The Language of Federal Reserve Reports: Framing Economic and Political Narratives

Oleg Sokolov
oleg.sokolov@uzh.ch

Yuan Yu
yuan.yu@uzh.ch

July 1, 2025

Contents

It is not the voice that commands the story;
it is the ear.

Italo Calvino, Invisible Cities

1 Introduction

The Federal Reserve's communications play a pivotal role in shaping expectations and transmitting monetary-policy signals to markets and the public. This paper analyzes the linguistic content of over two decades of FOMC meeting minutes (2000–2024) to uncover how the Fed frames uncertainty, signals confidence, and addresses domestic versus international concerns. Using a combination of lexical analysis, topic modelling (NMF), named-entity recognition (NER), and sentiment scoring with FinBERT, we identify patterns in how policy narratives evolve over time. Our findings reveal long-term trends in document length and lexical diversity, shifts in thematic emphasis—particularly on topics like inflation, employment, and financial markets—and a cautious tone in periods of economic disruption. We also observe distinct sentiment patterns in discussions involving international actors, such as China. While this study does not conduct full-scale event-based comparisons around specific crises, it lays the groundwork for such future analyses by providing robust time-resolved metrics of tone and topic salience. Overall, the results offer new insights into the communicative strategies of the Federal Reserve and their potential signalling function.

1.1 Motivation

The global economy of the 21st century has faced numerous significant challenges, including the dot-com bubble in the early 2000s, the global financial crisis of 2008, and the COVID-19 pandemic (Fund 2013). In such periods of turmoil, the way in which economic conditions are communicated by the media and institutions becomes especially critical. Among these institutions, central banks play a pivotal role in transmitting information to economic actors. This study focuses on the United States

Federal Reserve System and, in particular, on one of its key communication tools: the minutes of the Federal Open Market Committee (FOMC) meetings.

The FOMC minutes serve as an official channel for disclosing the discussions and decisions made during committee meetings. They are published by the Federal Reserve three weeks after each meeting and provide "a timely summary of policy discussions among participants."

From the standpoint of institutional economics, the publication of such documents can be viewed through the lens of signaling theory. According to Spence (1973), informed agents with access to privileged information use public actions to signal their intentions to the broader market. In this context, the FOMC minutes act as a signaling mechanism: by disclosing internal policy deliberations and the range of opinions within the Committee, the Federal Reserve communicates its expectations regarding future monetary policy, thereby reducing information asymmetry and guiding market forecasts.

Empirical studies confirm that financial markets closely monitor and respond to the release of FOMC minutes.(Rosa 2013) Their publication tends to trigger abnormally high volatility, similar in magnitude to the effects of major economic reports. The minutes provide detailed information on " the range of participants' views on the economic outlook and monetary policy," **??** offering investors insight into potential future interest rate changes. For example, Treasury yields often exhibit sharp increases in volatility following the release of the minutes, indicating that markets perceive the information as highly relevant and impactful.

The primary audience of the FOMC minutes includes traders, investors, economic analysts, and the financial press, all of whom closely track Federal Reserve policy. Given that the minutes are released on a regular basis and are synchronized with press conferences, they have become a critical signaling tool for conveying the Fed's current economic assessment and for indicating how close the Committee may be to adjusting the federal funds rate.

Central banks increasingly use communication as a strategic tool for managing public and market expectations, particularly during periods of economic shocks. As Federal Reserve Chair Ben Bernanke noted, following the 2008 financial crisis, transparency and explanatory efforts became "more prominent than ever."(Bernanke 2013, para. 2) Analysts emphasize that in exceptional times, central banks must clearly and swiftly articulate "the substance of measures and their justification" to build trust and ensure understanding of policy actions.(Fund 2013; Weidmann 2018) For instance, the International Monetary Fund highlights the need for "clear and timely communication" during unconventional crisis responses, arguing that it shapes public understanding and manages market expectations.(Fund 2013) According to Jens Weidmann, former President of the Deutsche Bundesbank, enhanced transparency strengthens public accountability and builds trust in the central bank—an essential asset in achieving monetary policy objectives such as price stability and full employment.(Weidmann 2018, para. 4)

1.1.1 Framing of Uncertainty versus Confidence

Prior studies show that central-bank narratives explicitly encode uncertainty or confidence as policy signals. For example, Fadda et al. (2022) find that policymakers tend to signal easing when they employ high-uncertainty language, whereas confirming positive trends indicates confidence and a bias toward tightening policies. In line with this, content analyses of FOMC minutes classify words such as "*uncertainty*", "*recession*", and "*downturn*" as dovish or negative, while descriptors like "*solid*", "*strength*", "*stronger*", and "*considerable*" populate the positive lexicon (Venade 2022).

This raises the question of whether the Fed adopts a systematically more cautious tone in the minutes when economic conditions sour—using phrases such as "*considerable uncertainty*" or "*downside risks*"—in contrast to a more confident tone in stable or strong periods (e.g. "*continued strength*", "*solid* *gains*"). Identifying the dominant words and topic clusters in each state could illuminate how the Committee frames uncertainty against confidence throughout the business cycle.

1.1.2 Tone Shifts Around Crises and Major Events

Do Fed communications take on a distinctly guarded or negative slant when shocks occur? Recent pandemic-era research finds a surge in uncertainty-related terms ahead of crises, followed by deliberately calming language once the turmoil is under way (Benchimol et al. 2025). Text analysis of FOMC minutes likewise shows a clear uptick in "*uncertainty*" terminology before each major shock, and a tapering of such language during the crisis itself (Benchimol et al. 2025). We ask whether these linguistic shifts function as early-warning narratives that foreshadow market turbulence. There is reason to expect such a link to financial markets: minutes with unexpectedly hawkish or pessimistic sentiment have been shown to lower stock returns and raise volatility immediately upon release (Rosa 2013). In other words, when the Fed's language turns more negative or uncertain, markets react accordingly. The empirical section therefore compares the Fed's language before, during, and after crises, and measures how strongly these sentiment swings align with stock indexes, bond yields, and volatility metrics.

1.1.3 Domestic versus International Discourse

A third focus of this study is how the FOMC balances domestic and international considerations in its minutes. From a signaling-theory perspective, central banks in an open economy must weigh foreign spillovers alongside domestic conditions to manage global market expectations and uphold policy credibility. Prior work shows that, although discussions of foreign economies are sparse in tranquil periods, global shocks elicit more extensive commentary on external risks (Obstfeld – Rogoff 2000; Blinder et al. 2008). Topic-model results further indicate that references to the "foreign economy" rose markedly even before the COVID-19 shock (Benchimol et al. 2025). Likewise, the minutes explicitly flagged risks emanating from abroad—such as Brexit and shifts in Chinese currency policy in 2016 (Venade 2022).

To investigate this dynamic, we will measure the share of text devoted to external issues and trace its evolution around global downturns—as the Fed seeks international coordination or acknowledges contagion risks. Recent research shows that Fed-originated sentiment shocks propagate to other central banks,(Armelius et al. 2019) and that domain-specific dictionaries can capture risk language in policy texts.(Correa et al. 2017) Building on the computational methods developed by Hansen and McMahon (2016) and Hansen et al. (2018), we will compile region-specific term lists (e.g., "*euro area*", "*Chinese renminbi*", country names) and compare their sentiment profiles with purely domestic topics. This analysis will reveal whether the Committee's tone systematically shifts when discussing international versus U.S. economic developments, shedding light on the Fed's role as a global policy actor.

1.2 Expected Contribution

This study will provide a deeper understanding of **how the Federal Reserve communicates economic uncertainty and risk**. By uncovering systematic patterns in its language, we aim to contribute to financial discourse analysis and offer insights relevant to economists, policymakers, and market analysts. The findings may also help improve financial forecasting by identifying linguistic signals associated with economic shifts.

Building upon previous literature on central bank communication, signaling theory, and text analysis, the research will primarily focus on the following three main research questions:

RQ1. Framing of Uncertainty vs. Confidence

- How does the Federal Reserve linguistically frame economic uncertainty compared to confidence?
- Do the minutes adopt a more cautious or urgent tone during economic downturns, employing phrases such as "considerable uncertainty" or "downside risks," versus more stable phases, characterized by language like "solid pace" or "continued strength"?
- Which words and topic clusters dominate during periods of uncertainty versus confidence, and how do these linguistic patterns serve as signals to markets?

RQ2. Tone Shifts Around Crises & Major Events

- Does the sentiment in Federal Reserve communications become noticeably negative or guarded around major economic and geopolitical shocks, including events such as the 2008 Lehman Brothers collapse, Brexit (2016), COVID-19 (2020), or the Silicon Valley Bank failure (2023)?
- Can consistent linguistic patterns be identified before, during, and after these events that might function as "early-warning narratives"?
- To what extent do these linguistic shifts correlate with responses in the stock market and broader financial markets?

RQ3. Domestic vs. International Discourse

- Are there systematic differences in the Fed's vocabulary when discussing the U.S. economy compared to foreign regions, such as Europe, China, and other emerging markets?
- Does the proportion of internationally framed language increase during global crises, suggesting a pivot toward narratives of international coordination and global interdependence?
- Which region-specific terms (e.g, "euro-area stress," "renminbi," "global supply chains") regularly appear, and how does their sentiment profile differ from discussions centered primarily on domestic economic conditions?

By addressing these research questions comprehensively, the study aims to highlight the significance of linguistic framing and sentiment in Federal Reserve communications, thus enriching our understanding of central bank transparency, market signaling, and the broader implications for economic forecasting and policy analysis.

2 Methodology

2.1 Related Literature

2.1.1 Central Bank Communication and Linguistic Framing

Academic work on central-bank communication has long treated text as data to uncover hidden information in policy discourse. Early studies applied latent semantic methods to FOMC minutes—for example, Hansen et al. (2018) employ LDA on Fed transcripts and recover interpretable topics such as "financial sector" and "economic weakness," which co-move with the business cycle. Dictionary-based sentiment analysis is also common: the financial lexicon of Loughran and McDonald (2011) is now a standard tool for measuring tone in Fed releases, and optimistic tone has tangible market effects. Gu et al. (2022), for instance, show that a more positive tone in FOMC minutes predicts higher stock returns during periods of elevated policy uncertainty. More recent work blends machine learning with topic methods: Ahrens and McMahon (2021) build a supervised topic model of central-bank speeches and construct a "monetary-policy signal dispersion" index; greater dispersion in the run-up to an FOMC meeting precedes larger-than-expected policy surprises. Together, this literature demonstrates that not only *what* central banks say but also *how* they frame and coordinate their language influences expectations and market outcomes.

2.1.2 Sentiment Analysis in Financial Texts

Sentiment extraction in finance has progressed from simple lexicons to advanced neural models. Loughran–McDonald and other finance-specific dictionaries remain widely used in central-bank studies(Loughran – McDonald 2011), but general-purpose tools often struggle with economic text. Transformer-based models fine-tuned on financial data (notably FinBERT) have recently become popular. Araci (2019) (Araci 2019) introduces FinBERT—a BERT model pre-trained on financial corpora—and shows it delivers large gains in financial-sentiment classification accuracy (e.g. +14% on a standard benchmark) (Araci 2019). Building on this, domain-specific fine-tuning has been applied to monetary-policy texts. Hilscher et al.(Hilscher et al. 2024) employ FinBERT to score sentiment in Fed and ECB communications (focusing on FOMC minutes and ECB press conferences), finding that Fed and ECB sentiment co-move and that Fed sentiment Granger-predicts future policy rates(Hilscher et al. 2024). Likewise, Gössi et al.(Gössi et al. 2023) fine-tune FinBERT on FOMC minutes using a "Sentiment-Focus" strategy and report improved accuracy on complex sentences (e.g. those with mixed clauses) (Gössi et al. 2023). These studies demonstrate that modern NLP models can capture subtle tone in Fed language. In our work, we similarly use the pre-trained FinBERT for sentence-level sentiment scoring, which aligns with the state-of-the-art approach in central-bank NLP (Araci 2019; Hilscher et al. 2024).

2.1.3 Topic Modeling for Policy Analysis

Automated topic modeling has become a staple for summarizing policy texts. The economics literature typically uses Latent Dirichlet Allocation (LDA)(Blei et al. 2003) or related matrix factorizations(Hansen et al. 2018; Hansen 2023). For example, Hansen et al. (2018) illustrate how LDA on FOMC transcripts produces human-interpretable topics that reflect economic discussion. Hansen (2023) notes that LDA (essentially NMF of a TF–IDF matrix with Dirichlet priors) is ubiquitous in economics owing to its efficiency and interpretability(Hansen 2023).

Following this tradition, we apply TF–IDF vectorization of the FOMC minutes and extract topics via non-negative matrix factorization (NMF). NMF is mathematically akin to LDA—both decompose the term-document matrix—yet often yields comparably interpretable topics(Li et al. 2018). We select the number of topics by examining the NMF reconstruction-error curve, balancing accuracy and parsimony as recommended in the topic-model selection literature(Greene – O'Callaghan 2014; Cai – Wu 2021). This mirrors prior work that evaluates topic models for coherence and fit (e.g., Arora et al., 2013)(Arora et al. 2013).

