# Guardians of the Balance Sheet: Determinants of IMF Executive Board Sentiment

Merih Angin[a], Michael Breen[b]

[a]*Department of International Relations; Computational Social Sciences, Koç University*
[b]*School of Law and Government, Dublin City University*

**Abstract**

How do the Executive Directors of the International Monetary Fund react to loan recipients? This paper builds on a new, sentence-level measure of Executive Board sentiment to capture how Directors assess borrowing arrangements at the moment of program approval. Drawing on four decades of transcript data, we show that Board tone is shaped less by the borrower's political or economic profile than by the financial exposure involved. When the Fund's liquidity is tight, larger commitments trigger a marked decline in positive sentiment—consistent with the Board's fiduciary role in guarding scarce resources. This effect disappears when the Fund's forward commitment capacity is strong. A complementary analysis using spline regressions reveals that the Board's most critical language is reserved not just for very large programs but also for very small ones, with mid-sized loan recipients receiving the most positive assessments. These patterns suggest that Executive Directors modulate their rhetoric to reflect both institutional constraints and the perceived justification for exceptional lending.

*Keywords:* IMF, Executive Board, Conditionality, NLP, Machine Learning

## 1. Introduction

What explains the language the International Monetary Fund's (IMF) Executive Board uses when it approves a lending program? Existing research tells us a great deal about the size, maturity, and conditionality of Fund arrangements, but it remains largely silent on the texture of Board deliberation: the praise, caveats, and warnings that Directors voice while they decide whether to commit the institution's money. This paper brings

those deliberations into view by analyzing a new dataset by Angin (2023) that includes sentiment scores coded line-by-line from Board transcripts covering four decades of lending activity. Our aim is to identify the forces that shape Board sentiment, the aggregate polarity of Directors' interventions at the moment a program is approved.

The prevailing empirical literature offers only indirect glimpses of the rhetorical dimension of IMF governance. Quantitative studies of IMF lending typically model economic distress, voting power, or domestic political institutions to explain who borrows, how much, and on what terms. Much of this work highlights the outsized influence of the United States (Oatley and Yackee, 2004; Stone, 2008, 2011), while also identifying the roles of other powerful creditors, including supplementary financiers (Gould, 2006), corporate lenders (Dang and Stone, 2021), former colonial powers (Stone, 2004), the G5 economies (Copelovitch, 2010), and, more recently, China (Kern and Reinsberg, 2022; Ferry and Zeitz, 2024b,a). While many of these studies imply a link to Executive Board decision-making, none of them systematically examine the content of Board deliberations. This paper fills that gap and, in doing so, opens a new front in the study of IMF governance.

Our theoretical point of departure is the Board's dual identity as crisis responder and fiduciary guardian of quota resources. Directors must mobilize large sums quickly enough to stabilize external accounts, yet they are also accountable to shareholders who expect the Fund's revolving capital to be preserved. When the resources placed at risk are large, two dangers loom: moral hazard—too generous a rescue may encourage future policy slippage—and institutional illiquidity—the Fund's own balance-sheet buffer can be depleted by a handful of outsized commitments. We argue that Directors manage these dangers by modulating the tone of their interventions. Cautious or skeptical language serves as an inexpensive but credible signal that they recognize the exceptional risk attached to a given program. This logic yields a straightforward testable claim: as the volume of resources at stake increases, Board sentiment becomes less positive. The remainder of the paper develops this sentiment framework formally, operationalizes its key variables, and subjects it to a battery of empirical tests.

Our findings support this view. We show that Board sentiment becomes more negative as program size increases, but only when the Fund's own liquidity is under strain. When resources are ample, large commitments do not prompt concern; when liquidity is tight, they do. We also find that the relationship is not strictly linear: using a piecewise spline, we show that Directors

express more negative sentiment for both the smallest and largest programs, while mid-sized arrangements receive the most positive assessments. These results suggest that Board sentiment is not simply an echo of borrower characteristics, but a reflection of institutional constraints and the perceived risks attached to lending decisions.

