient Few-shot Learning without Prompts. *arXiv preprint arXiv:2209.11055*.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). Attention Is All You Need. *Advances in neural information processing systems 30*.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. (2022). Chain-of-thought Prompting Elicits Reasoning in Large Language Models. *Advances in neural information processing systems 35*, 24824–24837.

Zhao, G. (2017). Confidence, bond risks, and equity returns. *Journal of Financial Economics 126*(3), 668–688.

Zhao, G. (2020). Ambiguity, nominal bond yields, and real bond yields. *American Economic Review: Insights 2*(2), 177–192.

Appendices

Appendix A Additional Tables and Figures

Table 5: List of Countries

Country	Region	Total
Senegal, Rwanda, Ivory Coast, Mozambique, Ethiopia, Kenya, Ghana, Benin, Zambia, Gabon, Angola, Cameroon	Sub-Saharan Africa	12
Panama, Honduras, Colombia, Jamaica, Peru, Costa Rica, Argentina, Barbados, El Salvador, Ecuador, Mexico, Chile, Paraguay, Suriname	South America	14
Ukraine, North Macedonia, Serbia	Europe	3
Egypt, Armenia, Tunisia, Pakistan, Jordan, Morocco, Lebanon, Tajikistan, Georgia	Middle East and North Africa	9
Papua New Guinea, Sri Lanka	Asia Pacific	2
TOTAL		40

Notes: This table report the list of countries in the sample.

Table 6: Descriptive Statistics

	Freq.	***N***	***Mean***	***Median***	***SD***	***Min***	***Max***	***Source***
Panel A. Textual data								**IMF**
Number of documents		447						
Total sentence count	Daily	12673	28	28	11	4	135	
Measure of tone 1	Daily	447	0.51	0.51	0.39	-0.95	1.74	
Measure of tone 2	Daily	447	0.29	0.30	0.20	-0.37	0.76	
Measure of tone 3	Daily	447	0.50	0.56	0.34	-1.00	1.00	
Panel B. Sovereign Spreads								**Bloomberg and IMF**
SPREAD bps	Daily	48296	1116.79	575.46	1775.10	41.00	15443.32	
Panel C. Macroeconomic Variables								**World Bank and IMF**
GDP (USD)	Annual	200	1.35E+11	4.85E+10	2.56E+11	2.91E+09	1.79E+12	
VIX	Daily	1277	21.36	19.61	8.24	11.86	82.69	
FCI (US)	Quarterly	20	-0.33	-0.56	0.65	-0.91	1.96	
Government Effectiveness	Annual	200	-0.30	-0.30	0.48	-1.58	0.79	
Political Stability and Absence of Violence Terrorism	Annual	200	-0.42	-0.36	0.67	-2.18	1.19	
Control of Corruption	Annual	200	-0.40	-0.47	0.59	-1.43	1.34	
Rule of Law	Annual	200	-0.40	-0.42	0.46	-1.28	0.93	
Regulatory Quality	Annual	200	-0.26	-0.31	0.52	-1.20	1.05	
Voice and Accountability	Annual	200	-0.18	-0.08	0.65	-1.78	1.16	
CPI	Annual	200	13.75	5.27	30.40	-1.55	229.82	