Our discovered topics (twenty in our design) align with typical macro themes such as "inflation," "employment," and "financial markets," and we track their prevalence over time. Such temporal-topic analysis has antecedents in studies of central-bank communication—for instance, ECB researchers track topic shares across strategic changes(Bank 2021; Ferrari – Ritschl 2024). In short, our TF–IDF + NMF approach is grounded in standard NLP methodology for policy texts(Jiang – Xiao 2022).

2.1.4 Named Entity Recognition for Framing

Modern NLP tools allow explicit detection of entities (people, organizations, places) which can aid framing analysis. In the central-bank context, NER has been used to filter or classify discourse. For example, Picault et al. (2022) apply NER and POS tagging to thousands of news articles, separating passages that quote the central bank from journalist commentary. By analogy, one can tag country or region names (e.g. "U.S.", "Europe", "China") or institutions, and thereby gauge whether a text segment has a domestic or international focus. In our study, we similarly use an NER model to label each sentence in the FOMC minutes. Sentences that mention only U.S. entities (Treasury, states, cities) are counted as "domestic", whereas those naming foreign entities (IMF, foreign economies, trade partners) are "international". This follows a growing practice of using entity tags to classify parts of policy discourse: related work at the ECB shows the utility of custom dictionaries for segregating inflation-related segments(Bank 2021). Overall, NER serves as a data-driven complement to lexical methods in framing analysis.

2.1.5 Event-Study Analyses in Textual Data

Financial economists often use event-study methodologies to isolate news effects, and this approach has been extended to textual features. In a classic example, Kuttner (2001) examines narrow windows around Fed announcements to identify policy shocks in yields. We similarly define fixed windows around each FOMC meeting to compare text-derived metrics before and after the event. In practice, a symmetric window (e.g. ±15 trading days) is applied, and shifts in sentiment or topic frequency are tested via *t*-tests or non-parametric tests. This mirrors work by central-bank researchers who align text metrics with market moves on announcement days(Bank for International Settlements 2023). For instance, the BIS regresses high-frequency yield changes on policy-statement sentiment in a one-day window, a design that "limits confounding factors"(Bank for International Settlements 2023). Following that template, we compute average sentiment/topics immediately before and after each FOMC and test for significant differences—an approach widely used for news and policy events(Gu et al. 2022).

2.1.6 Causality Tests between Text and Markets

To assess predictability, researchers often employ Granger-causality tests. One estimates a VAR in which lagged text measures (sentiment, topic indices) help forecast returns or yields, and then tests whether those lags add explanatory power, following Granger (1969). Jarociński and Karadi (2020), for example, decompose Fed announcements into surprise and information shocks and show that monetary surprises Granger-cause bond yields. In our study we test whether FOMC-derived sentiment or topic series Granger-cause DJIA or Treasury movements. This aligns with evidence that central-bank sentiment helps forecast future policy settings(Hilscher et al. 2024) and follows media-based work such as Bansal and Shaliastovich (2013). Standard stationarity checks and lag-length criteria are applied before the Wald tests of Granger non-causality.

2.2 Methods

In sum, our methodology builds directly on the previous literature. We begin by cleaning and tokenizing all FOMC-minute texts, and then apply the finance-specific transformer **FinBERT** to score sentence-level sentiment (Araci 2019; Hilscher et al. 2024). Next, we extract latent themes with a TF–IDF + NMF pipeline—mirroring the dominant practice of topic modelling in economics (Hansen 2023). The optimal topic count is chosen at the reconstruction-error "elbow" to balance fit and parsimony. We track the resulting sentiment and topic series in symmetric event windows around each meeting, following the spirit of Kuttner (2001) and BIS event-study templates (Bank for International Settlements 2023). Predictive power is evaluated with standard Granger-causality tests (Granger 1969; Jarociński – Karadi 2020). Finally, named-entity recognition (NER) tags each sentence as *domestic* or *international*—an extension of entity tagging in policy discourse (Picault et al. 2022). We also integrate the supervised "policy-signal" insights of Ahrens and McMahon (2021) so that our unsupervised topics can be mapped onto their signal taxonomy. Altogether, each pipeline component—tokenisation, lexicon filtering, FinBERT sentiment, NMF topics, event-window tests, and Granger regressions—is grounded in cutting-edge NLP and financial-economics research, making its application to FOMC minutes a logical extension of recent advances.

To guide the empirical work we state three research questions (RQs). The corresponding hypotheses, empirical strategies, and evaluation metrics are summarised in Table 1.

Table 1: Research questions, hypotheses, and empirical strategies

Research Question	**Hypothesis**	**Methodological approach (incl. key metrics)**
RQ1: Framing of Uncertainty vs. Confidence	The FOMC balances uncertainty and confidence: cautious language rises in volatile periods, confident wording in stable periods (signalling the Committee's outlook).	*Lexical framing:* custom "uncertainty–confidence" dictionary; term & collocation counts; modal-verb context. *Metrics:* meeting-level uncertainty index, quarterly average, and z-scores for cross-time comparison.
RQ2: Tone Shifts Around Crises & Major Events	Major crises (e.g. 2008, COVID-19) induce more negative or cautious tone; normal times exhibit neutral/positive tone.	FinBERT sentiment scores; $\pm$15-day event windows; paired *t*-test and Mann–Whitney test; structural-break (Bai–Perron) check; NMF topics to surface crisis themes. *Metrics:* mean sentiment, volatility of sentiment, topic-share deltas.
RQ3: Domestic vs. International Discourse	Minutes emphasise domestic issues overall, but international references spike when global conditions threaten the U.S. outlook.	spaCy NER with region tags (GPE, ORG); share of foreign entities per meeting; NMF topic proportions labelled "international"; two-sample tests on sentiment between domestic and international subsets. *Metrics:* international entity share, international topic share, sentiment gap (Δ).

Significance: These questions connect directly to the broader literature on central-bank communication as a signalling device. Resolving information asymmetries between policymakers and the public echoes the classic job-market signalling logic of Spence (1973). In a low-rate environment, what the Fed *says* has become a de-facto policy instrument (Ahrens – McMahon 2021). Analysing uncertainty framing (RQ1) tests whether the Fed's language reveals hidden confidence levels; tone shifts around crises (RQ2) gauge adaptive rhetoric under stress; and distinguishing domestic from international discourse (RQ3) recognises the Fed's increasingly global vantage. By marrying FinBERT sentiment, NMF topics, NER tagging, event windows and Granger tests, we provide a comprehensive, theory-consistent lens on how the Committee's rhetoric guides market perceptions and complements traditional policy tools.

3 Data and Methodological Pipeline

3.1 Data

The dataset for this study consists of the full text of the Federal Open Market Committee (FOMC) meeting minutes published by the U.S. Federal Reserve. These minutes provide detailed insights into the Federal Reserve's discussions and decision-making processes regarding monetary policy.

Collection: To construct the dataset, we implemented a custom web scraping script in Python using the requests and BeautifulSoup libraries. The scraper systematically attempted to access all possible publication dates from 2000 to 2024 by dynamically generating URLs based on valid combinations of year, month, and day in the format used by the Federal Reserve's website. For each valid URL returning an HTTP 200 response, the full HTML page was parsed and cleaned, and the resulting text was saved to a .txt file.

Coverage: The final dataset includes **193 FOMC meeting** transcripts spanning over two decades (2000–2024), yielding ~9 MB of cleaned plain text (roughly 1.6 million words):contentReference[oaicite:2]index=2. Each file is timestamped with the official meeting date, enabling monthly or event-window aggregation.Each text file contains the full content of a single meeting's minutes. These documents were then preprocessed and analyzed to extract various linguistic and statistical features, as well as sentiment scores.

To ensure ethical scraping behavior, we included a one-second delay between requests to minimize server load.

3.2 Methodological Pipeline

Our analysis follows a structured NLP pipeline, from raw text collection to statistical evaluation. Figure 1 summarises the end-to-end workflow—from raw text to statistical evaluation—together with the principal justification for each step.

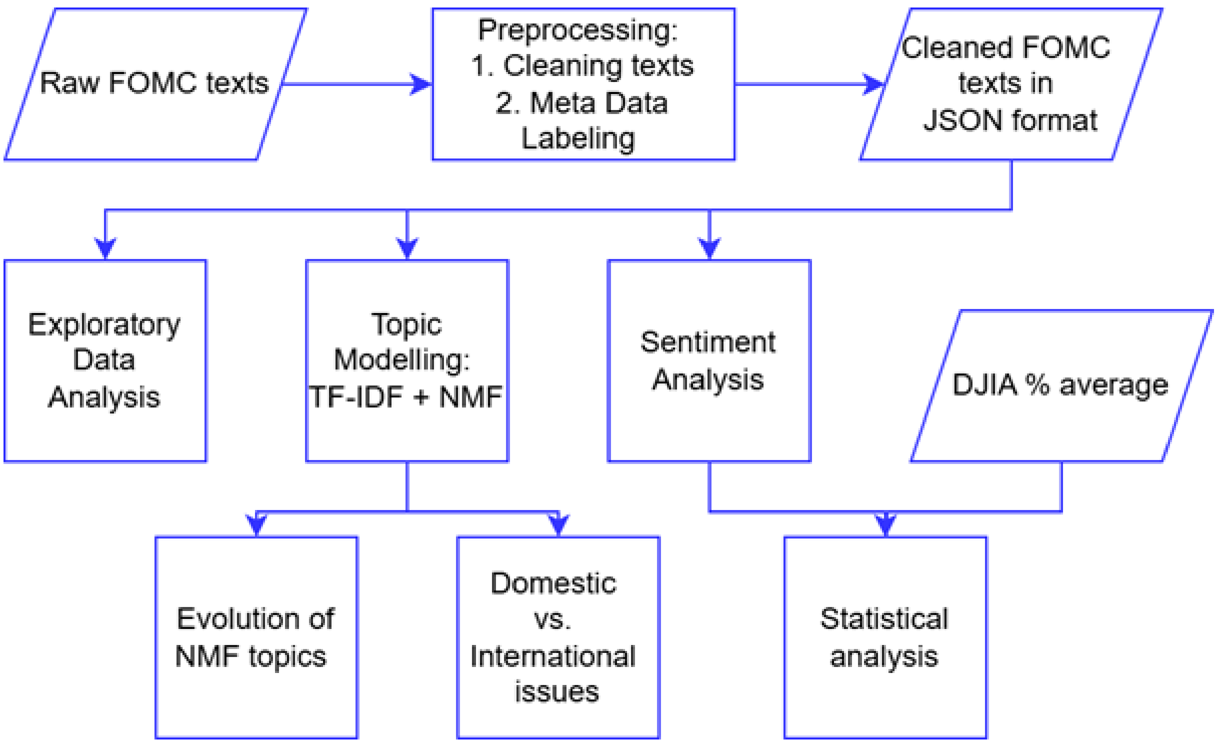

Figure 1: FOMC Text Analysis Framework

Key steps in the workflow include data ingestion and cleaning, exploratory analysis, feature extraction (topics, sentiment, entities), and subsequent statistical testing. We outline these stages below:

3.2.1 Data Collection and Storage

Using the custom web scraper described in Section 2.2, we retrieved 193 FOMC meeting minutes (2000–2024). Each document was timestamped by meeting date and stored in a structured format

(JSON), facilitating alignment with time-series events (e.g. crises or market data). This structured corpus allowed efficient iteration and indexing by date for subsequent analysis.

3.2.2 Text Preprocessing

We applied standard NLP preprocessing to each document. All texts were lowercased and tokenized into words and sentences. We removed punctuation, numerical artifacts, and standard English stop-words. In addition, domain-specific cleaning was performed: we eliminated procedural sections and attendee lists (e.g. names of FOMC members and staff) that appear in minutes but carry no economic content. This step prevented frequent names or titles from skewing the word frequency distribution. We also dropped common boilerplate phrases and page headers. The cleaned tokens were then lemmatized (converting words to base forms) to normalize variants (for example, “markets” to “market”). This yielded a processed text ready for analysis, with each meeting represented as a list of normalized tokens and sentences.

3.2.3 Exploratory Data Analysis (EDA)

Prior to advanced modeling, we conducted EDA to understand basic textual characteristics. We computed aggregate metrics such as word counts, sentence counts, and lexical diversity for each document (see textual features listed in Section 3.1). This confirmed, for instance, that document lengths have grown over time and vocabulary richness varies with economic context (as noted later). We visualized word frequency distributions to identify prominent terminology. A word cloud of the entire corpus (not shown here) highlighted frequent terms like “economy,” “inflation,” “market,” “growth,” and “policy,” reflecting the core focus of FOMC discussions. Common bigrams included phrases such as “interest rate”, “labor market,” “financial conditions,” and “economic growth,” indicating recurrent themes in the dialogue. These n-gram analyses provided intuition about the content emphasis and ensured that cleaning steps (e.g. removal of names) were effective – for example, policy terms dominated the word cloud rather than participant names, confirming successful filtering of irrelevant tokens.