Why should scholars and practitioners care about such rhetorical nuance? Sentiment is a forward-looking institutional signal with tangible consequences. Markets react to Directors' tone—filtered through post-meeting briefings—well before program implementation can be observed (Breen and Doak, 2023). Domestic political actors cite Board praise or skepticism when defending or contesting reforms. Fund staff themselves use the inaugural mood as a reference point in subsequent reviews. Analyzing what drives that mood therefore enriches our understanding of international crisis management in ways that parameter-only studies cannot.

Beyond its immediate contribution, the project speaks to broader debates in International Relations, International Political Economy and the study of International Organizations. It provides behavioral evidence on how shareholder power and institutional constraints shape multilateral discourse; it links that discourse to material stakes in global finance; and it demonstrates the scholarly dividends of the Fund's recent moves toward archival transparency. By treating Executive Board sentiment as both an outcome to be explained and an instrument of governance, the paper widens the analytical lens through which we view the world's most important crisis lending institution.

## 2. A Theory of Sentiment in IMF Lending

Recent scholarship has begun to examine not just what the IMF lends, but what it says. Studies of staff reports show that variation in tone and emphasis can influence domestic politics and market responses (Edwards, 2019; Breen et al., 2020; Ramos et al., 2022; Breen and Doak, 2023; Goes and Chapman, 2024). Other work has analyzed how the Fund frames economic ideas in official documents, documenting shifts in narrative that reflect changes in global politics, shareholder influence, and internal ideational trends (Ban and Gallagher 2015; Clift 2019; Kaya and Reay 2019). More recently, attention has turned to the Executive Board itself. Forster, Honig, and Kentikelenis (2025) use transcript length to measure Directors' voice and influence, while Arias, Clark, and Kaya (2025) examine how national preferences are ex-

pressed during Board discussions, particularly around climate policy. These studies challenge the view that the Board is merely a formality and suggest that its deliberations carry real political and institutional weight.

Yet one of the most consistent and visible signals to emerge from these discussions, namely the tone of Board sentiment, remains largely unexamined. Understanding that tone requires recognizing the Executive Board's dual role. It is both the Fund's lender of last resort, expected to act swiftly in moments of crisis, and the guardian of a finite pool of quota resources provided by 191 member states. That dual role creates two core concerns: moral hazard, where generous lending might encourage future policy failure, and institutional liquidity, where large commitments could strain the Fund's ability to support future borrowers. We argue that Executive Directors respond to these risks not just through formal program design, but also through language—modulating the tone of their interventions to signal concern or reassurance. The remainder of the paper sets out this framework and tests whether sentiment in Board discussions reflects the financial risks posed by the programs under review.

Moral hazard worries arise because generous rescue packages dampen the domestic political costs of past policy mistakes and may encourage future risk-taking. The larger the commitment, the more acute the reputational danger that the Fund will be seen to reward profligacy. Liquidity worries are mechanical. A single very large arrangement—let alone a cluster of them negotiated in close succession—can push the IMF's Forward Commitment Capacity sharply downward, raising the probability that the Fund itself will need to mobilise ad-hoc borrowing lines or quota increases. Both forms of risk motivate Directors to signal scepticism when programs place a disproportionate claim on common resources.

Language is the Board's first and most flexible signalling device. Before any funds are disbursed, long before the first review, Directors can broadcast a measured degree of caution—or, conversely, reassurance—through the tone of their interventions. Expressing doubts does not alter the legal parameters of the arrangement, but it does three useful things. It places reputational pressure on the authorities to deliver, it tempers market expectations about the speed of recovery, and it protects the Fund itself by documenting that the decision to lend was taken with eyes open to the risks involved. Because the reputational and liquidity stakes rise monotonically with the amount committed, a rational, guardianship-minded Board should calibrate its rhetoric accordingly.

From this logic we derive the core hypothesis that guides the empirical analysis:

***H1****: Holding constant the borrower's macroeconomic and political characteristics, Executive Board sentiment becomes less positive as the volume of Fund resources committed to a program increases, particularly when the Fund's own liquidity position is weak.*

The hypothesis implies a negative marginal relationship between loan size and the aggregate polarity score we observe in the transcript. If the Board is indeed acting as a judicious guardian, that relationship should survive controls for political divisions at the executive board, as well as economic and institutional conditions in the borrower, and it should strengthen when the Fund's own liquidity buffer is low or when global risk appetite is weak.

