Analyzing the Impact of FED Speeches on the Stock Market: A Sentiment Analysis Approach Using Financial BERT

Simone Zani, Riccardo Baudone, Fabio Barile, Lorenzo Pirozzi and Giulia Talà

Blackswan Quants Student Organization

ARTICLE INFO

Keywords:
Sentiment analysis
Financial News
Federal Reserve
Machine learning
S&P 500
SPY
Economic Indicators
Time Series Analysis
Financial Markets
Classification

ABSTRACT

Understanding the sentiment conveyed in Federal Reserve (Fed) speeches is crucial for anticipating market reactions and economic expectations. This study employs Natural Language Processing (NLP) techniques to analyze the sentiment of Fed speeches, leveraging FinBERT, a transformer-based model fine-tuned for financial text analysis. We retrieve historical Fed transcripts, pre-process the text, and apply FinBERT to extract sentiment scores. To enhance our analysis, we employ various statistical methods to examine sentiment trends, correlations with economic indicators, and potential predictive signals for monetary policy shifts. The results provide insights into how the Fed's tone fluctuates over time and its relationship with market conditions. This paper outlines our methodology, implementation details, and key findings, offering a data-driven approach to sentiment analysis in central bank communications.

1. Introduction

The increasing role of central banks in shaping economic and financial conditions has placed increasing importance on analyzing their communications. Central bank speeches, particularly those from the Federal Reserve, serve as a crucial channel through which policymakers convey their views on economic trends, monetary policy, and financial stability. Given the significant market reactions that can follow such communications, understanding the sentiment embedded in these speeches is a valuable endeavor for both policymakers and investors.

Recent advances in Natural Language Processing (NLP) have enabled the development of sophisticated sentiment analysis models capable of extracting meaningful insights from financial texts. Our research focuses on using these advances to assess the sentiment of Federal Reserve speeches and examine their potential influence on financial markets. By systematically analyzing historical speeches, we aim to determine whether sentiment patterns correlate with market movements, particularly in the context of the S&P 500 ETF (SPY), a widely used benchmark for U.S. equity markets.

Our study follows a structured approach that integrates NLP techniques, sentiment analysis, and financial data analysis. We employ FinBERT, a sentiment analysis model fine-tuned for financial text, to evaluate the tone of Federal Reserve speeches. In addition, we merge sentiment data with intraday price movements to investigate potential relationships between speech sentiment and market reactions.

By combining computational linguistics with financial analysis, this study aims to contribute to the broader field of financial NLP and central bank communication research. The findings will offer valuable information on how sentiment conveyed in monetary policy communications interacts with market dynamics, helping to refine strategies for both economic forecasting and investment decision making.

ORCID(s):

2. Literature Review

Recent research has explored the application of Natural Language Processing (NLP) models to analyze central bank communications and their impact on financial markets. One notable study is *CentralBankRoBERTa: A fine-tuned large language model for central bank communications* by Pfeifer and Marohl (1). They introduced a version of RoBERTa fine-tuned specifically for central bank texts, demonstrating how sentiment derived from these communications can provide valuable insights into market behavior.

Another significant work, *Measuring central banks' sentiment and its spillover effects with a network approach* by Fissaha and Gerding (2), investigates the broader impact of central bank sentiment using a network-based approach. This study emphasizes how shifts in sentiment from central bank statements can influence not only financial markets but also the global economic landscape.

Additionally, Alex G. Kim's *Financial Statement Analysis with Large Language Models* (3) explores how large language models like BERT and RoBERTa can be applied to corporate financial statements. While the focus is on corporate finance, Kim's work highlights the potential of these models to extract nuanced financial sentiment, relevant for forecasting market trends.

These studies illustrate the growing relevance of NLP and sentiment analysis in both central bank and corporate contexts, offering new insights into how economic narratives influence financial markets.

3. Hypotheses

3.1. Sentiment and Market Reactions

We hypothesize that the sentiment expressed in Federal Reserve speeches has a measurable impact on financial markets. Specifically, we assume:

- Positive sentiment in speeches correlates with an increase in stock prices, particularly in SPY (S&P 500

ETF), as optimistic outlooks may boost investor confidence.

- Negative sentiment leads to a decline in stock prices due to concerns over economic instability, inflationary pressures, or policy tightening.

- Market reactions to sentiment changes are more pronounced during periods of economic uncertainty or significant monetary policy shifts.

3.2. Temporal Alignment of Sentiment and Price Movements

Hypothesis 1. *We assume that the timing of market responses is not instantaneous but follows a structured pattern:*

- *There exists a short-term lag between speech sentiment and stock price movements, as markets take time to process and react to new information.*

- *Sentiment-driven market reactions may persist for several trading sessions, particularly when speeches provide new insights into monetary policy direction.*

- *Cross-correlation analysis should reveal whether sentiment leads or lags market movements over different time horizons.*

3.3. Speakers' Influence on Market Impact

Hypothesis 2. *Given the hierarchical structure of the Federal Reserve, we expect that different policymakers have varying levels of influence on financial markets:*

- *Speeches from the Chair of the Federal Reserve (or other high-ranking officials) have a stronger impact on market sentiment than those from lower-level officials.*

- *Frequent speakers may exhibit more predictable sentiment patterns, leading to diminished market sensitivity over time.*

- *The market response to sentiment depends not only on the tone but also on the credibility and past policy decisions of the speaker.*

3.4. Effectiveness of NLP-based Sentiment Analysis

Hypothesis 3. *A crucial assumption in this study is that FinBERT, as a financial sentiment analysis model, accurately captures the intended tone of Federal Reserve speeches:*

- *FinBERT correctly differentiates between positive, neutral, and negative sentiment in financial discourse.*

- *The model's sentiment scores align with qualitative assessments by financial experts and historical market reactions.*

- *Limitations of NLP models, such as contextual ambiguity and domain-specific jargon, do not significantly distort the sentiment classification process.*

These hypotheses provide a structured framework for our analysis, allowing us to assess the validity of sentiment-based market predictions and refine our approach accordingly.

3.5. Disclaimer

All source code and implementations presented in this paper have been developed by the authors. The code is the result of original work and was created specifically for the purpose of applying machine learning NLP model onto data. Any external libraries or packages utilized within the code (such as SciPy, scikit-learn, and others) are acknowledged and referenced accordingly. However, the underlying methods, algorithms, and codebase for data analysis, preprocessing and prediction have been fully developed and customized by the authors to suit the specific requirements of this study.

4. Data Retrieval and Data Pre-Processing

This section outlines the methodology for retrieving and preprocessing speech data from the Federal Reserve, including data scraping, timestamp adjustments, and speech segmentation.