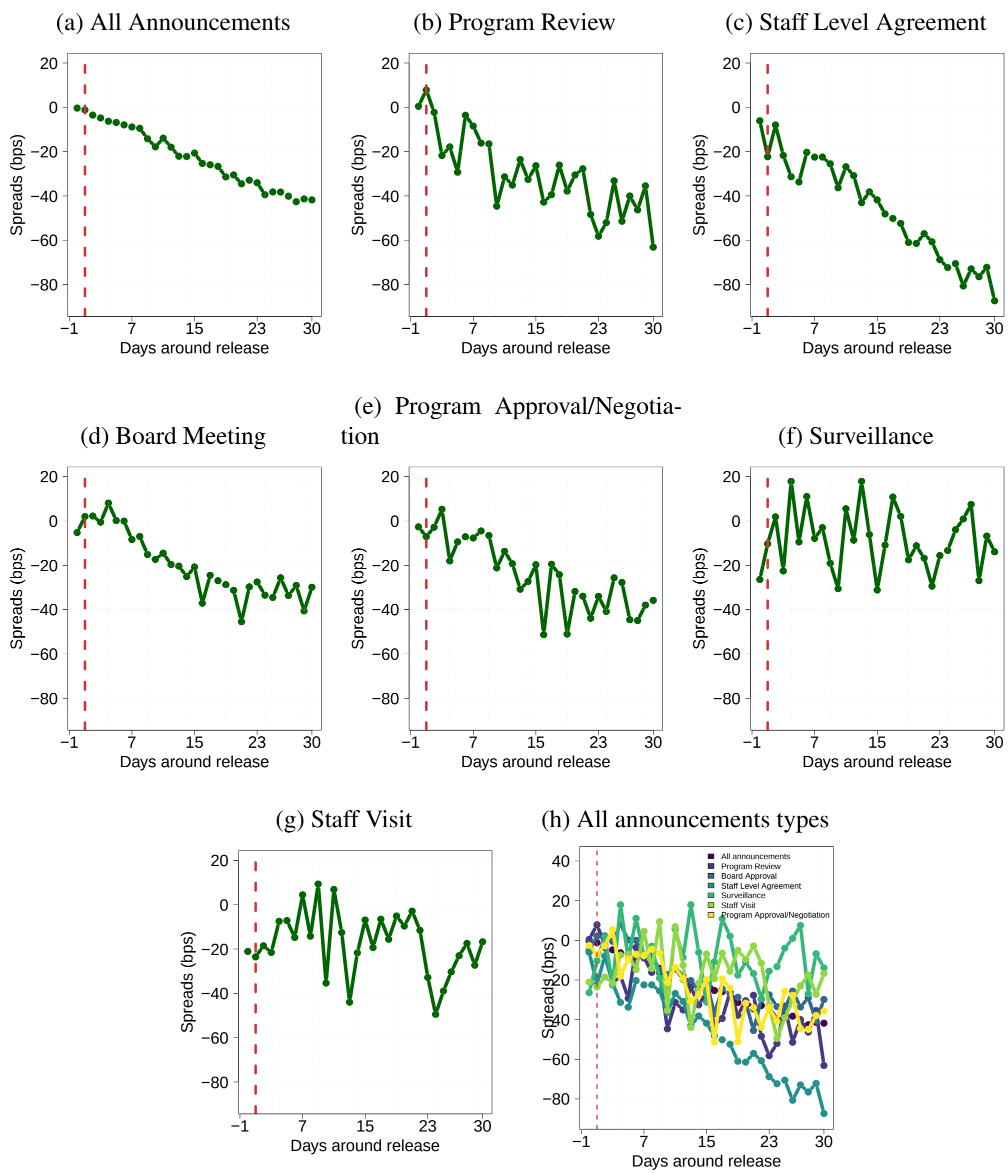
Notes: This table report the summary statistics of our sample including, the data frequency, some basics statistics and the sources of the data.

Table 7: Dominant Topics

Topic	Absolute Count	Share of Sentences
FX and Monetary	331	2.6%
FX and Reserve	345	2.7%
Debt	424	3.3%
Fiscal and Debt	526	4.2%
Climate	620	4.9%
Financial Stability	943	7.4%
Economic Growth	1105	8.7%
Structural Reforms/Governance	1269	10.0%
Monetary Policy	1301	10.3%
Fiscal	1737	13.7%
Others	4072	32.1%