## 3. Measuring Executive Board Sentiment (Data and Methods)

### *3.1. Provenance and Scope*

This study builds on the sentence-level Executive Board (EB) sentiment dataset developed by Angin (2023), constructed from verbatim IMF Executive Board meeting minutes (1976–2017) using the following pipeline (see Angin (2023) for full details):

Sentences are coded with respect to the program under discussion as *positive* (+1), *neutral* (0), or *negative* (−1). Formulaic thanks without evaluative content are coded 0; future-oriented commitments are 0 unless explicitly endorsed (then +1); critiques of timing/feasibility/monitoring are −1.

Standard natural language processing techniques—tokenisation, stop-word removal, and lemmatisation—prepare the text for analysis. To improve classification accuracy, the model incorporates two key innovations. First, context-refined word embeddings are used: each word vector is averaged with those of neighbouring words, allowing the model to distinguish terms used in different semantic contexts (e.g., "increase" in relation to inflation versus GDP). Second, sentiment enrichment is applied by integrating polarity scores from the VADER lexicon into the word vectors, helping to differentiate words that are semantically similar but sentimentally distinct.

Angin (2023) deploys a two-track classifier: (i) a feature-based support vector machines (SVM)/ logistic regression (LR) ensemble using context-refined embeddings with lexicon valence features, and (ii) a fine-tuned BERT-family transformer trained on the annotated minutes to capture IMF-specific

language. The fine-tuned transformer attains very high held-out accuracy; the ensemble serves as a robustness check.

The SVM/LR ensemble classifier achieves approximately 85 percent out-of-sample accuracy. When fine-tuned on the same data, the BERT-based model further improves performance, exceeding 97 percent accuracy, confirming the tool's ability to parse the formal and often elliptical language typical of Executive Board deliberations.

The trained model is then applied to the full corpus of transcripts to generate two sentiment indicators. The first is a negative-sentence ratio, defined as the share of negative sentences in each discussion (EB1). The second is a net sentiment index: the difference between the number of positive and negative sentences divided by the total number of sentences (EB2). Higher values indicate greater support; lower values reflect increased criticism or concern.

Because the sentiment index is based entirely on Directors' language, it captures variation in Board reactions that is invisible in financial metrics or economic aggregates. Understanding the sources of this variation—why some programmes generate endorsement while others provoke dissent—is the central explanatory task of this paper.

Sentences retain speaker role metadata. In our baseline, we follow the abovementioned study's ED-focused construction and restrict to Executive Directors when forming the program-level outcome.

### *3.2. Outcome Definition (Approval-Level DV)*

Our dependent variable captures the aggregate tone of EDs during the *approval* discussion for program $i$ in country $j$ and year $t$. Using the sentence-level predictions from Angin (2023), we compute:

$$\text{EB1}_{ijt} = \frac{\#\text{neg}_{ijt}}{\#\text{pos}_{ijt} + \#\text{neg}_{ijt} + \#\text{neu}_{ijt}} \in [0, 1], \tag{1}$$

$$\text{EB2}_{ijt} = \frac{\#\text{pos}_{ijt} - \#\text{neg}_{ijt}}{\#\text{pos}_{ijt} + \#\text{neg}_{ijt} + \#\text{neu}_{ijt}} \in [-1, 1], \tag{2}$$

where counts are over ED sentences in the approval segment of the meeting.[1] EB2 is our main DV (higher = more positive tone); EB1 is a comple-

[1]When an approval spans multiple sittings on the same program, we concatenate those sittings prior to aggregation. We exclude formulaic thanks in a robustness check.

mentary measure used for diagnostics and sensitivity analysis.

To limit leverage from very short discussions, we require a minimum of $S_{\min}$=15 EB sentences and winsorize the top/bottom 1% of the ratios in (2)–(1). As a presentation check, we also standardize EB2 within decades; results are unchanged.

### *3.3. Alignment with the Explanatory Framework*

Because sentiment is the *outcome* of interest here, we timestamp EB1 and EB2 to the approval year $t$ and relate them to contemporaneous institutional covariates (e.g., Fund liquidity, absolute SDR size of the commitment). Country-level controls enter at $t-1$ to avoid mechanical contemporaneous contamination from the approval discussion.