4.1. Data Retrieval

Federal Reserve speech data was collected from the official Federal Reserve website, covering speeches delivered by various speakers throughout the 2020 - 2024 time period. A custom web-scraping script was implemented using the Python libraries *requests* and *BeautifulSoup* to systematically extract speech titles and precise event start times from monthly calendar pages. URLs corresponding to each month were dynamically generated to facilitate a comprehensive extraction process for the entire year. Event start times were parsed into timezone-aware datetime objects localized to U.S. Eastern Time for consistency.

4.2. Data Preprocessing and Timestamp Adjustment

Initial speech transcripts lacked precise timestamps, instead having generalized date information. Therefore, preprocessing was required to segment the data accurately and align it with official event timings and stock market data.

First, each speech was assigned a default starting timestamp (10:00 AM Eastern Time) on its corresponding date. Next, the speech duration in minutes was estimated based on an average speaking rate of approximately 130 words per minute. Each transcript was then segmented into minute-long text portions based on this estimated duration, producing a structured dataset where each row represents one minute of speech content.

Using official event timings, we updated each speech's initial timestamp to the correct start time and incremented subsequent timestamps by one minute. Speeches shorter than five minutes or lacking a matching start time were excluded. This resulted in an overall dataset of 149 speeches.

Through this structured approach, we ensured that the final dataset accurately reflects the actual temporal sequence of speeches, which is critical for subsequent sentiment and market-impact analysis, as well as for ensuring consistency and reliability throughout the study.

5. Methodology

In this section, we will describe the methodology used in the project, from FinBERT model to further analysis on results and price action during speeches.

5.1. FinBERT Model

FinBERT is a transformer-based language model built upon BERT (Bidirectional Encoder Representations from Transformers) (7). It inherits the deep learning architecture of BERT, which consists of multi-layer bidirectional self-attention mechanisms, allowing it to capture contextual meaning in financial texts.

5.1.1. Architecture

The architecture is based on multi-layer transformer encoders. FinBERT employs 12 layers (for the base model), with 768 hidden units per layer and 12 self-attention heads, following the standard BERT-base architecture. It utilizes masked language modeling (MLM) instead of traditional left-to-right or right-to-left text processing. This masked token prediction improves contextual understanding. Additionally, FinBERT is designed for sentence pair classification, which analyzes relationships between sentences. This feature is particularly useful for understanding financial statements and policy communications.

This architecture enables FinBERT to effectively process financial sentiment analysis by considering both the direct meaning and the contextual significance of words within financial texts.

5.1.2. Pre-Training

FinBERT is pre-trained on a large corpus of financial texts, making it specialized for analyzing sentiment in finance-related contexts. Unlike general-purpose BERT models trained on Wikipedia and BookCorpus, FinBERT has been further trained on financial news articles, analyst reports, and regulatory filings.

The pre-training process involves unsupervised learning on financial data. The model was initially trained on large-scale financial documents, allowing it to develop a better understanding of domain-specific vocabulary and contextual nuances. It generates domain-specific word embeddings that differ from general NLP models, as they capture financial phrases such as "*rate hike*," "*quantitative easing*," or "*market volatility*" with better accuracy. This results in improved sentiment detection, as the pre-trained model has been evaluated on financial sentiment datasets, achieving state-of-the-art results in classifying financial statements as positive, negative, or neutral.

Since we are using FinBERT without additional fine-tuning, we leverage its pre-trained capabilities to analyze sentiment in Federal Reserve speeches directly.

5.1.3. Goals

The primary objective of using FinBERT in this study is to analyze the sentiment of Federal Reserve speeches and understand how monetary policy language affects financial markets. Specifically, we aim to classify the sentiment of different statements within Fed speeches into positive, negative, or neutral categories; identify shifts in central bank communication, particularly in response to economic conditions or policy changes; and compare sentiment trends over time, observing how policy language evolves under different macroeconomic circumstances.

By applying FinBERT to this task, we can extract valuable insights into the tone and implications of central bank communication, helping to assess the impact of monetary policy on market expectations.

5.1.4. Strategy of Training

Since we are not fine-tuning FinBERT, we adopt a zero-shot inference approach, meaning we apply the pre-trained model directly to our dataset without additional training. The main advantages of this strategy include immediate usability, as we can apply FinBERT without the need for additional labeled training data; domain relevance, as the model has already been trained on financial texts, making it well-suited for analyzing central bank communications; and efficiency, as without requiring computationally expensive fine-tuning, we can rapidly process and analyze large volumes of Fed speeches.

To ensure optimal results, we focus on preprocessing the text effectively before feeding it into FinBERT and interpreting sentiment outputs in the context of monetary policy.

5.1.5. Implementation

The implementation of our sentiment analysis pipeline using FinBERT follows these key steps:

Data Preprocessing: We extract and segment speeches from the Federal Reserve into minute-by-minute sentences, remove irrelevant content (e.g., headers, citations) to retain only the core speech text, and tokenize the text using the BERT tokenizer, which is designed to handle subwords efficiently.

Sentiment Analysis: We input each sentence into the pre-trained FinBERT model, retrieve sentiment probabilities for positive, negative, or neutral classifications, and assign the sentiment label based on the highest probability.

Post-processing and Interpretation: We aggregate sentiment scores across speeches or time periods, compare sentiment trends with macroeconomic indicators or market reactions, and analyze linguistic patterns associated with hawkish vs. dovish policy tones.

The use of FinBERT as a pre-trained model allows us to extract sentiment insights efficiently without the need for additional labeled training data or extensive computational resources.

5.1.6. Conclusion

In this study, we have chosen FinBERT as our sentiment analysis model due to its superior performance in handling financial texts. Unlike general-purpose NLP models such as BERT or RoBERTa, FinBERT is specifically pre-trained on financial data, making it particularly well-suited for analyzing Federal Reserve speeches, where monetary policy language plays a crucial role in shaping market expectations. Its domain-specific training enables it to better interpret financial terminology, policy statements, and sentiment nuances, such as hawkish vs. dovish tones, that are essential in central bank communication analysis.

5.1.7. Why We Chose FinBERT

Our decision to use FinBERT instead of other NLP models is supported by several key factors. First, FinBERT has proven superiority in financial sentiment analysis. Multiple studies have shown that FinBERT outperforms other models in this area. Araci (4) demonstrated that FinBERT improved sentiment classification accuracy by 14 percentage points over traditional methods. RAM AI (16) found that FinBERT outperformed dictionary-based approaches and other transformer models. Recent studies confirmed FinBERT's higher accuracy and F1 scores than models like DistilBERT, RoBERTa (11), and generic BERT (7) (17) (10).