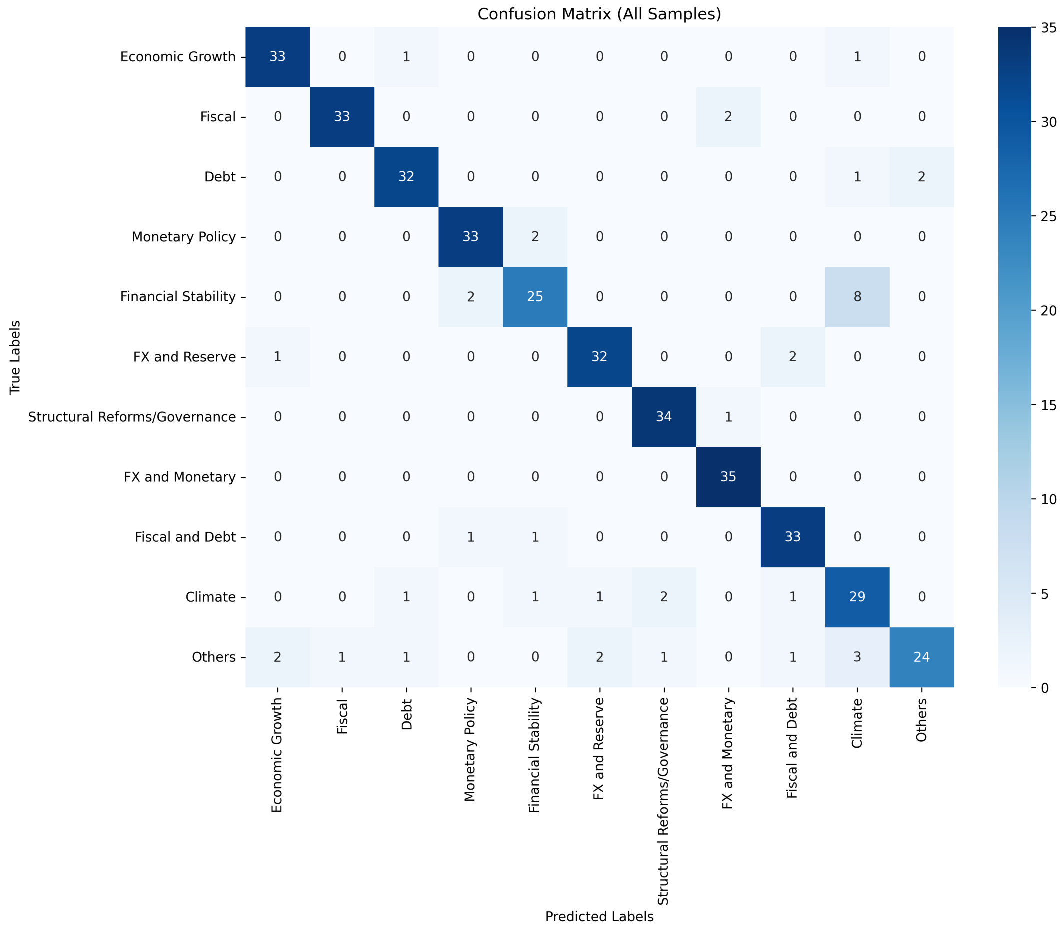
Notes: The table contains the share of topic sentence in our corpus. The sample comprises of 12673 sentences from 2020-2024 across 40 countries.

Figure 15: Cumulative Sovereign Spreads Changes around IMF press releases



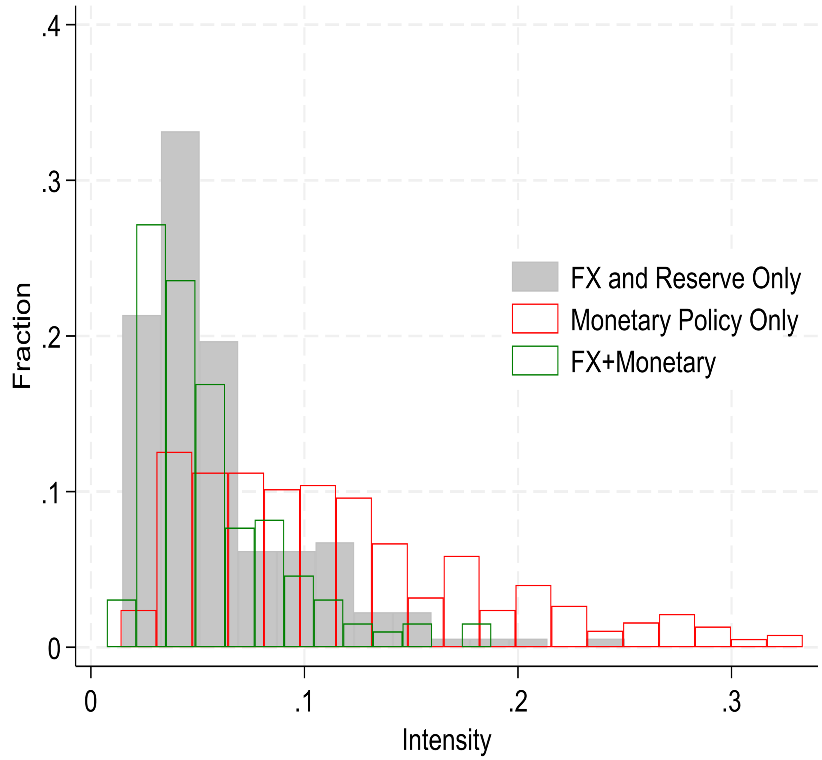
Notes: This figure shows the average sovereign spreads on a 30-day window after the IMF communication. The sample ranges from 2020-03-17 to 2024-11-19 and covers 40 countries.