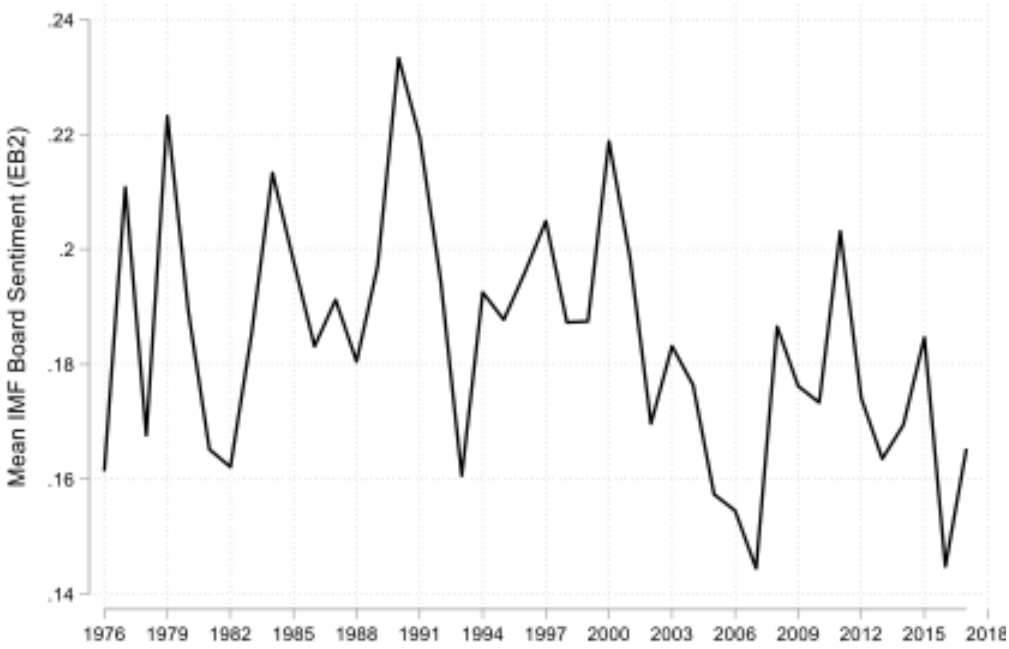
### *3.4. Validation*

Construct validity follows directly from Angin (2023)'s diagnostics (precision/recall, decade-stratified re-training, and historical face-validity around crisis episodes). In our setting, the approval-level aggregation reproduces the same cyclical patterns and yields stable distributions across decades; see Appendix for distributional and leverage diagnostics.

### *3.5. Descriptive Statistics*

Figure 1 shows a clear cyclical pattern in Executive Board sentiment across 640 programs from 1976 to 2017. Average sentiment begins at a moderate level in the late 1970s but declines through the early 1980s, reaching its lowest point around 1983–84, during the height of the Latin American debt crisis. It recovers through the late 1980s but dips again in the early 1990s as the Fund begins lending to transition economies following the collapse of the Soviet Union. From the mid-1990s to the mid-2000s, sentiment improves steadily, peaking just before the global financial crisis—during a period marked by strong global growth, ample IMF liquidity, and more precautionary lending. The crisis triggers a sharp drop in sentiment between 2007 and 2009, with a decline similar in scale to that of the early 1980s. Although sentiment recovers somewhat in the 2010s, with notable upticks during high-profile programs in the euro area and Ukraine, it never returns to its mid-2000s highs. This pattern is consistent with the paper's argument that Board tone reflects the broader context in which the Fund operates, with more cautious or negative language emerging during periods of heightened global or institutional risk.

Figure 1: IMF Executive Board Sentiment (EB2), 1976—2017




*3.6. Independent and Control Variables*

**Total SDR Commitments**

Our central independent variable is the face value of each program as approved by the Executive Board, measured in billions of Special Drawing Rights. This choice is motivated by the Fund's own budgeting logic: Directors vote on an absolute resource envelope, not a percentage of quota or a logarithmic transformation. When Directors weigh the merits of a program they confront a concrete opportunity cost—how many SDRs of the Fund's finite pool will be tied up for how long—so we treat the raw level of commitments as the clearest signal of balance-sheet exposure. Using the untransformed series preserves information at the top end of the distribution—precisely where we expect the political stakes, reputational risk, and potential moral-hazard concerns to be greatest.