Second, FinBERT is optimized for financial texts. It has been pre-trained on financial news, earnings reports, and market analyses, making it more context-aware in financial sentiment classification. General-purpose models trained on Wikipedia and BookCorpus do not capture domain-specific financial terminology as accurately.

Third, FinBERT demonstrates superior handling of neutral sentiment. Many central bank speeches contain neutral, non-committal language. FinBERT has been trained on datasets with a high proportion of neutral statements, making it more reliable in distinguishing neutral from positive/negative sentiment compared to generic models.

Fourth, FinBERT offers immediate usability without fine-tuning. Its pre-trained model performs well out-of-the-box, eliminating the need for expensive and time-consuming fine-tuning. This allows us to immediately apply it to Federal Reserve speeches without requiring additional labeled training data.

Finally, FinBERT provides computational efficiency. Fine-tuning requires significant computational resources (e.g., GPUs, large datasets). Using FinBERT without additional training enables efficient inference and large-scale analysis with minimal computational cost.

5.1.8. Why We Chose Not to Fine-Tune FinBERT

We opted not to fine-tune FinBERT and instead leverage its pre-trained capabilities in a zero-shot inference approach. This decision is driven by several key considerations.

Pretrained suitability is a significant factor. FinBERT has already been extensively trained on financial sentiment datasets, ensuring that it generalizes well to monetary policy speeches without requiring additional domain adaptation. There is also a lack of labeled data. Fine-tuning requires large, high-quality labeled datasets, which are often expensive and difficult to obtain for central bank communications.

Computational efficiency is another consideration. Fine-tuning is computationally intensive. By using FinBERT without additional training, we eliminate the need for expensive GPUs and reduce processing time. Additionally, FinBERT demonstrates robust generalization. Its financial-domain pretraining ensures it performs well across different financial contexts, minimizing the risk of overfitting to a limited set of Fed speeches.

By adopting this pretrained, zero-shot approach, we ensure a balance between accuracy, efficiency, and interpretability, allowing us to systematically extract sentiment insights from Federal Reserve communications. This methodology provides a reliable foundation for studying monetary policy sentiment trends and their potential implications on financial markets.

5.2. Statistical Methods Used for Analysis

5.2.1. Cross-Correlation

Cross-correlation measures the degree of similarity between two time series X_t and Y_t at different time lags. Formally, the cross-correlation function (CCF) at lag k is given by:

$$\rho_{XY}(k) = \frac{\mathrm{Cov}(X_t, Y_{t+k})}{\sigma_{X_t}\sigma_{Y_{t+k}}}, \quad k \in \mathbb{Z} \tag{1}$$

where

$$\sigma^2_{X_t} = \mathrm{Var}(X_t), \quad \sigma^2_{Y_{t+k}} = \mathrm{Var}(Y_{t+k}).$$

This statistic helps to identify potential lead-lag relationships between the variables. When applied to our dataset, we examine whether shifts in the sentiment series (derived from FinBERT) precede or follow movements in market prices.

5.2.2. Pearson Correlation

The Pearson correlation coefficient quantifies the strength and direction of the linear relationship between two random variables X and Y. Mathematically, it can be written in expectation form as:

$$r_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\sigma^2_X \, \sigma^2_Y}} \tag{2}$$

where

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right], \\
\sigma^2_X &= \mathbb{E}\left[(X - \mu_X)^2\right], \\
\sigma^2_Y &= \mathbb{E}\left[(Y - \mu_Y)^2\right].
\end{aligned}$$

The coefficient r_{XY} ranges from -1 to $+1$, with -1 indicating a perfect negative correlation, 0 indicating no linear relationship, and $+1$ indicating a perfect positive correlation. We use Pearson correlation to assess the linear dependencies between sentiment and price returns over various time windows, aiding in the identification of concurrent relationships.

5.2.3. *Skewness*

Skewness measures the asymmetry of a distribution around its mean. For a random variable X with mean μ and standard deviation σ, we express skewness as:

$$\text{Skewness}(X) = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3} \tag{3}$$

Positive skewness indicates a longer right tail (potential for extreme positive values), while negative skewness suggests a longer left tail (greater likelihood of extreme negative values). Within our framework, we evaluate skewness to detect asymmetrical distributions in both sentiment scores and market returns, illuminating the presence of non-symmetric risk or sentiment shocks.

5.2.4. *Kurtosis*

Kurtosis captures the “tailedness” of a distribution relative to the normal distribution. For a random variable X with mean μ and standard deviation σ, the kurtosis is given by:

$$\text{Kurtosis}(X) = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4} \tag{4}$$

A high kurtosis (leptokurtic) signifies fat tails and an elevated risk of extreme outcomes, whereas low kurtosis (platykurtic) indicates thinner tails and fewer outliers. By examining the kurtosis of market returns, we gauge whether events such as policy announcements introduce heavier-tailed distributions—implying heightened risks or unexpected market reactions.

5.2.5. *Integration of Statistical Measures*

In our analysis, these statistical measures jointly clarify how sentiment relates to price dynamics:

- **Cross-Correlation**: Detects if and when sentiment leads or lags price changes.
- **Pearson Correlation**: Captures linear dependence between the sentiment series and price returns.
- **Skewness** and **Kurtosis**: Reveals distributional nuances that might signal asymmetric risk profiles or outlier behaviors coinciding with central bank communications.

Together, these tools form a holistic view of the statistical behavior of both sentiment and market variables, enabling deeper insight into how Federal Reserve speeches may influence financial conditions.

5.3. Technical Indicator for Analysis

5.3.1. *Volume Weighted Average Price (VWAP)*

The Volume Weighted Average Price (VWAP) represents a quantitative trading benchmark that calculates the ratio of the value traded to the total volume traded over a specific time horizon. Mathematically, it is expressed as:

$$\text{VWAP} = \frac{\sum_{i=1}^{n} P_i V_i}{\sum_{i=1}^{n} V_i} \tag{5}$$

where P_i represents the price of the i-th transaction and V_i represents the volume of the i-th transaction.

VWAP serves as a critical reference point for institutional investors and algorithmic trading systems to evaluate execution efficiency. By incorporating volume data, VWAP provides a more nuanced representation of market activity compared to simple moving averages, as it weights price movements according to their trading intensity. This characteristic makes VWAP particularly valuable for analyzing market reactions to significant events such as Federal Reserve communications.