Figure 16: Confusion Matrix



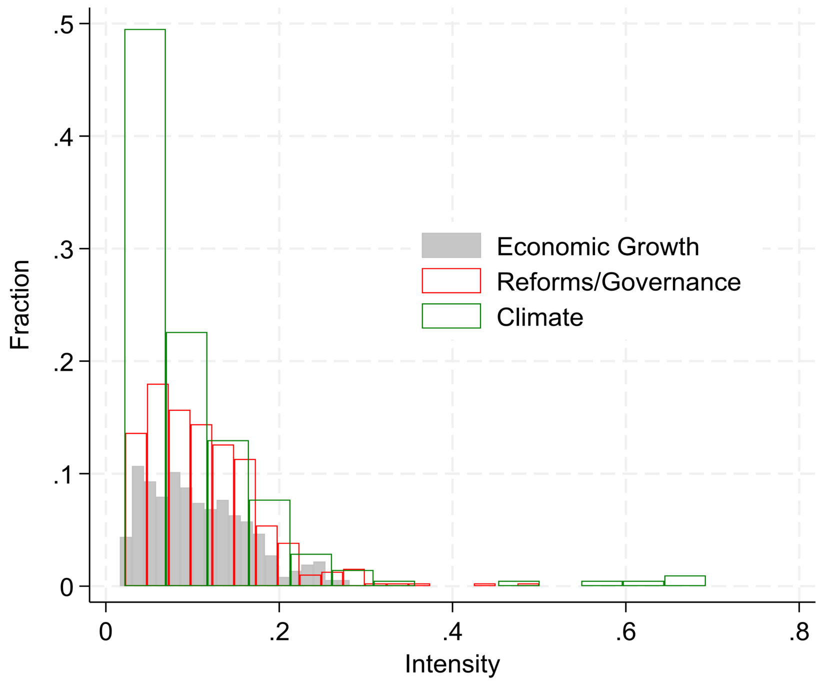
Notes: The confusion matrices plot the distribution of predicted labels by the "true" label from the validation sample which consists of 35 sentences in each class.

Figure 17: Distribution Monetary Policy, FX and Reserve Topics Intensity (I_{it}^{k})



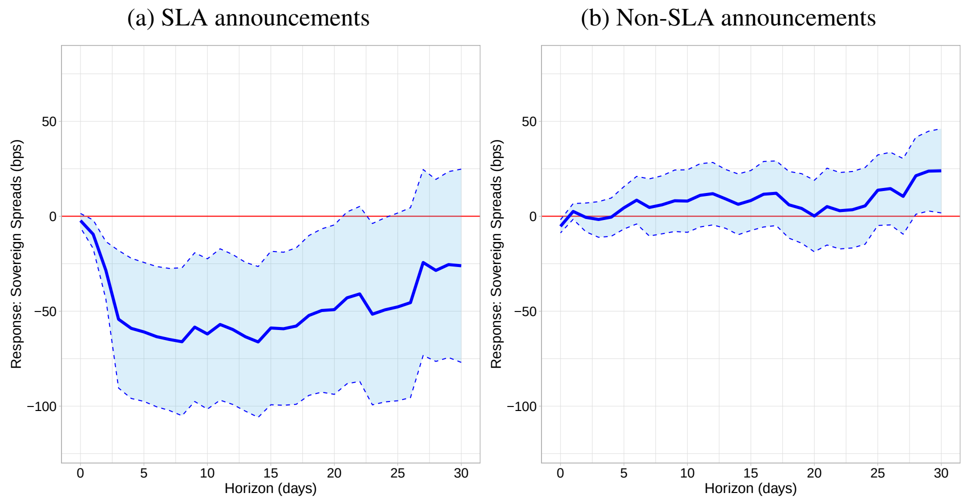
Notes: This Figure illustrates the intensity distribution of FX, Reserve and Monetary-related discussions within the IMF press release corpus, based on 12,673 sentences from 2020 to 2024 across 40 countries.

Figure 18: Distribution Economic Growth, Reforms/Governance and Climate Topics Intensity (I_{it}^{k})



Notes: This Figure illustrates the intensity distribution of Economic growth, Reforms and Governance and Climate-related discussions within the IMF press release corpus, based on 12,673 sentences from 2020 to 2024 across 40 countries.

Figure 19: SLA vs. non-SLA announcements



Notes: The outcome variable is the cumulative daily change of the sovereign spreads expressed in basis points. We use 7 lags of the daily changes in sovereign spreads, the daily VIX and the quarterly FCI and 1 lag for the six yearly WBGI and the yearly macroeconomic variable (growth, inflation and debt) as additional controls. The shock corresponds to the press release publication is equal to 1 where there is an announcement and zero otherwise. Dashed lines represent point-wise 68 percent significance bands based on Driscoll and Kraay (1998). The sample ranges from 2020-03-17 to 2024-11-19.

Appendix B Sentence Transformer Fine Tuning (SetFit) Training Steps

Step 1: Data Preparation

Choose a pre-trained Sentence Transformer model and prepare labelled data.