Two alternative measures, namely log total SDR and an exceptional access dummy, are reported only as robustness checks. The logged form compresses the right tail and effectively down-weights the largest operations, masking the very variation we aim to explain. The exceptional access in-

dicator (for borrowing that exceeds three times a country's quota), while useful for separating programs that breach the Fund's normal access limits, is dichotomous and fails to distinguish among the many large arrangements authorized below.

**Liquidity Ratio**

Because a given commitment means more or less depending on the Fund's remaining firepower, we supplement the absolute SDR measure with the Liquidity Ratio, defined as $100 \times \frac{\text{liquid resources}}{\text{liquid liabilities}}$. A higher ratio signals ample uncommitted resources and thus a looser budget constraint; a lower ratio indicates tighter conditions and, we theorize, heightens Directors' scrutiny of any large request. In the main models we include total SDRs and the Liquidity Ratio interactively, allowing the marginal effect of loan size on sentiment to vary with the Fund's own balance sheet stress.

**Controls**

The models include a range of control variables capturing program characteristics, macroeconomic context, domestic political institutions, and geopolitical alignment. To account for the intensity of IMF involvement, we control for the number of conditionality categories attached to the program, which serves as a proxy for its stringency. Macroeconomic performance is captured using the borrower's real GDP growth and per capita growth rates, both lagged one year. To measure domestic political capacity, we include the Polity2 score, which ranges from -10 (strongly autocratic) to +10 (strongly democratic), on the assumption that stronger democratic institutions may reduce scepticism about program implementation. Finally, we include a measure of geopolitical alignment—specifically, a borrower's proximity to the G7 bloc on UN General Assembly votes—using a principal component derived from the ideal-point distances between the borrower and each G7 country. While all countries in the sample have already secured a program, variation in alignment may still shape Executive Board sentiment, as closer alignment with the Fund's major shareholders could increase trust and reduce political friction. All variables are lagged where appropriate, and models include country and year fixed effects, and standard errors clustered at the country level.

## 4. Empirical Results

Table 1 presents two fixed effects models examining how the IMF's resource position influences the tone of Executive Board discussions. In the

baseline model (with DV EB2), the liquidity ratio is the only variable that consistently predicts sentiment: a one-point decrease in the Fund's forward liquidity ratio, equivalent to the decline seen in major crisis years, is associated with a 6.3 percentage-point drop in positive Board sentiment ($b = -0.0627$, SE $= 0.0167$, $p < .01$). Loan size, measured in billions of SDR, shows no significant association with sentiment in this model. However, the second model (with DV EB1), which includes an interaction between loan size and liquidity, shows that program size matters when liquidity is low. The coefficient on loan size is negative and marginally significant ($b = -0.067$, SE $= 0.034$, $p \leq .05$), and the interaction with liquidity is positive and statistically significant ($b = 0.010$, SE $= 0.005$, $p \leq .04$). This implies that large loans receive more critical commentary from the Board when the Fund's resources are under pressure, but this effect diminishes as liquidity improves. For example, a two-billion SDR program would be associated with a 13-point increase in negative sentiment when liquidity is near zero, but no such penalty when liquidity is at or above average levels. None of the other control variables, such as economic growth, political regime, or breadth of conditionality, have significant effects, and the models account for 34 to 37 percent of within-country variation in Board sentiment. Overall, the findings suggest that the tone of IMF Executive Board discussions is shaped not only by country characteristics but also by the Fund's own resource constraints, especially in the case of large lending programs.

To test whether the link between program size and Board sentiment is truly linear, we replaced the single SDR variable with a piecewise linear spline that allows the slope to vary across three data-driven ranges: small loans (below about 4 billion SDR, the 25th percentile), mid-sized loans (4–8 billion), and very large loans (above 8 billion), as shown in Table 2. In the model of overall sentiment (EB1), none of the spline segments is statistically different from zero, indicating that the Board's aggregate tone is essentially flat across the size distribution. The picture changes for the negative-tone measure (EB2). Additional financing is associated with significantly more negative language for the smallest programs, a more neutral or slightly positive tone for mid-sized arrangements, and a return to sharper criticism once loans exceed 8 billion SDR. The resulting shallow U-shape suggests that the Board applies its toughest scrutiny to the very smallest and the very largest operations, while treating mid-range programs more leniently. Because these breakpoints are anchored in the sample distribution rather than in formal IMF access limits, the spline should be viewed as descriptive; it flags a non-