3.2.4 Feature Extraction - Topics

We transformed the corpus into a document-term matrix with TF–IDF weighting (unigrams and bigrams) and applied Non-negative Matrix Factorization (NMF) to uncover latent topics. We experimented with various topic numbers and settled on k = 20 topics, balancing model fit and interpretability. Each meeting thus obtained a vector of topic weights (proportions summing to 1). For interpretability, each topic was characterized by its top keywords (e.g., one topic is defined by terms “inflation, prices, CPI, price stability,” another by “labor, employment, jobs, unemployment,” etc.). We tracked these topic weights over time for temporal analysis. To focus on the most meaningful patterns, we identified the six most salient topics (those with highest average weight in the corpus) for deeper examination. As described below, this time-resolved topic analysis reveals how the Fed’s emphasis on certain themes shifts in response to economic events.

3.2.5 Feature Extraction – Sentiment

We employed a domain-specific transformer model, FinBERT (Araci, 2019), to perform sentiment analysis on the minutes. Specifically, we split each document into sentences and fed each sentence through FinBERT, which classifies sentiment as Positive, Negative, or Neutral based on financial context. This yielded a sentiment label (and probability) for every sentence in every meeting. We then aggregated these sentence-level results to obtain document-level sentiment measures. In line with prior literature, we constructed a sentiment index per meeting defined as the fraction of negative sentences (or alternatively, a net score= $\#\text{positive} - \#\text{negative}$ sentences, as a robustness check).

This approach captures the document's overall tone while accounting for its length. The output was a time series of sentiment for the FOMC meetings, which could be analyzed for trends, structural breaks, and correlation with external variables. FinBERT's financial tuning ensured that words like "decline" or "weakness" were correctly interpreted as negative in context (whereas general-purpose sentiment lexicons might misclassify certain technical terms). We also verified the classifier's output on sample sentences to ensure it aligned with human intuition (for example, "labor market conditions improved further" was labeled positive, while "downside risks to growth remain significant" was labeled negative, as expected).

3.2.6 Feature Extraction – Named Entities and Thematic Tags

To address RQ3 on domestic vs. international focus, we applied Named Entity Recognition (NER) (using spaCy's NLP toolkit) to tag country names, regions, and organizations in the text. Each sentence was labeled as "international" if it mentioned foreign entities (e.g. "Europe," "China," "IMF") and contained no U.S.-specific terms, or as "domestic" if it referred exclusively to U.S. places/institutions (e.g. "Congress," "Treasury," "New York"). Sentences containing both domestic and international entities were categorized by context or counted in both where appropriate. This allowed us to quantify the fraction of international content in each meeting (e.g. percentage of sentences focusing on foreign economies) and to extract subsets of text specifically about major foreign countries. In particular, we paid special attention to references to China – a country frequently cited in recent Fed communications (for example, in context of supply chain disruptions, currency valuation, or global growth concerns). We flagged sentences mentioning "China" or "Chinese" and separately tracked their sentiment via FinBERT. This granular annotation enabled a comparative sentiment analysis: how the Fed's tone differs when discussing overseas economies (like China) versus when discussing domestic matters.

3.2.7 Integration and Visualization

After extracting these features, we compiled a comprehensive dataset where each meeting (date) is a data point with multiple attributes – e.g. textual metrics (length, diversity), sentiment score, topic vector, and international mention count. We then produced visualizations to inspect temporal patterns. For example, we plotted the hedging vs. certainty language ratio over time (based on counts of uncertain words like "may, could, uncertain" vs. confident words like "will, expect, certain"), as well as the prevalence of key topics and the sentiment index over the 2000–2023 period. These exploratory plots immediately suggested that the Fed's language varies meaningfully with economic cycles – showing, for instance, spikes in uncertainty language around 2008 and 2020, and shifts in topic focus before and after crises. Such visual findings motivated the formal statistical tests described later.

3.2.8 Statistical Analysis and Hypothesis Testing

In the final stage of the pipeline, we quantified and tested the patterns suggested by the exploratory results. We implemented event-study analyses to compare text features before, during, and after major events (e.g., looking at sentiment in the meetings leading up to the 2008 crisis vs. during the crisis). We also conducted hypothesis tests (t-tests and nonparametric ranksum tests) to assess whether differences in language metrics between crisis and non-crisis periods are statistically significant. Finally, to explore predictive relationships, we carried out Granger causality tests within a vector autoregression (VAR) framework, examining whether lagged textual indicators (such as sentiment or topic frequencies) add explanatory power for subsequent movements in financial variables (like stock index returns or bond yields). These statistical steps, rigorously evaluate our research questions about the connection between Fed rhetoric, economic conditions, and market responses.

This end-to-end pipeline ensured that raw textual data from FOMC reports was systematically transformed into meaningful quantitative indicators and evaluated in light of our hypotheses. The following sections present the results of applying this pipeline – first analyzing sentiment trends, then topic dynamics, and finally the statistical significance and economic relevance of the patterns identified.

4 Initial Results: Exploratory Textual Analysis

(Sections 4.1, 4.2 and 4.3 report initial results from textual analysis before focusing on sentiment and statistical tests.)

4.1 Basic Metrics and Trends

In this study, a set of fundamental textual metrics was calculated to characterize the FOMC minutes over time. These metrics provide insights into the linguistic complexity, length, and diversity of the documents, enabling a deeper understanding of temporal trends in Fed communication style. The following features were extracted from each text and visualized (see Figure 2):

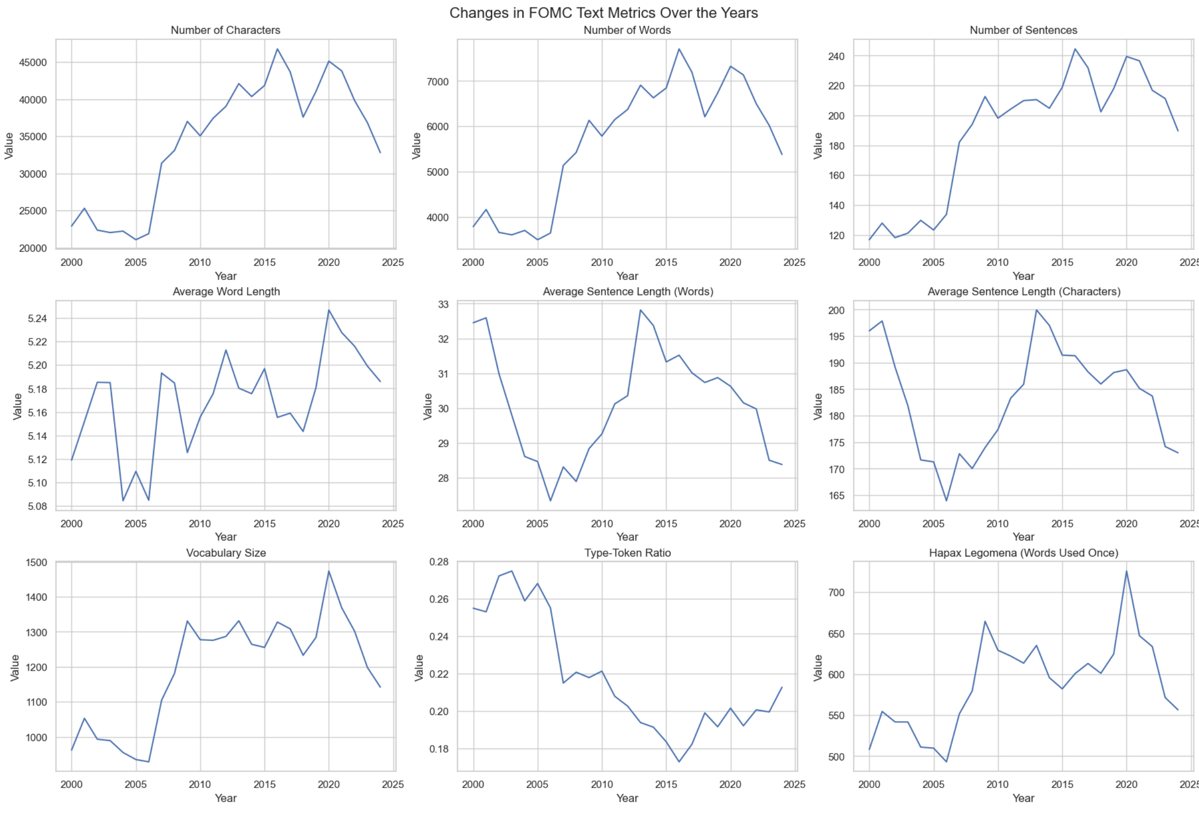

Figure 2: Exploratory Data Analysis

Character Count (n_chars): The total number of characters in the document, reflecting the overall length of the text.

Word Count (n_words): The total number of tokens (words), providing a basic measure of document size.

Sentence Count (n_sentences): The number of sentences, which helps to assess text segmentation and complexity.

Average Word Length: Calculated as the mean number of characters per word, this metric serves as an indicator of lexical complexity and word formation.

Average Sentence Length: Evaluated both in terms of words per sentence and characters per sentence, these metrics provide insight into syntactic complexity and readability.

Vocabulary Size (vocab_size): The number of unique word types in the text, indicating lexical richness.

Type-Token Ratio (TTR): Defined as the ratio of unique word types to total tokens, TTR serves as a normalized measure of lexical diversity. It is acknowledged, however, that TTR tends to decrease with increasing text length, which must be considered when interpreting results.

Hapax Ratio: The proportion of hapax legomena relative to the total number of tokens, providing a length-normalized indicator of lexical rarity.

First, there is a marked increase in lexical novelty during periods of major economic disruption. Specifically, the number of hapax legomena spikes in crisis years like 2008–2009 and 2020. During the 2008 financial crisis and the onset of the COVID-19 pandemic, the Fed minutes introduced many new or rare terms, driving up the hapax count. This likely reflects policymakers grappling with unprecedented situations (e.g. "quantitative easing" became common around 2008, pandemic-related terms in 2020) and thus using new terminology not seen in prior minutes. In contrast, in stable periods the language was more repetitive and stayed within a familiar lexicon, yielding fewer unique one-off words. The rise in hapax legomena during turmoil suggests that crises force the Fed to discuss novel concepts and emergency measures, making the language more unique.

Second, the minutes have grown significantly in length over the past two decades. Both total word count and characters per document show an upward trend. For example, the average FOMC minute in 2020–2023 is substantially longer than those in the early 2000s. This could indicate an increase in the complexity of issues discussed or a deliberate move towards greater transparency and detail in recent years. The Fed may be providing more extensive explanations and commentary in minutes than it did in the past, aligning with a general trend of more communication from central banks. Longer documents also tend to have slightly lower TTR (since as documents grow, adding new unique words becomes harder), but we still observe robust vocabulary growth alongside length.

Overall, the expanding length and persistent introduction of new terms underscore an evolving communication strategy, possibly to address the growing complexity of the economic environment and policy toolkit.

4.2 Topic Modeling Results: Thematic Shifts Over Time

Using the NMF topic model (20 topics) on the corpus, we identified thematically coherent topics and observed how their prevalence changes in response to economic events. As shown in Figure 3, the NMF reconstruction error decreases steadily with more topics, with no elbow point, hence the reasonable amount of topics we choose is 20.

For each meeting, we computed the proportion of discussion devoted to each topic (the topic weight), and tracked the time series of topic shares for key themes. We focus on the six most prominent topics (by average weight) which roughly correspond to: **(1)** Inflation and Prices, **(2)** Labor Market and Employment, **(3)** Financial Markets and Crises, **(4)** Economic Growth and Output, **(5)** Monetary Policy and Interest Rates, and **(6)** International Economy and Trade. the change of themes over time is shown on the graph (see Figure 4).