Step 2: Sentence Embedding and Attention Scores Matrix

Each sentence is first tokenized using the tokenizer associated with the transformer model. This converts the sentence into a sequence of token IDs. The tokenized input is passed through a transformer model, which outputs contextualized embeddings for each token:

$$H = [h_1, h_2, ..., h_N], \qquad h_i \in \mathbb{R}^d \tag{12}$$

The contextualized embeddings is achieve through a self-attention mechanism (Vaswani et al. (2017)) where each token attends to every other token in the sentence. The self-attention mechanism computes attention scores between all pairs of tokens, allowing each token to incorporate information from the entire sentence. This is crucial for capturing long-range dependencies, word sense disambiguation, and syntactic and semantic relationships. Each token is projected into three vectors: Query *(Q)*, Key *(K)* and Value *(V)*. These are computed as follows:

$$Q = XW^Q, \qquad K = XW^K \qquad and \qquad V = XW^V \tag{13}$$

where X is the input matrix (token embeddings), and W^Q,W^K,W^V are learned weight matrices. The attention score between tokens is computed using the dot product of queries and keys, scaled by the square root of the dimension d_k :

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{14}$$

This allows each token to incorporate contextual information from the entire sentence, enabling the model to understand meaning in context (e.g., resolving ambiguity, capturing dependencies). To convert these into a single sentence j embedding using pooling layer (e.g mean pooling).

$$e_j = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{15}$$

Step 3: Training Loop

The training loop is the iterative process where the model learns from data. It involves feeding batches of sentence pairs or triplets into the model, computing embeddings, calculating loss, and updating the model weights.

1. **Batching and few-short learning:** Use a DataLoader to create batches of sentence pairs for few-shot learning. To mitigate the problem of a few annotated samples on a small number of text pairs, a contrastive learning approach is used. The basic contrastive learning framework consists of selecting a data sample, called anchor a data point belonging to the same distribution as the anchor, called the positive sample (e_j^+), and another data point belonging to a different distribution called the negative sample (e_j^-) The contrastive loss function is designed to minimize the distance between the anchor and positive samples and maximize the distance between the anchor and negative samples (Gao et al. (2021)).

$$\mathcal{L}_{ij} = -log\frac{exp(cos(e_j^-, e_j^+)/\tau)}{\sum_{i=1}^{N} exp(cos(e_j^-, e_j^+)/\tau)} \tag{16}$$

2. **Forward Pass:** For each batch implement steps 1, 2 and 3 to compute embedding and loss function.

3. **Backward Pass and Optimization:** After computing the loss, the model needs to understand how each parameter contributes to that loss. This is done using backpropagation, an algorithm that applies the chain rule of calculus to compute the gradient of the loss with respect to each model parameter. These gradients tell the model how to change each weight to reduce the loss. Once gradients are computed, the optimizer (e.g., Adam) uses them to update the model's weights.

$$\theta \leftarrow \theta - \eta.\nabla_\theta \mathcal{L}_{cos} \tag{17}$$

   Where θ are the model parameters (embeddings layer weights, self-attention weights, forward pass network weights...), η is the learning rate and $\nabla_\theta \mathcal{L}_{cos}$ is the gradient of the loss with respect to θ. The empirical risk minimization objective is

$$\mathcal{L}_{cos} = \frac{1}{N}\sum_{i,j=1}^{N} \mathcal{L}_{ij} \tag{18}$$

4. **Learning Rate Scheduling:** Use warm-up to gradually increase learning rate at the start and decay strategies to reduce learning rate as training progresses.

Step 4: Evaluation

Compute the Accuracy, Precision, Recall and F1-score for evaluating the sentence embedding model.

1. Accuracy measures the proportion of total correct predictions (both true positives and true negatives) out of all predictions. It's useful when the classes are balanced. and its defined as follow:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

   Where TP (True Positives) is the correctly classified positive instances, TN (True Negatives) is the correctly classified negative instances and FP (False Positives) is the incorrectly classified negative instances as positive and FP (False Negatives) is the incorrectly classified positive instances into negative.

2. The Precision measures how many of the predicted positive cases were actually positive.

$$Precision = TP/(TP + FP)$$

3. Recall measures how many actual positive cases were correctly predicted.

$$Recall = TP/(TP + FN)$$

4. The F1-score is the harmonic mean of precision and recall.

$$F1 - Score = 2 \times (Precision \times Recall)/(Precision + Recall)$$

PUBLICATIONS

Speaking to the Markets: The Role of IMF Announcements in Investors’ Confidence
Working Paper No. WP/25/258