In our analytical framework, we implemented VWAP volatility bands at standard deviation levels of 1.0 and 1.5 to establish dynamic support and resistance thresholds. These bands are calculated as:

$$\text{VWAP}_{\text{upper}} = \text{VWAP} + \sigma \times k$$

$$\text{VWAP}_{\text{lower}} = \text{VWAP} - \sigma \times k$$

where σ represents the standard deviation of price from VWAP and k represents the multiplier (1.0 or 1.5 in our analysis).

The empirical significance of these volatility bands lies in their capacity to identify statistical anomalies in price action. When asset prices approach or breach the upper volatility band (VWAP $+1.5\sigma$), this suggests a statistically significant deviation from equilibrium pricing that may indicate temporary overvaluation. Conversely, penetration of the lower band (VWAP -1.5σ) may signal potential undervaluation. These statistical boundaries provide contextual frameworks for quantifying market reactions to Fed communications with respect to established trading patterns.

By integrating VWAP analysis with the sentiment outputs from FinBERT, we establish a methodological bridge between quantitative market dynamics and qualitative linguistic signals, enabling a more comprehensive assessment of how Federal Reserve communications influence market behavior across various volatility regimes.

6. Implementation Details (Key Functions in Python Code)

This section aims to illustrate how we implemented and organized the code (which is freely available in its entirety on our GitHub repository). In this section, we will present only two aspects: the process of data scraping for the metadata of the FED speeches and the pipeline for utilizing FinBERT. For the functions and files illustrated, it will be possible to visualize the GitHub code by just clicking on them.

6.1. Data Retrieval Process and Implementation

A crucial aspect of our research project involves the automatic retrieval of historical Federal Reserve speeches from the official Federal Reserve website. To achieve this, we developed a Python function called `breakdown_html()`, which is responsible for parsing the website's calendar pages and extracting relevant information such as the date, time, title, and speaker of each speech. Below, we explain the purpose, structure, and logic of this function.

6.1.1. Overview of `breakdown_html()`

The function `breakdown_html()` accepts a URL corresponding to a specific month and year from the Federal Reserve's speech calendar. It performs the following steps:

1. Sends an HTTP GET request to retrieve the webpage's HTML content.
2. Parses the HTML using the `BeautifulSoup` library, which allows easy navigation through the document tree.
3. Identifies the section of the page labeled "Speeches" using its specific HTML class attributes.
4. Iteratively extracts key details for each listed speech, including:
    - **Title**: The descriptive title of the speech event.
    - **Date**: The day of the month when the speech occurred.
    - **Time**: The scheduled time for the speech.
5. Returns these elements in the form of three lists, namely `titles_list`, `dates_list`, and `times_list`.

6.1.2. Technical Explanation

The core logic is implemented as follows. The function starts by making a request to the specified URL.

This ensures the request mimics a browser visit, avoiding common scraping blocks. If the page loads successfully (HTTP status 200), the function proceeds to parse the HTML and locate all calendar sections marked with the class:

```
soup.find_all('div', class_='row cal-nojs__rowTitle')
```

Next, the function looks specifically for the section titled "Speeches". Once located, it scans the subsequent HTML blocks (siblings) to collect all speeches listed under this section. Each speech entry is structured into:

- A div block containing the **title** (including the speaker's name embedded in HTML paragraphs),
- A div block containing the **time** in AM/PM format,
- A div block containing the **date** as a numerical day of the month.

The collected information is stored into the three lists: `titles_list`, `times_list`, and `dates_list`. Each list may contain extraneous entries (due to irregular page formatting), so basic post-processing is applied to filter valid data:

```
titles_list = titles_list[0:]
times_list = times_list[1:]
dates_list = dates_list[1:]
```

Finally, these lists are returned to the calling function, enabling downstream processing where the text is further cleaned, speaker names are isolated, and times are converted into a standard format.

6.1.3. Function Usage in the Pipeline

The `breakdown_html()` function is used within a larger loop that iterates over all months and years in the study period. Each month's URL is built dynamically, passed into the function, and the extracted data is immediately processed into a structured DataFrame using additional helper functions such as `handle_titles()`, `time_handling()`, and `handle_dates()`. This modular approach ensures that data retrieval is decoupled from subsequent parsing and formatting steps.

6.1.4. Rationale for Web Scraping

Since the Federal Reserve does not offer a structured API for historical speech data, web scraping is necessary to automate data collection. By directly parsing the calendar pages, we ensure comprehensive coverage and minimize manual intervention. To comply with ethical scraping guidelines, we use appropriate headers to simulate regular browser behavior, avoid excessive requests, and limit scraping to publicly accessible data.

6.2. Sentiment Analysis Using FinBERT

The FinBERT model is employed to analyze the sentiment of Federal Reserve speeches. This section describes the implementation details of the functions responsible for executing FinBERT and handling its output data.

6.2.1. Implementation of Sentiment Analysis Functions

The sentiment analysis pipeline consists of several functions that load the FinBERT model, process textual data, and store the results in structured formats for further analysis.

6.2.2. FinBERT Sentiment Computation

The function `finbert_sentiment()` processes a given speech excerpt and outputs the probability scores for three sentiment classes: positive, negative, and neutral. The function structure is as follows:

1. Tokenizes the input text using the FinBERT tokenizer.
2. Passes the tokenized input to the FinBERT model to obtain logits.
3. Applies a softmax function to compute probability scores for each sentiment category.
4. Determines the sentiment class with the highest probability.
5. Returns the sentiment probabilities and the dominant sentiment label.

6.2.3. Applying Sentiment Analysis to Speeches

The function `analyze_sentiment()` is a simplified wrapper that processes textual data using a sentiment pipeline. It ensures that text input remains within the model's 512-token limit:

```
def analyze_sentiment(text, nlp):
    result = nlp(text[:512])
    return result[0]["label"], result[0]["score"]
```

6.2.4. Aggregating Sentiment Scores

To facilitate downstream analysis, the function `aggregate_sentiment_confidence()` groups sentiment confidence scores by date and speaker. It computes the total confidence scores for each sentiment category.

A complementary function, *aggregate_sentiment_iterator()*, extends this process over multiple years by iterating through preprocessed pickle files containing sentiment data:

6.2.5. Executing the Sentiment Analysis Pipeline

The function `compute_sentiment()` orchestrates the sentiment analysis workflow:

1. Loads the FinBERT model and tokenizer.
2. Applies `finbert_sentiment()` to extract sentiment scores for each speech minute.
3. Computes the expected sentiment score as `finbert_pos - finbert_neg`.
4. Uses `analyze_sentiment()` for additional sentiment classification.
5. Saves the processed data as a CSV file and a pickle object for later retrieval.