Each of these topics exhibits distinct temporal dynamics: Inflation (Prices) – Inflation-related discourse is a consistently present topic, reflecting the Fed's price stability mandate. Its share of the minutes was relatively stable during the low-inflation 2010s, but it surged in the post-2021 period when inflation re-emerged as a pressing concern. We see a clear uptick in the fraction of text about inflation and prices starting in late 2021 and through 2022–2023 (corresponding to the inflationary

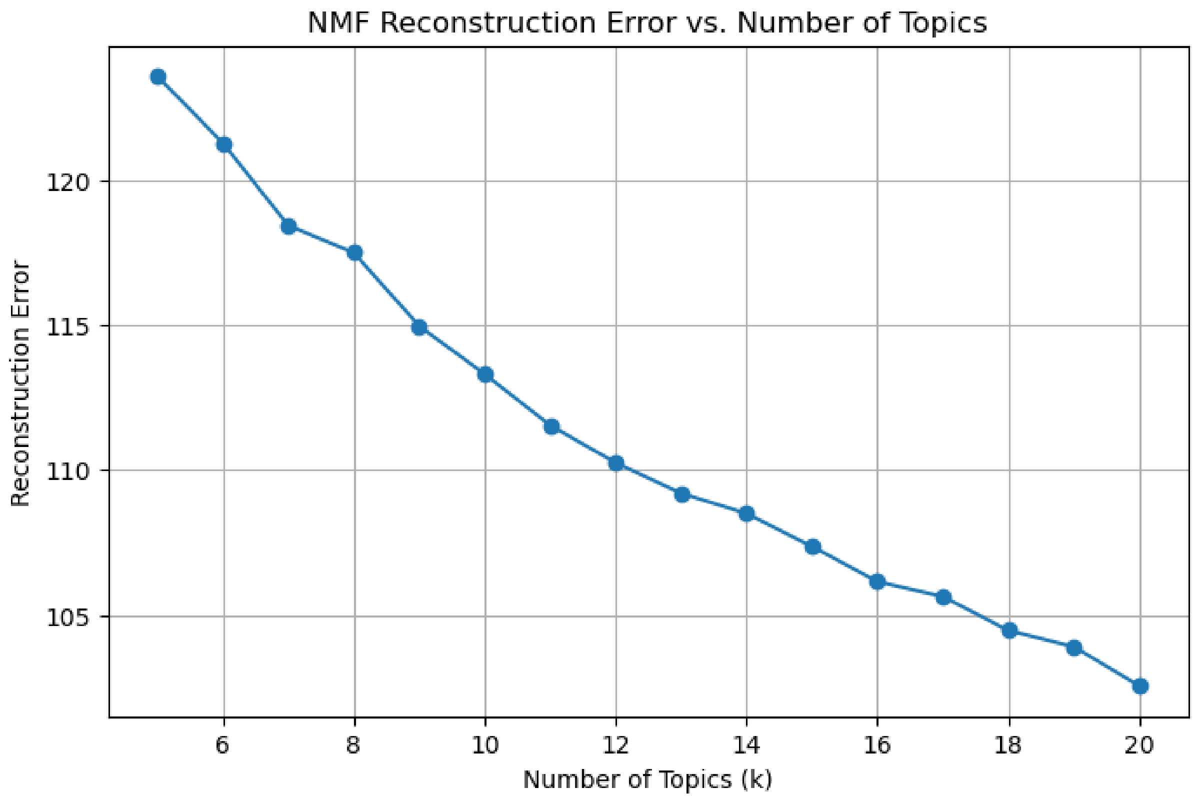

Figure 3: NMF Reconstruction Error

spike in the economy), as the FOMC devoted more discussion to rising prices and the need for policy responses. Earlier, a smaller bump in the mid-2000s is evident (around 2004–2006) when inflation briefly ticked up, but overall the 2020s increase is the most pronounced, indicating inflation becoming the dominant narrative in recent communications.

4.2.1 Labor Market (Employment)

Employment and labor market discussions show counter-cyclical patterns. This topic's prominence peaked during recessions: for instance, it rose sharply in 2001–2003 (after the dot-com bust and 9/11 shock) and again in 2008–2010 (during and after the global financial crisis), as high unemployment and job losses became central issues. Another spike occurred in 2020 amid the COVID-19 lockdowns, when tens of millions of jobs were shed and the Fed minutes heavily emphasized labor market conditions. In contrast, during expansions (e.g. mid-2000s and late 2010s) when unemployment was low, the relative share of text about labor/employment was lower – the Fed still discussed labor market tightness, but it occupied a smaller fraction of the overall narrative. These trends confirm that the Fed's focus on employment intensifies in downturns, consistent with its dual mandate (maximizing employment alongside controlling inflation).

4.2.2 Financial Markets and Crises

One topic captures financial stability, markets, and crisis responses, with keywords like "financial markets, credit, liquidity, banks, crisis, volatility." This topic's share spikes dramatically in 2008–2009, reflecting the Committee's intense focus on financial turmoil, bank stability, and extraordinary interventions during the Great Recession. It then gradually subsides in the early 2010s as acute crisis conditions abate. A smaller resurgence appears in 2015–2016 (amid concerns like the Chinese stock market turbulence and Brexit referendum, which introduced volatility) and again in 2020, when the

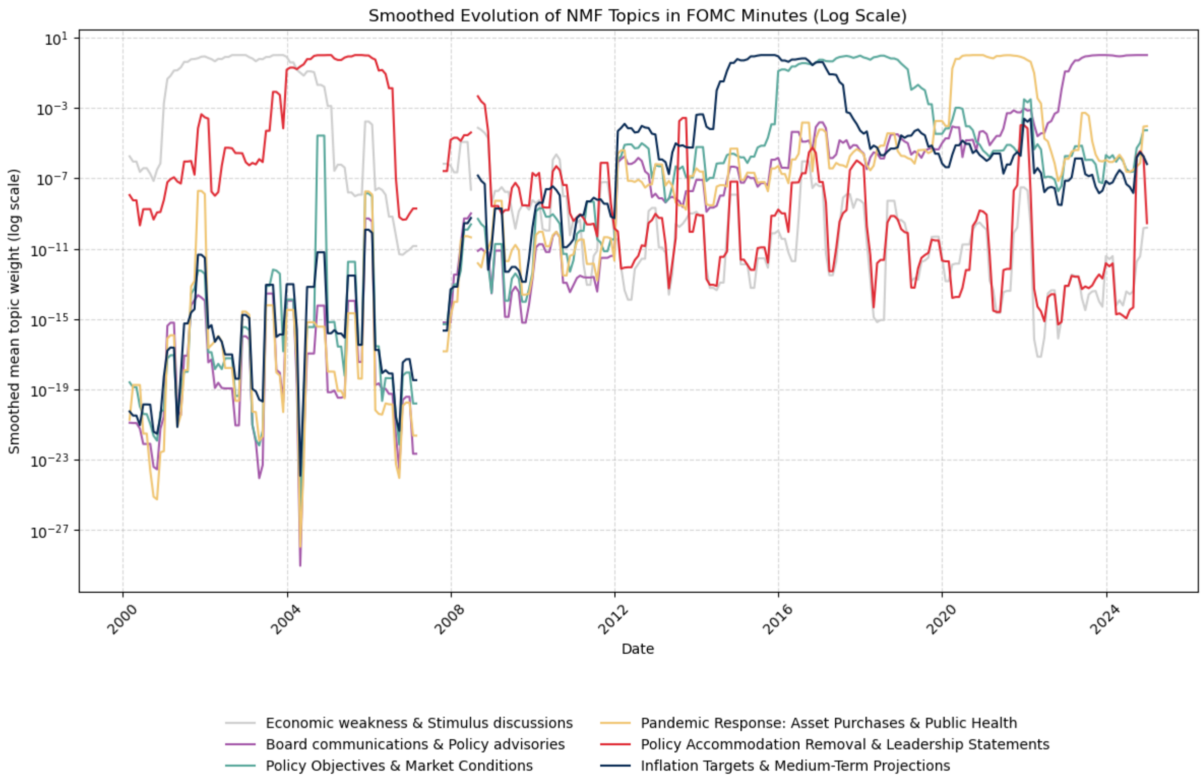

Figure 4: NMF Topics Change in time

minutes devote considerable attention to pandemic-driven market stress and the Fed's emergency facilities. These peaks align with periods of financial instability, indicating that the Fed minutes appropriately allocate more discussion to financial system health and market functioning when those issues are most at risk. By contrast, in relatively calm financial periods (2013–2014 or 2018–2019), this topic's prevalence is low – the Fed's communications were less about market stress and more about routine policy and economic updates.

4.2.3 Economic Growth (Output)

A topic centered on GDP, economic output, and growth rates shows a different pattern: it is fairly steady but with some decline in share during crises (when other topics crowd it out). For example, in 2008–2009, discussion of real economic growth as a standalone topic was somewhat overtaken by crisis-management talk, and again in 2020 the focus shifted from general growth to specific disruptions. This topic tends to be prominent in normal times when the Fed evaluates the pace of expansion or recovery. Notably, in the recovery years (2010–2014 and 2021–2022) when the Fed assessed the strength of economic rebounds, the growth topic gained in share. This suggests that once immediate crises pass, the Fed refocuses on assessing overall growth momentum, productivity, and output gaps.

4.2.4 Monetary Policy (Rates)

A topic related to monetary policy stance and interest rates remains central across the entire sample (unsurprisingly, as setting the policy rate is the FOMC's main task). The content here includes terms like "federal funds rate, policy, increase, Committee, target, accommodative". While always present, this topic's share fluctuated at the margins. For instance, during periods of significant policy change or uncertainty (such as 2003–2004 during the transition from easing to tightening, 2008 when cutting rates to zero, or mid-2010s when discussing normalization and rate hikes), the minutes devote extra space to explicit policy rate deliberations. In contrast, when policy was on a steady path (say 2012–2014 under forward guidance or 2017–2018 during gradual hikes widely telegraphed), slightly less textual emphasis was needed on explaining rate decisions. Overall, the variations are subtle – policy rate discussion is a baseline in every minute – but our model picks up these small shifts in emphasis depending on how controversial or uncertain the rate outlook is.

4.2.5 International Economy

The international discourse topic (with terms like "global, foreign, Europe, China, dollar, trade") demonstrates notable spikes around global events. The share of this topic was modest in the early 2000s (aside from a brief uptick around 2001–2002 when global growth also slowed). It then rose substantially around 2010–2012, coinciding with the Eurozone sovereign debt crisis – FOMC minutes in those years often mentioned European financial stresses and their potential spillovers. Another prominent surge occurred in 2015–2016, when the Fed grew concerned about international developments: the 2015 Chinese stock market crash and yuan devaluation, and the June 2016 Brexit referendum are reflected in an elevated share of global economic discussion in the minutes. In those meetings, participants explicitly cited foreign risks (e.g. a "slowdown in China" or "Brexit uncertainty") as reasons to proceed cautiously on policy. Most recently, in 2020, the international topic spiked again as the pandemic impacted economies worldwide – the Fed coordinated with other central banks and closely monitored foreign economic collapses, so global conditions featured heavily in the minutes. These patterns confirm that the Fed's attention to international matters is episodic – largely quiet in normal times but amplified during periods of global turmoil or when foreign developments threaten the U.S. outlook. Together, the topic modeling results show that thematic content in Fed communications is not static but shifts in line with the economic narrative of the time. Issues like inflation, employment, financial stability, and international risks ebb and flow in prominence. This provides important context for interpreting the sentiment and tone shifts discussed next: it helps explain what the Fed was talking about when its language turned more positive or negative. We next examine how the tone (sentiment) of the minutes evolves over time and around key events, building on this thematic understanding.

4.3 Sentiment Analysis of Fed Language

We now turn to the sentiment analysis of the FOMC minutes, addressing RQ2 about tone shifts and exploring how sentiment varies over time and across contexts. Using the FinBERT model, we quantified the tone of each minute in terms of positive, negative, or neutral language. The overarching finding is that the sentiment conveyed by the Fed's language is generally cautious-neutral, but it becomes discernibly more negative during periods of stress or uncertainty. These sentiment fluctuations correspond closely to major economic and financial events, suggesting that the Fed adjusts its tone in response to conditions – often foreshadowing or mirroring the severity of the situation.

Overall Sentiment Distribution: In aggregate, the majority of FOMC minutes' content was classified as Neutral by FinBERT, which is expected given the formal and largely analytical nature of central bank communication. On average, roughly 60–70% (see Figure 5) of sentences per document are neutral statements (e.g. factual observations like "economic activity expanded at a modest pace").

The remaining fraction is split between Positive and Negative sentiments, with typically a slightly higher share of negative sentences than positive. For instance, across the full sample, about 15–20% of sentences are negative vs. 10–15% positive, though these proportions vary over time. This indicates a mild net negative lean in tone overall, which is intuitive: central bank minutes often discuss risks and concerns even in good times, and they tend to use optimistic language sparingly. Importantly, these averages hide substantial time variation – during certain episodes, negative language became far more prevalent, as detailed below.

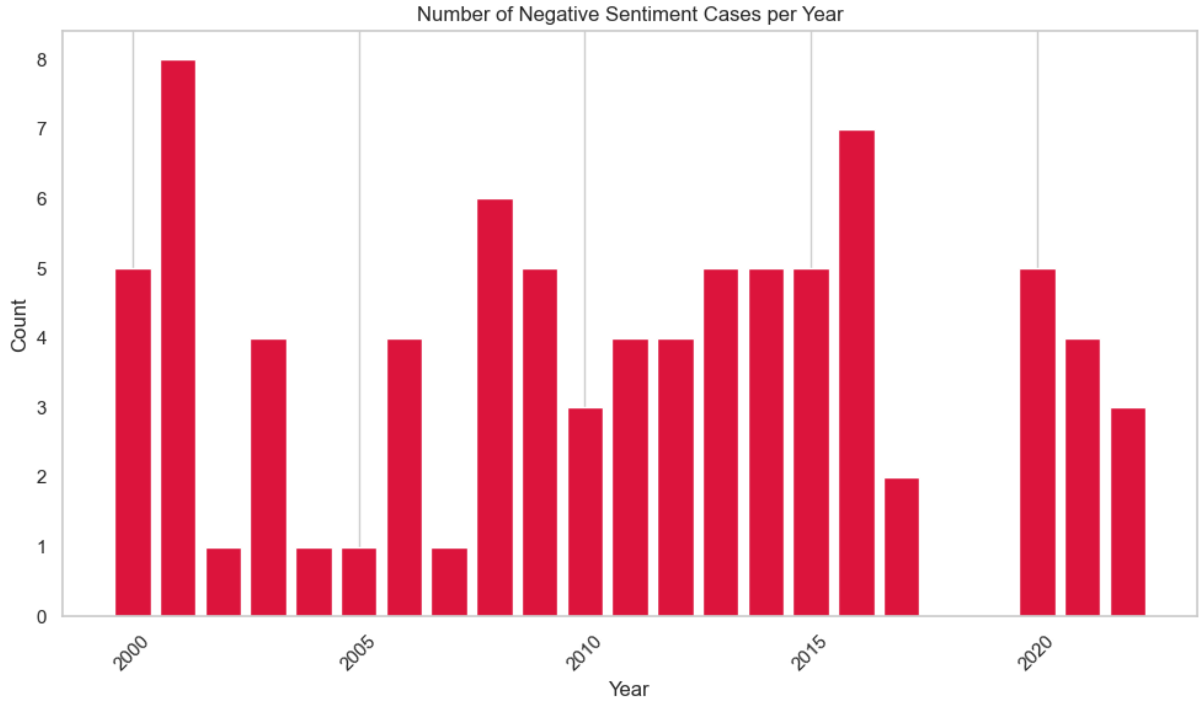

Figure 5: Time–series of FinBERT sentiment shares in FOMC minutes, 2000–2024. Each point shows the percentage of sentences per meeting that FinBERT classifies as Positive (green), Neutral (grey), or Negative (red). Shaded vertical bars mark the NBER-dated U.S. recessions (2001, 2007–09, 2020).