Table 1: IMF Executive Board Sentiment Towards Borrowers

| **VARIABLES** | (1) EB2 | (2) EB1 |
|---|---|---|
| Total SDRs (billions) | -0.001 | -0.067** |
| | (0.004) | (0.034) |
| Liquidity Ratio | -0.062*** | -0.003 |
| | (0.017) | (0.030) |
| Total SDRs × Liquidity Ratio | | 0.010** |
| | | (0.005) |
| Number Cat. Conditions | -0.001 | -0.001 |
| | (0.003) | (0.004) |
| GDP growth (t–1) | -0.001 | 0.001 |
| | (0.001) | (0.001) |
| GDP per capita (t–1) | 0.000 | 0.000 |
| | (0.000) | (0.000) |
| Polity2 (t–1) | 0.001 | -0.002 |
| | (0.002) | (0.002) |
| G7 Voting Proximity (t–1) | 0.004 | 0.004 |
| | (0.005) | (0.007) |
| Observations | 487 | 487 |
| R-squared | 0.342 | 0.369 |
| Year FE | YES | YES |
| Country FE | YES | YES |

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

monotonic pattern that future work, ideally using program-type or quota-based variables, can trace to the institutional rules that govern concessional and exceptional lending.

Table 2: Piece-wise Linear Spline

| **VARIABLES** | (1) EB1 | (2) EB2 |
|---|---|---|
| Small Loans | 0.009 | -0.015** |
| | (0.017) | (0.007) |
| Medium Loans | -0.012 | 0.031*** |
| | (0.021) | (0.007) |
| Large Loans | 0.002 | -0.014*** |
| | (0.005) | (0.003) |
| Observations | 579 | 579 |
| R-squared | 0.126 | 0.096 |
| Number of id | 123 | 123 |
| Year FE | YES | YES |
| Country FE | YES | YES |

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Two alternative measures (log total SDR and an exceptional access dummy) were tested and found to be statistically significant. In future iterations of the paper, we plan to incorporate additional covariates, including more detailed economic and geopolitical controls. These will include program type, refined indicators of UN voting alignment with the United States (focusing on votes salient to U.S. interests), income group classifications, and temporary membership on the UN Security Council.

## 5. Conclusion

This paper opens the "black box" of IMF decision making by showing that what Executive Directors say when they approve a program is systematically related to what the Fund is being asked to spend and to the depth of its own pockets at the time. Drawing on four decades of verbatim Board transcripts and a high-accuracy NLP sentiment model, we document three principal findings. First, when the Fund's forward-commitment capacity is tight, larger commitments elicit markedly more negative language—precisely

Table 3: Summary Statistics

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| EB1 | 0.112962 | 0.1263153 | -0.2330097 | 0.6172839 |
| EB2 | 0.1874941 | 0.0821343 | 0 | 0.4809524 |
| Total SDRs (billions) | 0.7282407 | 2.658096 | 0.00075 | 26.4329 |
| Liquidity Ratio | 5.289762 | 0.8266043 | 3.49 | 7.11 |
| Number Cat. Conditions | 5.762173 | 2.357057 | 1 | 13 |
| GDP growth | 3.405544 | 6.765948 | -64.05 | 149.97 |
| GDP per capita | 6446.032 | 14104.77 | 1.1 | 195772.7 |
| Polity2 | 1.655488 | 7.207333 | -10 | 10 |
| G7 Voting Proximity | -5.99e-10 | 2.367986 | -7.256482 | 4.700697 |

the signalling pattern predicted by the Board-as-guardian framework. Second, once we relax the assumption of linearity, negativity follows a shallow U-shape: Directors voice the most scepticism over both very small, often concessional, loans and the handful of mega-packages that dominate the right tail of the size distribution, while mid-range operations draw the mildest scrutiny. Third, these rhetorical adjustments are driven neither by borrowers' macrofundamentals nor by domestic political institutions; they are an endogenous response to the Fund's balance-sheet stress and to the reputational risks that accompany outlier cases at either end of the lending spectrum.