6.2.6. Data Storage and Management

The processed sentiment data is stored in both CSV and pickle formats to facilitate analysis. The CSV file `2020-2024sentiment.csv` provides an easily accessible tabular representation, while the pickle file allows for efficient data retrieval in Python workflows. The `mh.PickleHelper` module is used to handle serialization and deserialization of the sentiment dataset.

7. Consideration About The Results

This section presents and discusses the results obtained from the sentiment analysis of Federal Reserve speeches, shedding light on the effectiveness of the proposed methodology and the key insights derived from our analysis. Understanding the relationship between the Fed's tone and market or economic responses is not only a central objective of our work but also a relevant contribution to the broader field of central bank communication studies.

Our analysis is divided into two complementary parts. First, we showcase visual representations of sentiment dynamics over time, as well as comparisons between sentiment scores and selected market indicators. These plots, created using the matplotlib library, serve to highlight trends and potential correlations in an intuitive manner. Second, we present statistical measures and correlation coefficients to quantitatively assess the strength and significance of these relationships. This part of the analysis relies primarily on SciPy and NumPy, which allowed us to compute key statistics and test hypotheses regarding sentiment's predictive power for market movements and monetary policy shifts.

Beyond presenting numerical results, we aim to interpret these findings through a critical and discretionary lens, comparing them with our initial expectations and the existing literature. This interpretative effort is fundamental to contextualizing the results and understanding their economic relevance. In particular, we will discuss potential causes for deviations from our hypotheses, such as unexpected external shocks, evolving communication strategies by the Fed, and limitations inherent to textual sentiment analysis models like FinBERT.

By combining visual, statistical, and interpretative approaches, this section seeks to extrapolate useful insights and offer a comprehensive understanding of how the sentiment conveyed in Federal Reserve communications interacts with financial markets and economic conditions.

7.0.1. The choice of SPY

The choice to focus on SPY, the ETF tracking the S&P 500 index, stems from several considerations. SPY offers a high level of liquidity, ensuring reliable price data across different time intervals. Moreover, SPY accurately reflects the overall performance of the U.S. equity market, making it an appropriate proxy for assessing the relationship between Fed communications and market reactions. Finally, historical data for SPY is readily accessible, facilitating consistent data retrieval and ensuring compatibility with the analysis framework.

7.0.2. Speeches under study

We analyzed a total of 261 Federal Reserve speeches. After applying filtering criteria related to speech timing, market opening hours, and data availability (excluding speeches with missing price information), the final dataset consisted of 149 speeches. The average speech length was approximately 18.5 minutes, with durations ranging from a minimum of 4 minutes to a maximum of 47 minutes.

The distribution of speeches across years is as follows:

- 2020: 18 speeches
- 2021: 43 speeches
- 2022: 20 speeches
- 2023: 31 speeches
- 2024: 37 speeches

7.1. Density Distribution of Sentiment Scores

We decided to study the density distribution of the sentiment scores produced by the FinBERT model. We plotted the density distribution over the entire time span (2020-2024) and separately for each year, allowing us to better isolate year-specific behaviors. Additionally, we identified the

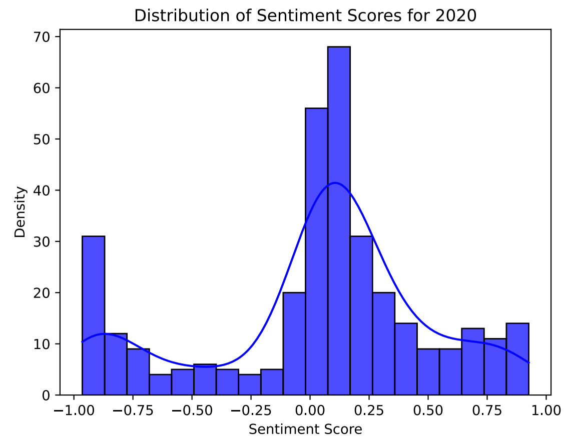

Figure 1: Sentiment Distribution 2020

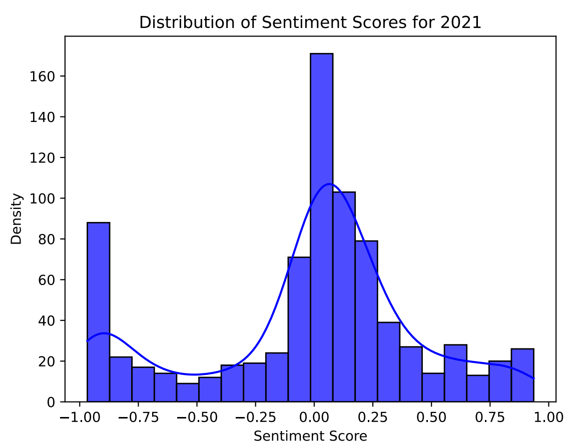

Figure 2: Sentiment Distribution 2021

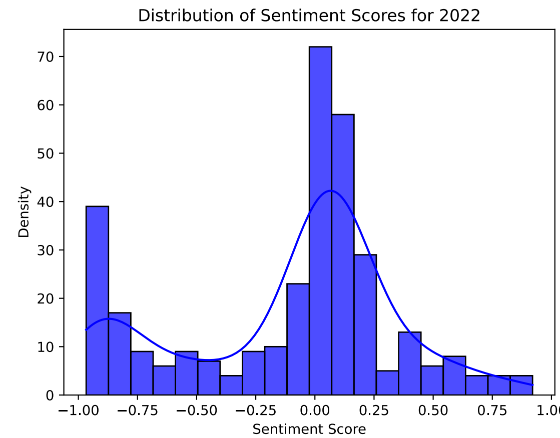

Figure 3: Sentiment Distribution 2022

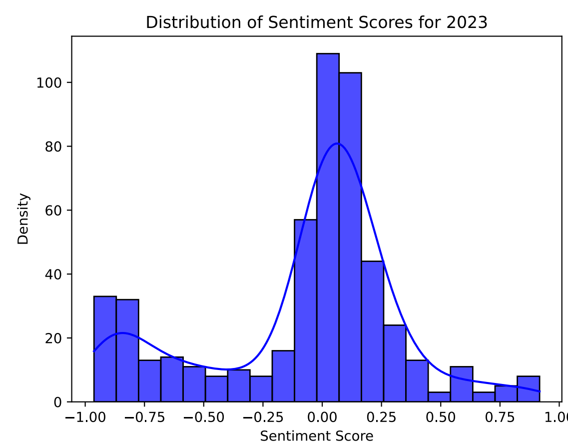

Figure 4: Sentiment Distribution 2023

most frequently speaking governors to assess the sentiment distribution of their speeches specifically.

7.1.1. Yearly Density Distribution

Kurtosis 5.2.4 and skewness 5.2.3 provide valuable insights into the shape and behavior of the return distributions.