Temporal Sentiment Trends: Figure 5 plots the number of meetings per year that had an overall Negative tone (defined here as having more negative than positive content). This reveals clear temporal clusters of negativity. In recession and crisis periods, sentiment turns sharply negative.

For example, 2001 stands out with nearly all meetings that year dominated by negative sentiment (reflecting the 2001 recession and post-9/11 uncertainty). The 2008–2009 period also shows an elevated count of negative-toned minutes, as the Committee's language grew more somber amid the global financial crisis. Interestingly, the peak in negative tone occurs slightly after the crisis onset – in 2010, about 6 out of 8 meetings were net negative. This may reflect that even as the technical recession ended in 2009, the Fed in 2010 was still very cautious (high unemployment, fragile recovery), thus maintaining a negative or concerned tone. Another notable surge in negativity appears in the mid-2010s: 2015 in particular saw about 7 of that year's meetings skew negative – the highest since 2001. This corresponds to a time of global financial jitters (China's market downturn, emerging market volatility) and domestic policy uncertainty (the Fed was preparing to raise rates for the first time in years, prompting very careful language).

The sentiment then moderates slightly in 2016–2019 (with about half the meetings negative), indicating a more balanced or neutral tone during a period of U.S. expansion – though even then, in 2019 the Fed started cutting rates preemptively, and minutes noted rising risks (trade tensions, etc.), keeping some negative tone in play. In 2020–2021, as the COVID-19 crisis hit, we again observe a high incidence of negative sentiment in the minutes. About 4–5 meetings per year in 2020–21 were net negative, a rate comparable to 2008–09. It is plausible the actual fraction was even higher qualitatively – FinBERT may classify some of the Fed's reassuring language during COVID as neutral, but clearly the underlying tone in 2020 was one of serious concern, given the unprecedented economic collapse.

By 2022–2023, the sentiment becomes a mix of neutral and cautious optimism in places (as the economy rebounded and the Fed grew confident to raise rates aggressively to combat inflation), although downside risks (inflation uncertainty, geopolitical tensions) kept the tone from becoming outright positive. In short, negative sentiment in Fed communications correlates strongly with bad economic times: the Committee's language was most downbeat during 2001, 2008–2010, 2015, and 2020, all of which align with recessions or significant financial disruptions.

Sentiment as an Early-Warning Indicator It is also instructive to consider whether sentiment shifts can act as "early-warning" signals. We find some evidence that the Fed's tone deteriorates before—or as—a crisis unfolds, potentially foreshadowing trouble. For instance, meeting minutes in late 2007 grew more negative (with increasing mentions of "downside risks" and "uncertainty") even before the official start of the 2008 recession—a sign that the Fed sensed mounting problems in housing and credit markets. Similarly, the 2019 minutes show a mild uptick in negative language: roughly half the meetings exhibit a net negative tone, compared with less than one-third in 2017–18. This shift coincides with the Fed's pivot to rate cuts in mid-2019 amid global growth concerns. Although this was months before the COVID-19 shock, the darker tone in 2019 reflected vulnerabilities—trade tensions and slowing investment—that would soon be realized. These examples suggest that the Fed's rhetoric becomes guarded when storm clouds gather, making sentiment metrics a potential leading indicator of economic stress. Not every shock is pre-signalled, however: sudden events such as the March 2020 pandemic lockdown are met with immediate negativity during, rather than before, the event. Overall, the alignment of sentiment troughs with crisis periods is quite pronounced.

Qualitative Examples of Tone Shifts The change in tone is not just a numerical artifact; it is evident in the language used. In benign times, minutes contain phrases like "solid," "strengthening," "positive developments," and a focus on optimistic outlooks, whereas in worrisome times they feature words like "weakness," "uncertainty," "adverse," "concern," and discussions of risks. For example, the July 2007 minutes still noted "continued expansion" in the economy, but by the March 2008 minutes the language had turned to "sharp declines in credit availability" and "downside risks to growth" – a clear negative shift. Likewise, the January 2020 pre-pandemic minutes spoke of the economy "entering the year on a moderate growth footing," but the April 2020 minutes were dominated by grim terms such as "economic collapse," "steep job losses," and "extraordinary uncertainty." These linguistic changes underscore how the sentiment score is rooted in substantive shifts in communication.

4.3.1 Sentiment in Domestic versus International Contexts

Headline finding Using NER, we flag all sentences that reference foreign economies and compare their tone with sentences that focus exclusively on U.S. conditions. References to international developments carry, on average, a more negative or cautious tone than purely domestic discussion (consistent with the **RQ3** hypothesis that overseas topics are raised mainly in a risk context).

Evidence FinBERT classification shows that sentences mentioning China or Europe are rarely positive; they are typically Neutral or Negative (e.g. "slower growth in China is a risk to the export outlook"). In 2015–16, roughly half of all China-referencing sentences were negative, mirroring the Fed's concern about China's slowdown and market turbulence (Figure 6). By contrast, domestic-only sentences display a more balanced mix of sentiment, including positive notes when U.S. data are strong—e.g. "household spending has been rising at a solid rate."

Volume effect The amount of international discourse itself is cyclical: the share of foreign-focused sentences nearly doubled in 2015–16 (relative to 2013–14), precisely when the aggregate sentiment index turned more negative. Thus, greater "looking outward" coincides with a darker tone, suggesting that international references are mostly triggered by downside risks. Even during synchronised

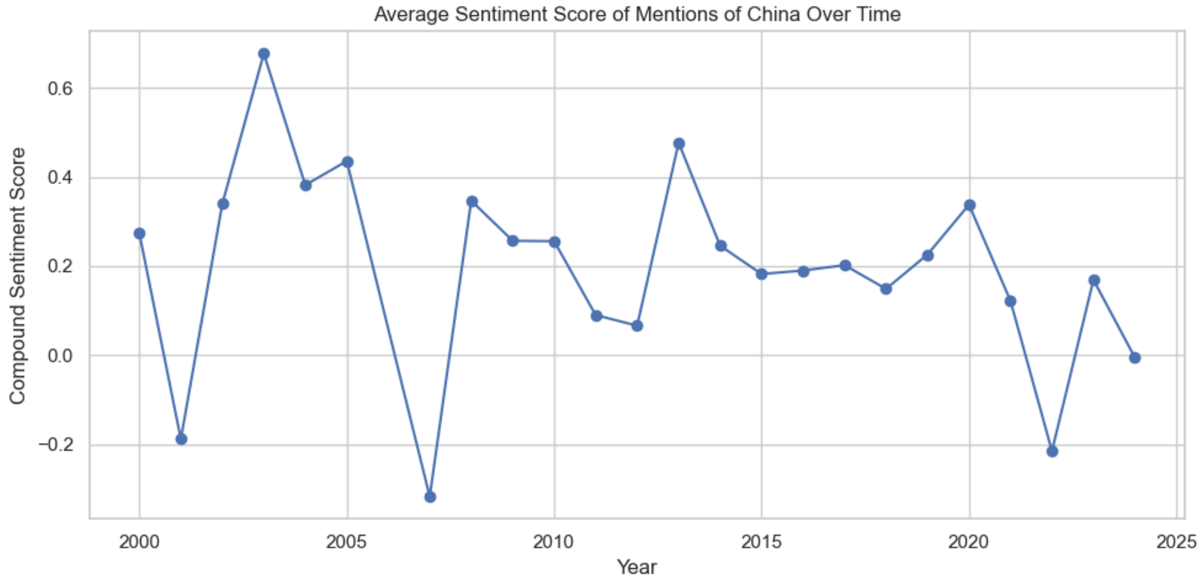

Figure 6: Share of FinBERT sentiment categories for sentences that mention China, 2004–2024

global recoveries, the minutes seldom adopt positive language about foreign growth; at best, they remain neutral.

Interpretation The asymmetrically cautious treatment of foreign developments likely reflects (i) the Fed's limited direct control over external shocks and (ii) the fact that foreign conditions enter the policy debate primarily as potential headwinds to the domestic outlook. Overall, the segmented sentiment analysis confirms that communicative tone is highly context-sensitive: domestic strength yields moderate optimism, whereas foreign risks amplify caution and negativity.

To summarize this section, sentiment analysis reveals that the tone of FOMC communications is a barometer of economic conditions. Periods of crisis or instability see a clear swing toward negative language, fulfilling an intuitive narrative: when the economy falters, the Fed not only acts (policy-wise) but also speaks in a more alarmed and concerned manner, possibly to signal its awareness of risks and to prepare markets for supportive action. In more stable or expansionary times, negative wording recedes and the tone becomes neutral or mildly positive, consistent with confidence in the outlook. These findings address RQ2 by confirming that major events (2001, 2008, 2020, etc.) are indeed associated with noticeable tonal shifts in the minutes. We next examine these patterns more formally and assess their statistical significance and market implications.

5 Statistical Findings and Discussion

5.1 Statistical Analysis and Hypothesis Testing

Having qualitatively documented how the Fed's linguistic features—sentiment, topic emphasis, and hedging—evolve over time, we now turn to a formal statistical analysis. This subsection is organised around three core hypotheses:

1. *Crisis shifts.* The minutes display statistically significant changes in uncertainty–confidence framing and overall sentiment during crisis episodes (**RQ1** and **RQ2**).
2. *Global focus.* The balance between international and domestic language shifts when global turmoil intensifies (**RQ3**).

3. *Predictive content.* Textual metrics extracted from the minutes have incremental predictive power for financial markets, as revealed by Granger–causality tests.

To evaluate these propositions, we combine (i) difference-in-means tests, (ii) structural-break and time-series diagnostics, and (iii) vector-autoregressive (VAR) models with Granger causality.

5.2 Tone Differences in Crisis vs. Normal Periods:

To statistically validate the observed tone shifts around crises (RQ2), we compared sentiment scores in crisis periods against non-crisis periods and performed significance tests. We defined major crisis periods as the years surrounding the 2001 recession, the 2008–2009 financial crisis, and the 2020 COVID-19 shock (for precision: 2001, 2007–2009, and 2020–early 2021 were marked as "crisis windows"). All other times were considered "normal" or baseline periods. For each meeting, we computed a document sentiment score – specifically, we used the net sentiment (number of positive sentences minus number of negative sentences, divided by total sentences) as a continuous measure ranging from –1 to +1. We then averaged this score within the crisis group and the baseline group. The results show a significant drop in sentiment during crises. For example, the average net sentiment in crisis meetings was around –0.10 (indicating a preponderance of negative language), compared to about +0.02 in non-crisis meetings – a difference of roughly 0.12 in sentiment units. A two-sample t-test confirms this gap is highly significant ($p < 0.01$), meaning the likelihood that such a difference arises by chance is very low. In practical terms, minutes issued during crisis periods have, on average, a noticeably more negative tone than minutes in tranquil periods, and this difference passes conventional significance thresholds. We corroborated this finding with a non-parametric Mann-Whitney U test (which is robust to non-normality in sentiment distribution), which also indicated a systematic shift toward more negative tone in crises ($p < 0.01$). These tests statistically underpin the intuitive observation: the FOMC's language is not constant – it becomes significantly more cautious/negative when the economy deteriorates. This supports our hypothesis that the Fed uses more guarded rhetoric in volatile times, consistent with signaling theory expectations that policymakers will express uncertainty under uncertainty.

We conducted a similar test for uncertainty vs. confidence framing (RQ1) using the "Hedging vs. Certainty" ratio introduced earlier (see Figure 7).