These results make three contributions. Empirically, they introduce a publicly replicable measure of Board sentiment that complements existing data on program size, conditionality and voting. Theoretically, they demonstrate that Directors use tone—not just legal conditionality—to balance their dual mandate as crisis responders and fiduciary stewards, lending modest support to "guardian" models of multilateral oversight. Substantively, they show that the Fund's internal liquidity constraint has behavioral consequences that ripple outward: borrowers contemplating large requests, and markets pricing their sovereign paper, should expect a colder rhetorical reception when the institution's own reserves are low.

At the same time, the analysis is consciously modest in scope. The spline thresholds are descriptive rather than institutional; they hint at, but do not prove, distinct regimes of concessional and exceptional access. Because we lack program-type identifiers, we cannot say whether the U-shape reflects concessional facilities that target the poorest members, the optics of bail-

outs for systemic cases, or both. Nor can we observe whether tonal shifts at approval translate into tougher oversight during subsequent reviews. Addressing these gaps will require new data—on facility type, quota shares, staff assessments, and post-approval performance—that can anchor the rhetorical patterns we uncover in the Fund's formal lending architecture.

Future work can extend the logic and the measurement strategy well beyond the IMF. Central bank swap networks, multilateral development banks, and even regional safety nets all confront the same tension between rapid crisis response and balance-sheet prudence; analyzing their deliberations could reveal whether "guardian sentiment" is a general feature of lender-of-last-resort governance. Within the Fund, pairing Board and staff rhetoric would illuminate how internal principal-agent dynamics shape the public face of multilateral advice. Finally, real-time event-study designs could test whether the tonal signals we identify move sovereign spreads, domestic reform trajectories, or public support for adjustment.

For scholars of international organization, the broader message is that words matter—and that they are deployed strategically when money is scarce. For policymakers, the findings underscore a practical point: requests that would draw little comment in flush times may face a chillier rhetorical climate when the Fund's liquidity buffer is thin. Understanding, measuring and ultimately explaining that climate is essential to a full account of how multilateral surveillance and crisis lending work.

## References


Angin, M., 2023. IMF Decides, Machine Learns: An AI Approach to IOs. APSA Annual Meeting 2023, Los Angeles.

Breen, M., Doak, E., 2023. The IMF as a Global Monitor: Surveillance, Information, and Financial Markets. Review of International Political Economy 30, 307–331. doi:10.1080/09692290.2021.2004441.

Breen, M., Hodson, D., Moschella, M., 2020. Incoherence in Regime Complexes: A Sentiment Analysis of EU-IMF Surveillance. JCMS: Journal of Common Market Studies 58, 419–437.

Copelovitch, M.S., 2010. The International Monetary Fund in the Global Economy: Banks, Bonds, and Bailouts. Cambridge University Press.

Dang, T.A., Stone, R.W., 2021. Multinational Banks and IMF Conditionality. International Studies Quarterly 65, 375–386.

Edwards, M., 2019. The IMF, The WTO & The Politics of Economic Surveillance. Routledge, London and New York.

Ferry, L.L., Zeitz, A.O., 2024a. China, The IMF, and Sovereign Debt Crises. International Studies Quarterly 68, sqae119.

Ferry, L.L., Zeitz, A.O., 2024b. The Power of Having Powerful Friends: Evidence from a New Dataset of IMF Negotiating Missions, 1985-2020. The Review of International Organizations 19, 411–442.

Goes, I., Chapman, T.L., 2024. Can "Soft" Advice from International Organizations Catalyze Natural Resource Sector Reform? International Studies Quarterly 68, sqae048. doi:10.1093/isq/sqae048.

Gould, E.R., 2006. Money Talks: The International Monetary Fund, Conditionality and Supplementary Financiers. Stanford University Press, Stanford, CA.

Kern, A., Reinsberg, B., 2022. The Political Economy of Chinese Debt and International Monetary Fund Conditionality. Global Studies Quarterly 2, ksac062.

Oatley, T., Yackee, J., 2004. American Interests and IMF Lending. International Politics 41, 415–429.

Ramos, L., Gallagher, K.P., Stephenson, C., Monasterolo, I., 2022. Climate Risk and IMF Surveillance Policy: A Baseline Analysis. Climate Policy 22, 371–388.