These metrics are particularly relevant in financial analysis as they provide a deeper understanding of risk beyond standard deviation, helping to assess the likelihood of extreme market movements.

Descriptive Statistics Table 1 summarizes the key statistical measures — mean, skewness, and kurtosis — for the sentiment scores across the speeches delivered by the Federal Reserve between 2020 and 2024.

Table 1
Descriptive Statistics of Sentiment Scores in Federal Reserve Speeches (2020-2024)

Year	Mean	Skewness	Kurtosis
2020	0.0278	-0.5351	-0.1880
2021	-0.0243	-0.4544	-0.1984
2022	-0.1110	-0.4683	-0.5190
2023	-0.0749	-0.6045	0.0146
2024	-0.0592	-0.3883	0.0262
Overall (2020-2024)	-0.0473	-0.4526	-0.1391

Interpretation and Contextualization The overall mean sentiment score for the 2020-2024 period stands at −0.0473, indicating a slightly negative average tone across all speeches. This suggests that, on balance, Federal Reserve communications during this timeframe tended to lean more cautious or pessimistic, likely reflecting the challenges posed by the economic and financial environment over these years — including the COVID-19 pandemic, inflationary pressures, and evolving monetary policy stances. This is consistent with Priola-Lorenzini (18) on central banks sentiment across 2000-2020, which showed a negative trend in the regulators' tone.

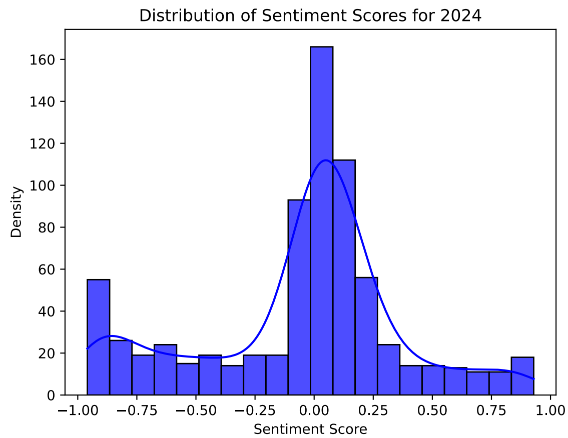

Figure 5: Sentiment Distribution 2024

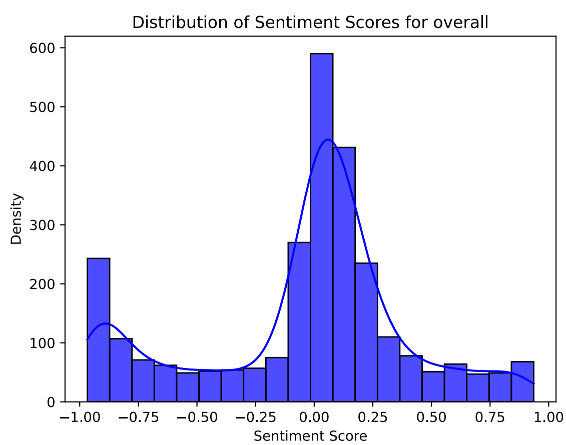

Figure 6: Sentiment Distribution Overall

Yearly Trends. Breaking down the sentiment by year reveals important shifts in tone. In 2020, the mean sentiment score was positive (0.0278), possibly reflecting the initial accommodative stance adopted by the Federal Reserve to stabilize markets and support economic recovery during the early stages of the pandemic. However, starting from 2021, the average sentiment score turned negative, with the sharpest drop occurring in 2022 (-0.1110). This aligns with the Federal Reserve's pivot towards a more restrictive monetary policy stance as inflation surged, necessitating repeated interest rate hikes. The tone in 2023 and 2024 remained negative, albeit less extreme, signaling a cautious and still somewhat restrained communication style, as economic uncertainty persisted.

Skewness and Distributional Asymmetry. The skewness values are consistently negative across all years, indicating a left-skewed distribution of sentiment scores. This skewness highlights that, in each year, a substantial proportion of speeches conveyed sentiment more negative than the average. Such a pattern is particularly pronounced in 2023 (-0.6045), when the Federal Reserve faced heightened scrutiny over the risk of recession triggered by its tightening cycle.

Kurtosis and Tail Behavior. The kurtosis values fluctuate around zero, with slightly negative values for 2020, 2021, and 2022, and slightly positive values for 2023 and 2024. These values suggest that the distribution of sentiment scores does not exhibit extreme tail behavior; in other words, the occurrence of highly positive or highly negative speeches is not significantly more (or less) frequent than in a normal distribution. The absence of strong excess kurtosis implies that, despite changes in tone over the years, the Federal Reserve maintained a relatively balanced approach, avoiding overly extreme communication styles.

Contextual Reflection The observed evolution of sentiment scores reflects broader macroeconomic dynamics. The relatively positive tone in 2020 coincided with unprecedented monetary support and optimism regarding recovery. As inflation pressures emerged in 2021, the tone shifted into negative territory, reflecting growing concerns over price stability and the need for corrective action. The particularly low sentiment in 2022 is consistent with the aggressive policy tightening cycle launched that year, accompanied by more cautious and sometimes explicitly hawkish communication from Federal Reserve officials.

In 2023 and 2024, sentiment remained negative but showed slight improvements, potentially linked to initial signs of inflation moderation and growing market expectations of a possible policy pivot. Overall, these statistical patterns highlight how sentiment analysis can serve as a valuable quantitative tool for tracing the evolving stance and communication strategy of central banks under varying economic conditions.

7.2. Top Speaker Distribution

The speakers chosen for the study are **Governor Christopher J. Waller** (19% of the dataset, nominated by President D. Trump, still in charge) , **Governor Michelle W. Bowman** (19% of the dataset, nominated by President D. Trump , still in charge) , **Vice Chair for Supervision Randal K. Quarles** (10% of the dataset , nominated by President D. Trump , his duty ended on 2021).

Descriptive Statistics by Speaker Table 2 summarizes the descriptive statistics for the sentiment scores associated with speeches delivered by each of these individuals.