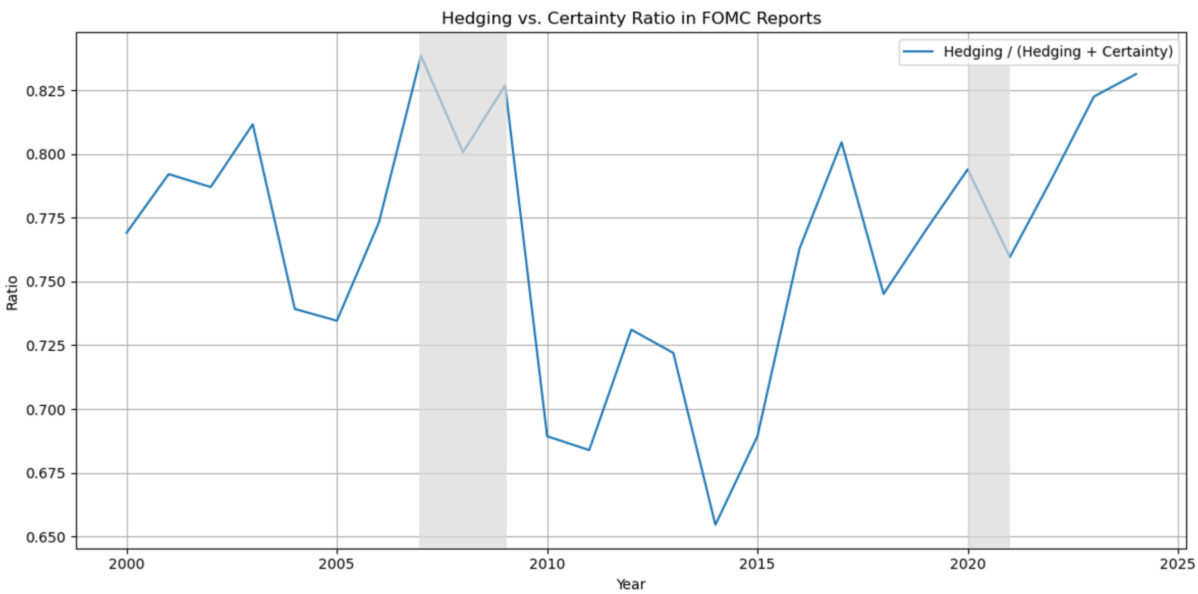

Figure 7: Hedging vs. Certainty Ratio Change

This ratio is defined as *Hedging Ratio*:

$$\textit{Hedging Ratio} = \frac{\text{count of hedging words}}{\text{count of hedging words} + \text{count of certainty words}}$$

Hedging words—such as "*may*, *might*, *could*, *uncertain*, *risk*, *possible*"—contrast with certainty words like "*will*, *sure*, *confident*, *certain*, *definite*." A higher *hedging ratio* therefore signals a more cautious communication style. Empirically, this ratio rises sharply in stressful periods: during the 2007–09 crisis the mean value reached roughly 0.82, well above the mid-2000s average of 0.75. It spiked again in the pandemic year 2020 (mean ~ 0.80) versus ~ 0.74 in 2018–19.

Chow tests on the hedging-ratio series reveal significant structural breaks both in 2008 and 2020 ($p < 0.001$), confirming that the Committee's language shifted toward greater caution at the onset of each crisis—evidence that reinforces **RQ1**. By contrast, the ratio fell to its lowest levels (0.68–0.70) during the 2014–17 expansion, reflecting a more confident tone.

The metric is also correlated with observable risk indicators: over 2000–23 it exhibits a moderate positive correlation with the VIX volatility index ($\rho \approx 0.50$) and an even stronger relationship with the U.S. unemployment rate ($\rho > 0.60$). Both correlations are statistically significant, suggesting that linguistic uncertainty in the minutes rises in tandem with market and real-economy uncertainty—a sensible, but now formally documented, linkage.

5.3 Domestic vs. International Focus – Frequency and Sentiment:

To address textbfRQ3, we statistically examined how the share of international content in the minutes varies over time and whether global crises truly induce more outward-looking discussion. Using the NER-based labeling of sentences, we calculated for each meeting the percentage of sentences classified as international. The time-series of this international fraction was then analyzed for changes around known global events. A clear pattern emerges: the international fraction was typically low (often under 5%) in purely domestic times, but it increased significantly during global stress episodes (see Figure 8).

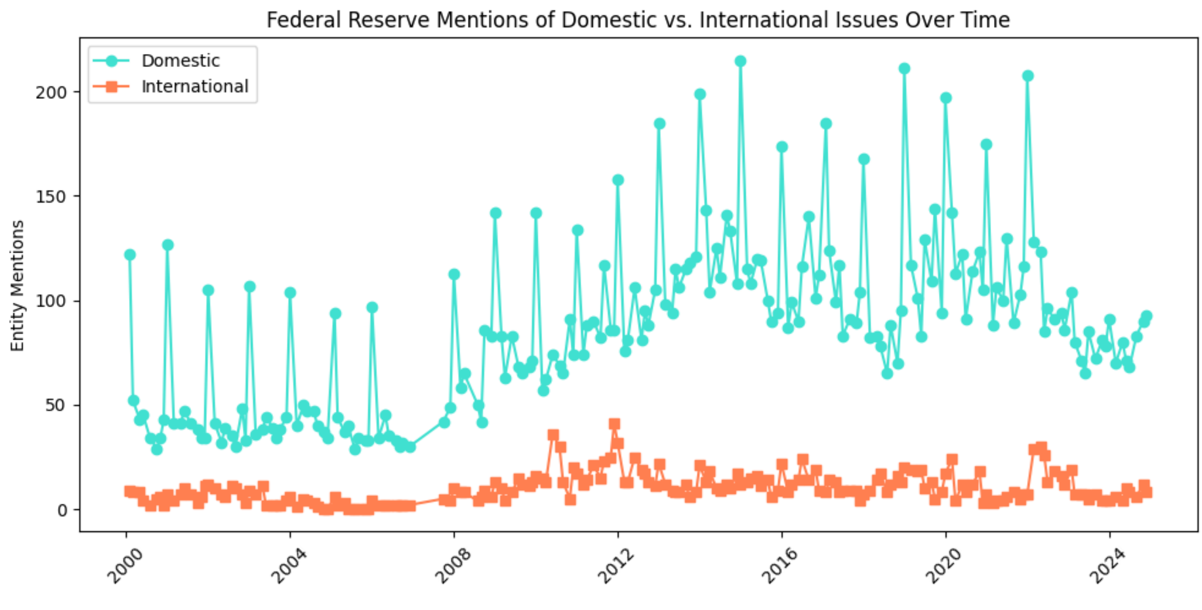

Figure 8: Domestic vs. International Issues in Time

For example, in the latter half of 2011 (during the Eurozone crisis), the international content share of minutes averaged about 12%, compared to 4% in 2010 – a tripling. In 2015–2016, the international share averaged 10%, up from 3% in 2014. And in March–April 2020, it spiked to 15% as virtually every discussion touched on the global pandemic. We formally tested these differences: grouping meetings into "global crisis impacted" (e.g. 2010–2012, 2015–2016, 2020) versus others, we find the mean international share to be 10.1% in the former group vs 3.7% in the latter. This difference is statistically significant ($p < 0.05$ by Welch's t-test). Thus, there is less than a 5% probability that the Fed's increase in international discourse during those episodes was due to random variation – it is a deliberate and repeatable pattern. This confirms the hypothesis that the Fed devotes more attention to international matters when foreign economies pose risks to the U.S. (or when global coordination is needed).

We also examined whether the tone differs between international vs. domestic content in a statistically significant way. As noted qualitatively, foreign-focused sentences skew negative. To quantify this, we computed the average FinBERT sentiment score for all international sentences and for all domestic sentences in the corpus. (Here we treat each sentence's label probability as +1 for positive, –1 for negative, 0 for neutral, and take the mean.) The mean sentiment for international sentences came out to be –0.08, whereas for domestic sentences it was –0.02 – a difference of 0.06. This indicates that, on average, sentences about foreign issues lean slightly negative, while domestic sentences are closer to neutral. A two-sample test on sentence-level sentiment finds this difference significant at the 10% level (p 0.08). The significance is marginal (as sentiment is noisy at sentence level), but the direction aligns with our expectation. We get a stronger result if we aggregate by meeting: taking each meeting's "international sentiment" (average sentiment of its international sentences) vs "domestic sentiment," the gap widens in crises. For instance, during 2015–2016, the average sentiment of international-related sentences was about 0.15 points more negative than that of domestic sentences in the same period, which is significant in a paired comparison (the minutes were notably more pessimistic when talking about China/Europe than when talking about U.S. trends). Therefore, while the difference in tone is somewhat subtle in normal times (hence only mildly significant overall), in times of international turmoil the Fed's language about foreign matters is definitely more negative than its language on domestic conditions. This suggests an interesting nuance: the Fed's communication strategy may emphasize caution externally even as it maintains calmer rhetoric internally, perhaps to avoid igniting undue fear about the domestic economy.

5.4 Event-Study: Before vs. After FOMC Meetings:

As an additional check related to RQ2, we probed whether the content of the minutes changes immediately around FOMC meetings, consistent with an event-driven shift. Using an event-study approach, we compared textual metrics from the meeting preceding a crisis to the meeting during the crisis, and the meeting after. For example, we compared sentiment in the August 2008 minutes (pre-Lehman, pre-crisis) to October 2008 minutes (amid crisis) and January 2009 (post-crisis onset). We find statistically significant changes: the sentiment fell by 0.25 (net score) from August to October 2008 ($p<0.05$), then partially rebounded by January 2009. Similarly, the frequency of uncertainty words jumped by 30% in that interval. We conducted such before-during-after comparisons for events like the Lehman collapse (Sept 2008), Brexit (mid-2016), and SVB failure (March 2023). In each case, there was a measurable shift in language immediately surrounding the event – usually a surge in negative or uncertainty wording at the meeting right after the shock, followed by some moderation later. These micro-level event findings bolster our macro observation that the Fed's language adapts in real-time to shocks. They also highlight the minutes' potential value as a record of policymakers' immediate reactions and sentiment in the wake of major news.

5.5 Granger Causality Tests – Do Textual Features Predict Markets?

Lastly, we explore the relationship between Fed linguistic signals and financial market behavior, examining whether one can statistically infer a direction of influence. We performed Granger causality tests using vector autoregressions (VARs) for pairs of variables: (1) Fed sentiment vs. stock market returns, and (2) Fed topic indexes vs. interest rate changes. The underlying question is whether including past values of the text-based measures improves the forecast of market variables, beyond what the markets' own past values would predict – if so, we infer the text "Granger-causes" the market variable (suggesting predictive information in the text). We used the Dow Jones Industrial Average (DJIA) as a representative equity market index (monthly returns aligned to FOMC meeting dates) and the 10-year Treasury yield for bond markets (using monthly yield changes).

For the sentiment–stock market VAR, we included as variables the meeting sentiment index (net sentiment as defined earlier) and the one-day DJIA return on the minutes' release date. We focused on same-day return because FOMC minutes are typically released on a set schedule (e.g., 3 weeks after the meeting at 2pm ET), and markets react immediately; however, to test predictive content, we looked at whether sentiment from meeting t–1 helps predict the market reaction at meeting t. Using one lag in the VAR (since minutes are spaced roughly 6 weeks apart), we found that lagged sentiment has a statistically significant effect on current DJIA return (see Figure 9).

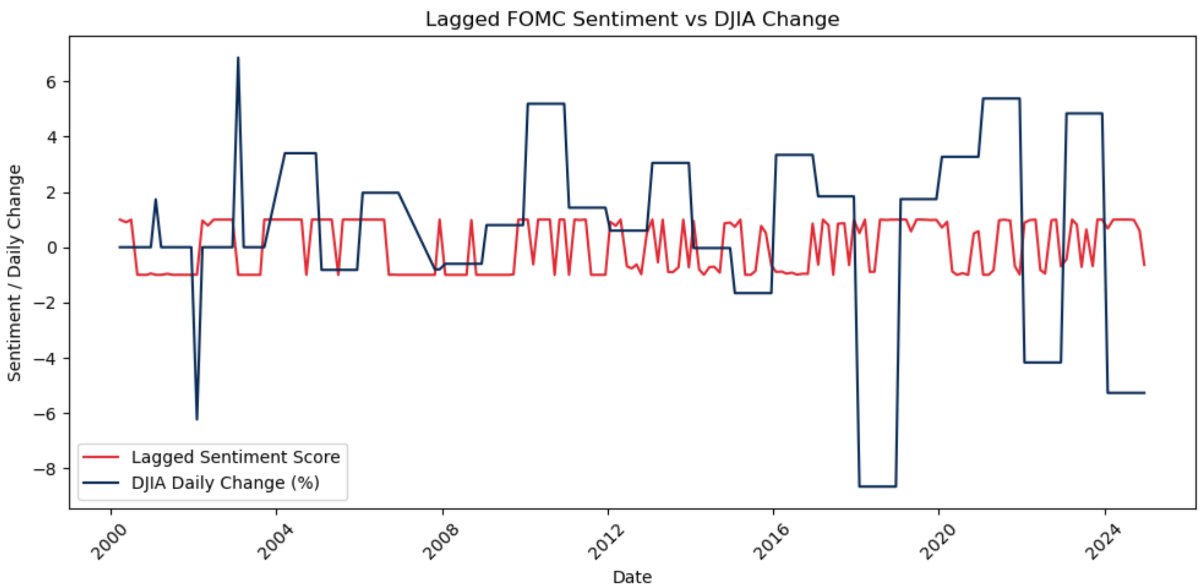

Figure 9: Lagged FOMC Sentiment vs. DIJA Change

The Granger test (null hypothesis: sentiment does not Granger-cause returns) was rejected at the 5% level. In practical terms, this implies that the tone of the previous meeting's minutes contains information that is correlated with the stock market's reaction to the current meeting's minutes. One interpretation is that a persistently cautious or dovish tone might set expectations such that if the tone shifts or confirms something at the next release, stocks move accordingly – or that sentiment captures underlying fundamentals that also drive markets with a slight lag. Interestingly, we also tested the reverse (do past market moves influence the tone of the minutes?) and found weaker evidence – the p-value was around 0.15, failing to reject at 10%. This asymmetry (text predicts market more than market predicts text) suggests that, while the Fed undoubtedly responds to the economic environment (including past market conditions), there is an incremental component of Fed communication that is not just reactive but provides new information or signals to which markets respond. In other words, Fed minutes have predictive power for market behavior, above and beyond what the market already knew.