Stone, R.W., 2004. The Political Economy of IMF Lending in Africa. American Political Science Review 98, 577–591.

Stone, R.W., 2008. The Scope of IMF Conditionality. International Organization 62, 589–620.

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

## Appendix A. Figures

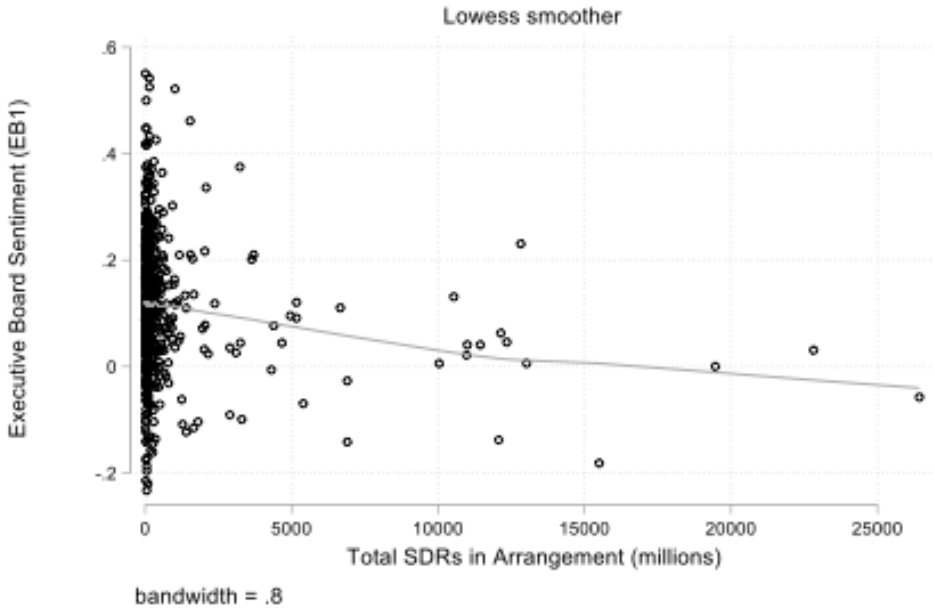



Figure A.2: EB1 vs. program size (total SDRs), 1976–2017 — LOWESS fit

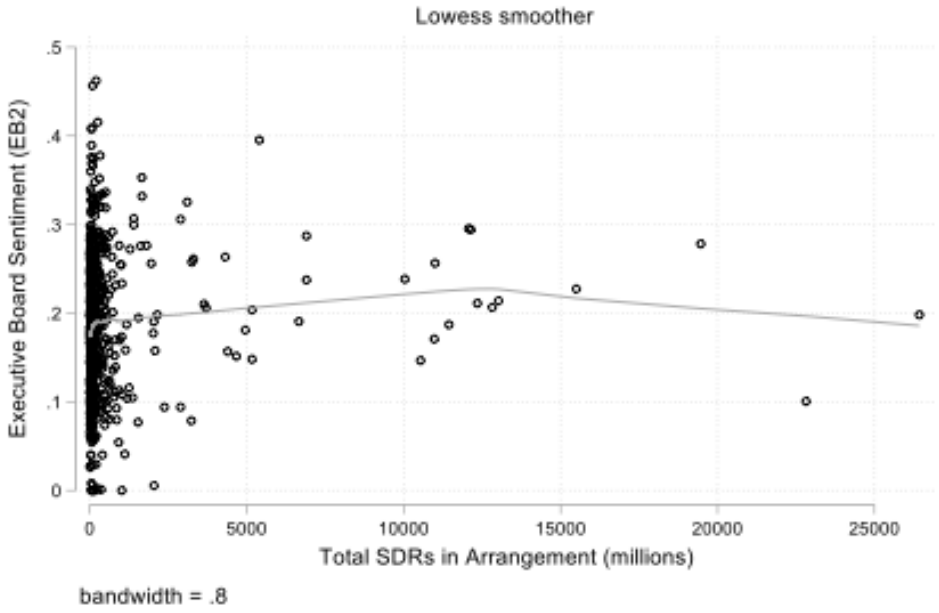



Figure A.3: EB2 vs. program size (total SDRs), 1976–2017 — LOWESS fit