Table 2
Descriptive Statistics of Sentiment Scores by Federal Reserve Speaker (2020-2024)

Speaker	Mean	Skewness	Kurtosis
Christopher J. Waller	-0.1179	-0.4258	-0.2459
Michelle W. Bowman	-0.0010	-0.8495	1.2508
Randal K. Quarles	0.0103	-0.4506	0.1042

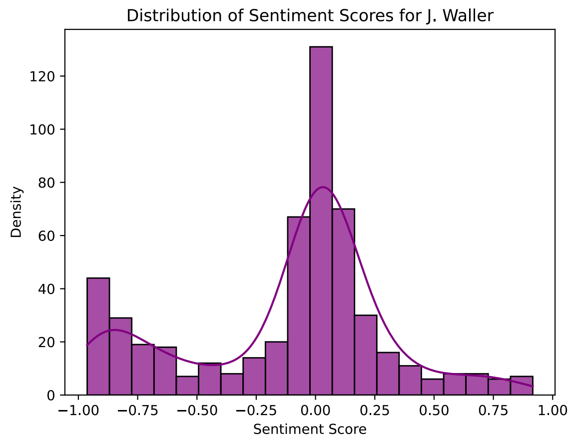

Figure 7: Sentiment Distribution Gov. Waller

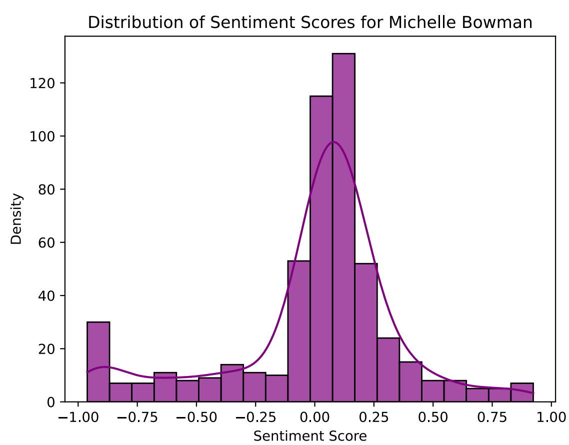

Figure 8: Sentiment Distribution Gov. Bowman

Interpretation and Contextualization The results reveal substantial heterogeneity in communication tone across the selected speakers, highlighting the importance of accounting for individual communication styles when interpreting Federal Reserve messaging.

Speakers Analysis The sentiment profiles of Federal Reserve officials reveal distinct communication styles. **Governor Christopher J. Waller** exhibits the most negative average sentiment score (-0.1179) , aligning with his data-dependent, risk-aware policy stance. His speeches show a slight negative skewness (-0.4258) and small negative kurtosis (-0.2459), indicating a broadly even distribution with fewer extreme values. In contrast, **Governor Michelle W. Bowman** maintains a nearly neutral mean sentiment (-0.0010) but stands out for her highly negative skewness (-0.8495) and high positive kurtosis (1.2508), reflecting a tendency for more extreme negative sentiment episodes. This pattern aligns with her focus on structural risks in financial stability. **Vice Chair Randal K. Quarles**, whose tenure ended in 2021, presents a slightly positive sentiment

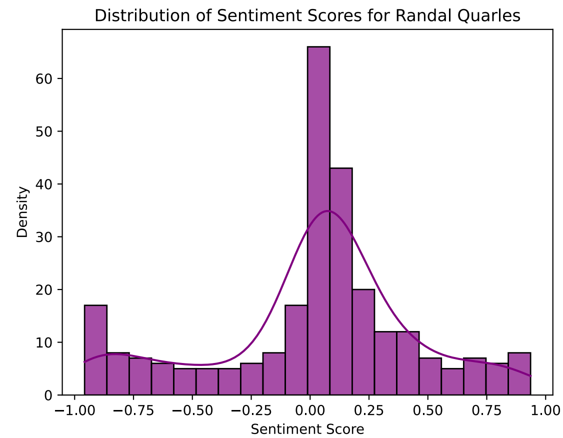

Figure 9: Sentiment Distribution Vice Chair Quarles

score (0.0103), consistent with his emphasis on regulatory resilience. His skewness (-0.4506) suggests a mild left skew, while his kurtosis (0.1042) indicates a near-normal sentiment distribution. These profiles illustrate how each official's communication reflects their policy priorities and risk perceptions

Contextual Reflection The distinct sentiment profiles for each speaker reflect differences in their individual roles, policy focuses, and communication strategies. **Waller**, as a key voice on monetary policy implementation, naturally conveys a more reactive and cautious tone tied to economic data developments. **Bowman**, with a remit that includes community banking and regulatory issues, adopts a more diverse communication strategy, potentially explaining the elevated kurtosis and frequent swings in sentiment. **Quarles**, in his supervisory role, maintained a relatively constructive and optimistic tone, reflecting the stability-focused nature of his speeches.

These differences underscore the importance of disaggregating sentiment analysis not only by time but also by speaker identity. Central bank communication is inherently heterogeneous, and aggregate sentiment measures risk obscuring these crucial individual variations. This insight enhances the interpretability of sentiment-based indicators and offers a richer understanding of how individual policymakers contribute to the broader communication landscape of the Federal Reserve.

7.3. Plot Sentiment versus Price Metrics

7.3.1. Which speeches we plotted

We confined our plot analysis to the four speeches characterized by the highest market **volatility** and the four **longest** speeches. Initially, we plotted the sentiment score against the percentage change in the price, followed by the sentiment score against the price of the ETF. Additionally, we incorporated the **VWAP** bands 5.3.1 corresponding to volatility levels of 1.0 and 1.5. Finally, we generated a **word cloud** to provide context regarding the primary topics

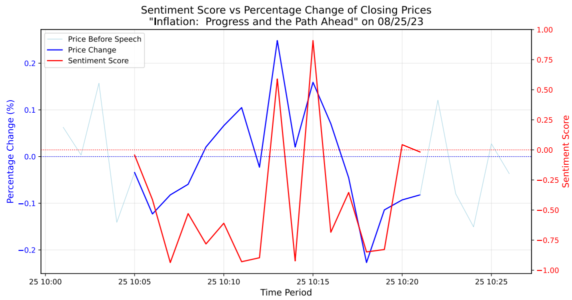

Figure 10: Sentiment Score vs Percentage Change of Closing Price - Top Volatility (0.94)

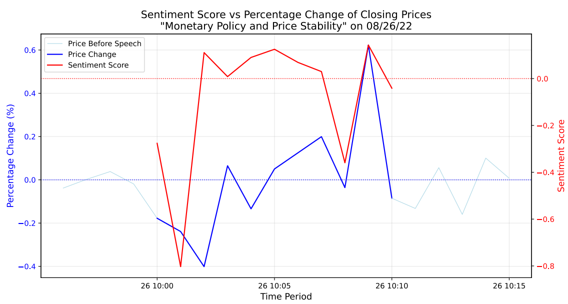

Figure 11: Sentiment Score vs Percentage Change of Closing Price - Top Volatility (1.42)