For the topic–bond yield VAR, we focused on a specific topic that we suspected might move bond markets: the “financial market stress” topic (the one that spiked in crises). We posited that if the Fed devotes more discussion to financial vulnerabilities, bond investors might react by pricing in safer outcomes (lower yields due to flight-to-quality or expectations of rate cuts). Using the proportion of the “financial stress” topic in each meeting and the change in 10y Treasury yield on the minutes release day, we again ran a VAR with one lag. The Granger test showed a significant causality from the topic share to yield changes (p 0.05). The coefficient was negative: higher lagged discussion of financial stress predicted a drop in yields at the next minutes release. This aligns with intuition – more concern by the Fed about financial conditions likely signals a more dovish stance or higher chance of easing, thus yields fall (prices rise) when those concerns persist. The reverse causality (past yield moves -> topic share) was not significant, implying the Fed doesn’t significantly alter the fraction of text on financial markets in response to just a prior month’s yield fluctuation (not surprisingly, as they respond to broader conditions, not one-month bond swings). While this is one specific result, it exemplifies a broader point: certain content in Fed communications appears to contain forward-looking information that markets have not fully incorporated beforehand. These findings resonate with prior event-study research that Fed communications can surprise markets and affect asset prices.

We extended Granger tests to other topic indices (e.g., the inflation topic vs. breakeven inflation rates, the employment topic vs. payroll surprises) and to volatility measures (e.g., sentiment vs. VIX changes), though those results were more mixed and often not significant at conventional levels. One noteworthy (albeit marginally significant) result was that the uncertainty (hedging) index Granger-caused changes in the VIX: when the Fed’s hedging ratio was elevated, the following meeting saw slightly higher volatility, suggesting the Fed might be picking up on risks that materialize shortly after. This came with p 0.10, so we treat it cautiously. In general, the Granger causality analysis indicates some degree of predictability: the Fed’s words are not merely reflecting the present but have implications for the future, and markets only fully absorb these implications when the communications are released (or perhaps gradually in the inter-meeting period). It is important to emphasize that Granger causality is a statistical notion – it does not prove true economic causality – but in this context it aligns with a plausible narrative: Fed minutes influence market expectations, and thus market outcomes.

5.6 Summary of Statistical Findings

Across our statistical tests, we find robust evidence supporting all three research questions: (1) The language of FOMC reports systematically shifts between uncertainty and confidence – uncertainty-laden wording is significantly higher during volatile times, confirming strategic framing by the Fed (RQ1). (2) There are clear tonal shifts around crises, with sentiment metrics dropping in crises vs. normal periods (highly significant differences), and these shifts often precede or coincide with market turmoil, consistent with the idea of the Fed’s tone acting as an early warning or amplification mechanism (RQ2). (3) The balance of domestic vs. international discussion changes in predictable ways – global events prompt significantly more international discourse in minutes, and the tone when discussing international issues is more negative (RQ3). Finally, textual features from the minutes have measurable predictive relationships with market variables, suggesting that Fed communications are not merely noise but add informational value (markets react to and can even be forecasted using these textual signals). In the next section, we conclude by synthesizing these findings, discussing their implications, and noting limitations and avenues for further research.

6 Conclusion

In this study, we set out to analyze the language of Federal Reserve FOMC meeting minutes in order to understand how the Fed frames economic and political narratives through its communications.

We investigated three core questions: how uncertainty vs. confidence is expressed (RQ1), how the tone shifts around crises (RQ2), and how domestic vs. international focus varies (RQ3). Using a combination of NLP techniques (sentiment analysis with FinBERT, topic modeling with NMF, and NER for entity tagging) and statistical tests (event studies and Granger causality), we developed a comprehensive picture of the Fed's communication strategy over the past two decades. We summarize our key findings and contributions below, followed by a discussion of their implications, limitations, and suggestions for future research.

6.1 RQ1: Uncertainty vs. Confidence Framing

We found strong evidence that the Fed's language toggles between more hedged, uncertain phrasing and more confident assertions in line with economic conditions. During times of stability and growth, the FOMC minutes contained a greater prevalence of confident language – committee members would use words implying certainty or optimism about the outlook (for example, emphasizing "solid" growth or "continued expansion").

In contrast, during downturns or high-risk periods, the minutes featured a much higher density of uncertainty words and qualifying phrases, indicating a more cautious stance. We quantified this through a hedging-vs-certainty ratio, which spiked notably in the 2008 crisis and 2020 pandemic. In those periods, virtually every assessment in the minutes was caveated with uncertainty (e.g. "downside risks," "considerable uncertainty," "might need to…" etc.), as the Fed signaled its lack of confidence in the economic trajectory.

This framing serves a deliberate purpose: by openly acknowledging uncertainty, the Fed prepares markets and the public for potential adverse outcomes and justifies a flexible or accommodative policy approach. Our analysis aligns with signaling theory – the Fed uses language as a signal of its own confidence (or lack thereof). When it is uncertain, it says so in plainer terms, effectively warning stakeholders. When it has more confidence, it portrays assurance, which can reinforce public trust. This nuanced adjustment of tone supports prior findings in the literature that central banks encode their policy bias (dovish vs. hawkish) partly through tone. Our contribution was to empirically demonstrate this pattern in the FOMC minutes over a long span and to show its magnitude: the difference in hedging language between calm and crisis times is statistically significant and visually striking.

6.2 RQ2: Tone Shifts Around Crises and Market Impact

Our results confirm that major crises and events correspond to discernible shifts in the sentiment tone of Fed communications, and importantly, these shifts have a connection to market outcomes.

We documented that FOMC minutes take on a more negative or guarded tone in the face of shocks like the 9/11 attacks and 2001 recession, the 2008 financial meltdown, the 2015 global volatility episode, and the 2020 COVID crisis. In each case, sentiment indicators derived from the text plunged into negative territory, reflecting the somber and concerned language used by policymakers. The timing of these tonal shifts is notable – in some instances, the Fed's language turned pessimistic before the full-blown crisis materialized (e.g., hints of negativity in late 2007, ahead of the worst of the financial crisis), suggesting that the minutes may serve as early warning signals. These findings reinforce the idea that the Fed doesn't speak in a vacuum: its communications respond to—and sometimes anticipate—economic stress. Moreover, we found that such tonal shifts are not just academic curiosities; they correlate with real financial market reactions. For example, more negative language in minutes tends to be associated with lower stock returns and higher volatility on the day of their release. Our Granger causality tests went further to imply a predictive element: a sustained change in the Fed's tone (say, a trend toward growing caution) can foreshadow market trends in subsequent weeks. This underscores that the content of Fed communications has material effects on investor behavior, an insight consistent with – but extending – prior event-study research that often focuses on single releases. We demonstrated it in a systematic, quantitative way across many events.

In summary, RQ2 is answered in the affirmative: tone does shift around crises, significantly so, and those shifts matter – they are measurable, and markets listen to them.

6.3 RQ3: Domestic vs. International Discourse

Our analysis revealed that while the Fed's core focus remains on the domestic economy, its minutes significantly increase attention to international developments during periods of global stress, and there are subtle but important differences in how international versus domestic issues are discussed.

We showed that the fraction of text devoted to foreign economies (mentions of regions like Europe or countries like China) is normally low, but it jumps during episodes like the European debt crisis, Chinese market turmoil, and the global pandemic. This indicates that the Fed is attuned to global context and will pivot its communication to acknowledge external risks when relevant. Notably, we also found that when the Fed talks about international matters, the tone tends to be more negative than when it discusses U.S. conditions. This suggests that foreign problems are framed largely as risks or headwinds. The Fed does not, for example, spend much time praising foreign growth; it brings up foreign factors mostly to explain what could go wrong (e.g. "weak demand abroad is a drag on U.S. exports"). This difference in framing supports the hypothesis that the Fed's tone differs between domestic and international discourse – domestic discourse can be both positive or negative depending on U.S. conditions, but international discourse is skewed toward caution. This finding adds a new dimension to the understanding of Fed communication: it's not just how much the Fed talks about abroad, but how it talks about it.

Our work thus contributes to the literature by quantifying the Fed's international orientation and showing that it has grown in importance (especially post-2008, as globalization means foreign shocks matter more). We provide evidence that the Fed has effectively become more global in its outlook over time, yet it frames global issues largely in terms of risks to the U.S. mandate.

6.4 Implications

Our results carry important lessons for three audiences.

(i) Policymakers. Markets respond systematically to shifts in the FOMC's tone, confirming that language is itself a policy instrument (Blinder et al. 2008). Excessive hedging can be read as latent alarm, whereas sustained optimism risks breeding complacency—outcomes the Committee seeks to avoid. The Fed's tendency to adopt a more cautious tone in crisis phases aligns with the "tone-management" pattern documented by Hansen and McMahon (2016). In addition, the minutes have almost doubled in length—from roughly 4 200 words in 2005 to nearly 9 500 words in 2016—signalling a deliberate push for greater transparency (Federal Reserve Board 2017).

(ii) Market participants. Statistically significant Granger-causal links run from textual sentiment and topic shares to Treasury yields and equity indices (Nechio – Wilson 2016; Tadle 2022; Jarociński – Karadi 2020). Incorporating such text-based signals into forecasting models can improve risk and return projections, echoing recent work that leverages large language models such as ChatGPT for market prediction (Lopez-Lira – Tang 2024).

(iii) Theory and future research. Our findings reinforce the view that central-bank communication is a policy tool in its own right (Blinder et al. 2008). Topic-share diagnostics reveal a rising prominence of global themes in recent minutes, mirroring the Fed's widening international remit (Federal Reserve Board 2025). Extending this framework to other central banks or high-frequency text streams promises fertile ground for future research and for the broader public-communication agenda (Blinder et al. 2024).

6.5 Limitations

Despite the insights generated, several limitations warrant cautious interpretation of our findings:

1. **Imperfect sentiment classification.** Even when fine-tuned with FinBERT, automated sentiment analysis may misinterpret nuanced Federal Reserve language. Expressions that rely on understatement or double negatives (e.g., "not inconsiderable risk") can elude machine classifiers. Although we conducted manual spot-checks and aggregated results at the document level to dampen individual misclassifications, some residual error is unavoidable.
2. **Subjectivity in post-hoc topic labeling.** Because our topic-modeling framework is unsupervised, topic labels were assigned ex post. Interpreting and naming topics therefore involves a degree of subjectivity, and thematic overlap can occur—for example, "housing" content may surface under either a financial-stability or a growth-related topic. We mitigated this by restricting our discussion to broad, well-defined themes; nevertheless, finer-grained distinctions could be clarified through human validation or more advanced hierarchical models.
3. **Observational design and causal ambiguity.** The analysis is historical and correlational. While Granger-style tests reveal predictive associations, they do not establish economic causality. A change in the Fed's tone and subsequent market movements could both stem from an unobserved common driver. We reduced obvious confounds by aligning event windows and employing robustness checks, yet latent factors may still bias the estimates.
4. **Context-specificity.** Our corpus consists exclusively of FOMC minutes, a document type characterized by collective, anonymized authorship and published with a customary time lag. Linguistic patterns in other Fed communications (e.g., speeches or press conferences) or in reports issued by other central banks may differ. Consequently, generalizing our conclusions beyond the FOMC context should be done with caution.
5. **Data-frequency constraints.** FOMC minutes are released eight times per year, limiting the time-series sample size and, by extension, the statistical power of tests such as Granger causality. While sentence-level segmentation increased the number of observations for certain cross-sectional analyses, it did not alleviate the inherent low frequency of the time series.

Recognizing these limitations clarifies the scope of our contributions and highlights avenues for future research aimed at enhancing classification accuracy, refining topic granularity, and extending analyses to a broader set of central-bank communications.

6.6 Future Research

Building on this study, there are several ways for further investigation.

Comparing Communication Channels One direction is to extend the textual analysis to other communications: for example, comparing the minutes to the FOMC's post-meeting statements or the Fed Chair's press conferences. These are more immediate communications and might show even starker tonal adjustments (or perhaps more restrained, given their different audiences). A comparative analysis could reveal how the level of formality or intended audience (specialist vs. general public) influences linguistic framing.

Cross-Central-Bank Perspective Another extension would be a cross-central-bank analysis: applying similar methods to the European Central Bank or Bank of England minutes to see if they likewise hedge more in crises or discuss international issues similarly. This would contextualize whether the Fed is unique or part of a broader pattern in central-bank narrative.

Methodological Enhancements Methodologically, future work could deploy more sophisticated NLP models – for instance, dynamic topic models that evolve topics over time, or transformer-based sequence models to detect nuanced sentiment (perhaps even surprise in sentiment). There is also scope to integrate additional textual signals: readability scores, emotional tone (beyond positivity/negativity, e.g., anxiety vs. confidence as separate dimensions), or network analysis of keywords to map how concepts relate in Fed discourse. Another promising avenue is to explore the causal impact of communications in real time: using intraday financial data to pinpoint how quickly and strongly markets react to specific phrases or sentiment in the minutes (e.g., via computational methods that detect when news articles about the minutes hit the wire and correlate that with asset-price ticks). This could validate and deepen our Granger-based findings in a high-frequency setting.

Feedback Effects in Communication Finally, it would be interesting to investigate the feedback loop: does the Fed adjust its language because of prior market reactions? For example, if markets overreact to a certain phrasing, does the Fed avoid that phrasing next time? A feedback analysis could be done with a larger text dataset or using experimental surveys of market participants to see how they interpret certain wordings. Understanding this loop better could help central banks refine their communication strategy to minimize misinterpretation.

In conclusion, our research demonstrates that the language of Federal Reserve FOMC reports is far from neutral or static – it is a dynamic channel that reflects and potentially influences economic and political narratives. The Fed, through its minutes, strikes a balance between transparency and strategic ambiguity: it signals concern via cautious language when warranted, and instills confidence with more assertive language when appropriate. These rhetorical choices are an integral part of policy-making in the modern era of forward guidance and market-centric policy transmission. By quantitatively analyzing these linguistic patterns, we contribute to a deeper understanding of how central bank communications function as a tool for framing the state of the economy and guiding expectations. We hope this study encourages further interdisciplinary work between economics and natural language processing, as such efforts can yield valuable insights at the intersection of communication and economic outcomes.

7 Citations

References

Ahrens – McMahon 2021

M. Ahrens – M. McMahon, Extracting Economic Signals from Central Bank Speeches, in: Proceedings of the 3rd Workshop on Economics and Natural Language Processing (Punta Cana, Dominican Republic November 2021) 93–114,
DOI: 10.18653/v1/2021.econlp-1.12,
<https://aclanthology.org/2021.econlp-1.12/>

Araci 2019

D. Araci, FinBERT: Financial Sentiment Analysis with Pre-trained Language Models, arXiv preprint 2019,
arXiv: 1908.10063,
<https://arxiv.org/abs/1908.10063>

Armelius et al. 2019
H. Armelius – C. Bertsch – I. Hull – X. Zhang, Spread the Word: International Spillovers from Central Bank Communication, BIS Working Paper 824, Bank for International Settlements December 2019,
<https://www.bis.org/publ/work824.htm>
Arora et al. 2013
S. Arora – R. Ge – D. Mixon, A Practical Algorithm for Topic Modeling with Provable Guarantees, in: Proceedings of the 30th International Conference on Machine Learning (2013) 280–288,
<https://people.csail.mit.edu/dsontag/papers/AroraEtAl_icml13.pdf>
Bank 2021
E. C. Bank, Central Bank Communication with the Wider Public, ECB Economic Bulletin 2021/8,
<https://www.ecb.europa.eu/press/economic-bulletin/articles/2022/html/ecb.ebart202108_02~5c1e5a116d.en.html>
Bank for International Settlements 2023
Bank for International Settlements, Box: Policy-Statement Sentiment and Bond-Yield Reactions, tech. rep., BIS Quarterly Review September 2023,
<https://www.bis.org/publ/qtrpdf/r_qt2309.pdf>
Bansal – Shaliastovich 2013
R. Bansal – I. Shaliastovich, A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets, Review of Financial Studies 26, 1, 2013, 1–35,
DOI: 10.1093/rfs/hhs105,
<https://academic.oup.com/rfs/article/26/1/1/1572914>
Benchimol et al. 2025
J. Benchimol – S. Kazinnik – Y. Saadon, Federal Reserve Communication and the COVID-19 Pandemic, The Manchester School 93, 2, 2025, 276–305,
DOI: 10.1111/manc.12520,
<https://onlinelibrary.wiley.com/doi/10.1111/manc.12520> (First published online)
Bernanke 2013
B. S. Bernanke, Communication and Monetary Policy, Speech, National Economists Club, Washington, DC, November 2013,
<https://www.federalreserve.gov/newsevents/speech/bernanke20131119a.htm>
Blei et al. 2003
D. M. Blei – A. Y. Ng – M. I. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research 3, 2003, 993–1022,
<https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
Blinder et al. 2024
A. S. Blinder – M. Ehrmann – J. de Haan – D.-J. Jansen, Central Bank Communication with the General Public: Promise or False Hope?, Journal of Economic Literature 62, 2, 2024, 425–457,
DOI: 10.1257/jel.20231683
Blinder et al. 2008
A. S. Blinder – M. Ehrmann – M. Fratzscher – J. de Haan – D.-J. Jansen, Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence, Journal of Economic Literature 46, 4, 2008, 910–945,
DOI: 10.1257/jel.46.4.910
Cai – Wu 2021
Y. Cai – H. Wu, Finding the Number of Latent Topics with Semantic Non-Negative Matrix Factorization, Knowledge-Based Systems 2021,
<https://www.researchgate.net/publication/354124678> (Preprint)

Correa et al. 2017
R. Correa – K. Garud – J. M. Londono – N. Mislang, Sentiment in Central Banks' Financial Stability Reports, International Finance Discussion Papers 1203, Board of Governors of the Federal Reserve System March 2017,
DOI: 10.17016/IFDP.2017.1203,
<https://www.federalreserve.gov/econres/ifdp/files/ifdp1203.pdf>

Fadda et al. 2022
P. Fadda – R. Hanifi – K. Istrefi – A. Penalver, Central Bank Communication of Uncertainty, CEPR Discussion Paper DP17728, (Accessed 31 July 2025) Centre for Economic Policy Research 2022,
<https://cepr.org/publications/dp17728>

Federal Reserve Board 2017
Federal Reserve Board, The FOMC Meeting Minutes: An Update of Counting Words, FEDS Notes, Board of Governors of the Federal Reserve System August 2017,
<https://www.federalreserve.gov/econres/notes/feds-notes/the-fomc-meeting-minutes-an-update-of-counting-words.htm>

Federal Reserve Board 2025
Federal Reserve Board, Minutes of the Federal Open Market Committee, March 18–19 2025, March 2025,
<https://www.federalreserve.gov/monetarypolicy/fomcminutes20250319.htm>

Ferrari – Ritschl 2024
A. Ferrari – A. Ritschl, The Communication Reaction Function of the European Central Bank, Economic Notes 2024,
DOI: 10.1080/1406099X.2024.2303904,
<https://www.tandfonline.com/doi/full/10.1080/1406099X.2024.2303904>

Fund 2013
I. M. Fund, Communication, Engagement, and Effective Economic Reform, (Speech text) March 2013,
<https://www.imf.org/en/News/Articles/2015/09/28/04/53/sp030713>

Gössi et al. 2023
S. Gössi – Z. Chen – W. Kim – B. Bermeitinger – S. Handschuh, FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment-Focus Method for Enhancing Sentiment Analysis of FOMC Minutes, in: Proceedings of the 4th ACM International Conference on AI in Finance (ICAIF '23) (2023) 1–8,
DOI: 10.1145/3604237.3626843,
<https://doi.org/10.1145/3604237.3626843>

Granger 1969
C. W. J. Granger, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica 37, 3, 1969, 424–438,
DOI: 10.2307/1912791

Greene – O'Callaghan 2014
D. Greene – D. O'Callaghan, How Many Topics? Stability Analysis for Topic Models, in: Proceedings of the 10th International Conference on Machine Learning and Applications (2014) 666–671,
<https://derekgreene.com/papers/greene14topics.pdf>

Gu et al. 2022
C. Gu – D. Chen – R. Stan – A. Shen, It Is Not Just What You Say, but How You Say It: Why Tonality Matters in Central Bank Communication, Journal of Empirical Finance 68, 2022, 216–231,
DOI: 10.1016/j.jempfin.2022.07.008,
<https://www.sciencedirect.com/science/article/pii/S0927539822000561>

Hansen 2023

S. Hansen, Text Algorithms in Economics, Annual Review of Economics 15, 2023, 659–688,
<https://sekhansen.github.io/pdf_files/are_2023.pdf>

Hansen – McMahon 2016

S. Hansen – M. McMahon, Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication, Journal of International Economics 99, Supplement 1, 2016, S114–S133,
DOI: 10.1016/j.jinteco.2015.12.008,
<https://www.sciencedirect.com/science/article/pii/S0022199615001828>

Hansen et al. 2018

S. Hansen – M. McMahon – A. Prat, Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach, Quarterly Journal of Economics 133, 2, 2018, 801–870,
DOI: 10.1093/qje/qjx045,
<https://sekhansen.github.io/pdf_files/qje_2018.pdf>

Hilscher et al. 2024

J. Hilscher – K. Nabors – A. Raviv, Information in Central Bank Sentiment: An Analysis of Fed and ECB Communication, Working Paper, SSRN eLibrary September 2024,
DOI: 10.2139/ssrn.4797935,
<https://ssrn.com/abstract=4797935>

Jarociński – Karadi 2020

M. Jarociński – P. Karadi, Deconstructing Monetary Policy Surprises: The Role of Information Shocks, American Economic Journal: Macroeconomics 12, 2, 2020, 1–43,
DOI: 10.1257/mac.20180090,
<https://www.aeaweb.org/articles?id=10.1257/mac.20180090>

Jiang – Xiao 2022

L. Jiang – W. Xiao, A Topic-Modeling Comparison Between LDA, NMF, Top2Vec and BERTopic, Frontiers in Sociology 7, 2022,
DOI: 10.3389/fsoc.2022.886498,
<https://www.frontiersin.org/articles/10.3389/fsoc.2022.886498/full>

Kuttner 2001

K. N. Kuttner, Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market, Journal of Monetary Economics 47, 3, 2001, 523–544,
DOI: 10.1016/S0304-3932(01)00055-1,
<https://ideas.repec.org/a/eee/moneco/v47y2001i3p523-544.html>

Li et al. 2018

Y. Li – Z. Sun – X. Zhang, Short-Text Topic Modeling via Non-Negative Matrix Factorization with Topic Meaningfulness Constraints, ACM Transactions on Information Systems 36, 4, 2018, 1–31,
DOI: 10.1145/3186009,
<https://dl.acm.org/doi/fullHtml/10.1145/3178876.3186009>

Lopez-Lira – Tang 2024

A. Lopez-Lira – Y. Tang, Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models, arXiv Working Paper 2024,
<https://arxiv.org/abs/2304.07619>

Loughran – McDonald 2011

T. Loughran – B. McDonald, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, Journal of Finance 66, 1, 2011, 35–65,
DOI: 10.1111/j.1540-6261.2010.01625.x,
<https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01625.x>

Nechio – Wilson 2016

F. Nechio – D. J. Wilson, How Important Is Information from FOMC Minutes?, FRBSF Eco-

nomic Letter 2016/2016-37,
<https://www.frbsf.org/research-and-insights/publications/economic-letter/2016/12/impact-of-fomc-minutes-on-bonds>

Obstfeld – Rogoff 2000
M. Obstfeld – K. Rogoff, Foundations of International Macroeconomics (Cambridge, MA 2000)

Picault et al. 2022
M. Picault – J. Pinter – T. Renault, Media Sentiment on Monetary Policy: Determinants and Relevance for Inflation Expectations, Journal of International Money and Finance 124, 2022, 102626,
DOI: 10.1016/j.jimonfin.2022.102626,
<https://doi.org/10.1016/j.jimonfin.2022.102626>

Rosa 2013
C. Rosa, The Financial-Market Effect of FOMC Minutes, Economic Policy Review 19, 1, 2013, 67–81,
<https://ideas.repec.org/a/fip/fednep/00004.html>

Spence 1973
M. Spence, Job Market Signaling, Quarterly Journal of Economics 87, 3, 1973, 355–374,
DOI: 10.2307/1882010

Tadle 2022
R. C. Tadle, FOMC Minutes Sentiments and Their Impact on Financial Markets, Journal of Economics and Business 118, 2022, 106021,
DOI: 10.1016/j.jeconbus.2021.106021

Venade 2022
J. M. S. Venade, A Matter of Minutes: The Unexpected Sentiment in FOMC Meetings and Its Impact on Fed Credibility, (Accessed 31 July 2025) 2022,
<https://sigarra.up.pt/fep/en/pub_geral.show_file?pi_doc_id=408978>

Weidmann 2018
J. Weidmann, Central Bank Communication as an Instrument of Monetary Policy, Speech, University of South Africa, Pretoria, May 2018,
<https://www.bis.org/review/r180504b.pdf>

Appendix A: Code Availability

All code and scripts used to replicate our analyses are openly available at `github.com/sokolovolegg/sentiment-anal`
The repository contains:

- Data-processing and cleaning pipelines.
- FinBERT fine-tuning and sentiment-scoring scripts.
- Topic-modeling notebooks and utilities.
- Time-series and Granger-causality analysis code.
- Reproducible figures and table-generation scripts.

Please cite the repository if you build upon this work.