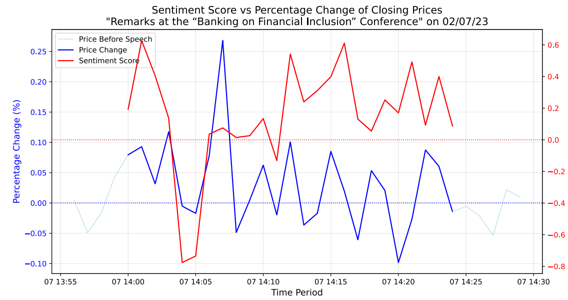

Figure 12: Sentiment Score vs Percentage Change of Closing Price - Top Volatility (0.95)

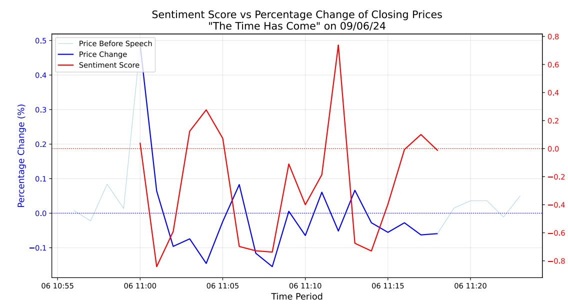

Figure 13: Sentiment Score vs Percentage Change of Closing Price - Top Volatility (1.19)

discussed in the selected speeches and to identify potential topics to which the **market** may be particularly **sensitive**.

Top Volatility We calculated volatility during each speech by computing the minute-by-minute standard deviation of SPY closing prices. The assumption is that heightened volatility reflects increased uncertainty and potential new information being incorporated into prices, in line with the **Efficient Market Hypothesis**.

Longest Speeches We repeated the same set of plots for the four longest speeches, which had lengths of 47, 40, 37, and 33 minutes respectively. This selection follows the assumption that longer speeches might contain more information, leading to more pronounced market responses. We calculated the length by counting the number of 'minute-by-minute' script, considering an average of 120 words per minute.

Sentiment vs Percentage Change

Observations and Interpretation. The visual inspection of these time-aligned plots reveals that, in certain periods, the two curves — sentiment score and price change — appear to move in the same direction. This alignment is particularly visible in speeches focusing on topics such as *monetary policy and price stability*, where shifts towards a more positive (or negative) tone seem to coincide with upward (or downward) adjustments in market prices. Such patterns are intuitively consistent with the idea that investors actively

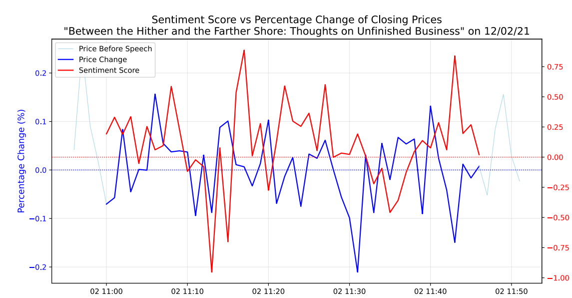

Figure 14: Sentiment Score vs Percentage Change of Closing Price - Longest Speech (47 mins)

parse central bank communication for signals relevant to future monetary policy actions and economic outlooks.

However, this relationship is neither constant nor uniform. In several cases, price movements appear to *lag* changes in sentiment, suggesting that certain parts of a speech — especially those conveying nuanced or mixed messages — require more time for market participants to fully process and incorporate into trading decisions. This is particularly relevant for complex topics, such as regulatory changes or systemic risk, where the sentiment signal may be less immediately interpretable.

Interestingly, this alternating pattern of alignment and divergence is also visible when focusing on the longest speeches within the dataset — speeches that, by their nature, could be expected to contain a higher volume of information

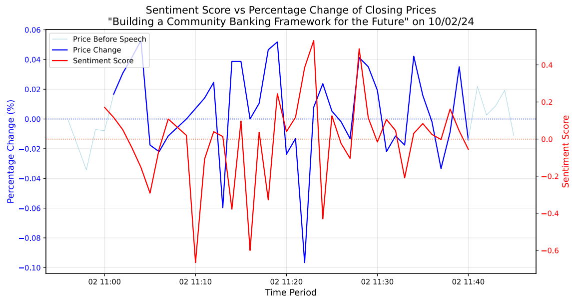

Figure 15: Sentiment Score vs Percentage Change of Closing Price - Longest Speech (40 mins)

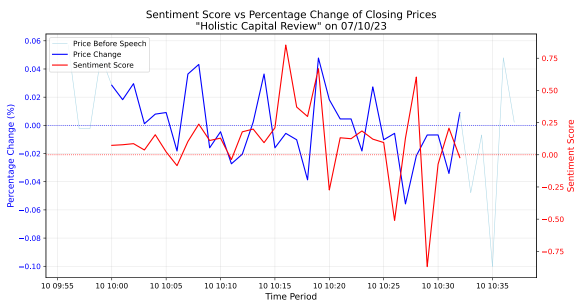

Figure 16: Sentiment Score vs Percentage Change of Closing Price - Longest Speech (33 mins)

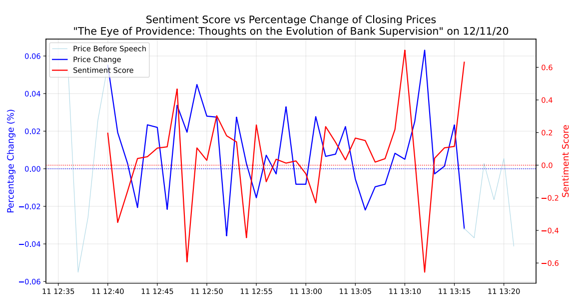

Figure 17: Sentiment Score vs Percentage Change of Closing Price - Longest Speech (37 mins)

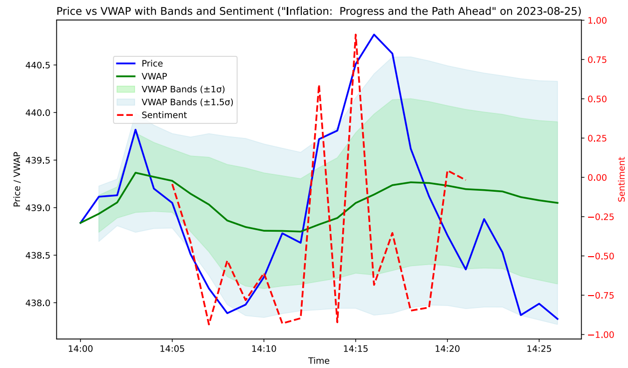

Figure 18: Sentiment Score vs Closing Price along with VWAP bands - Top Volatility (0.94)

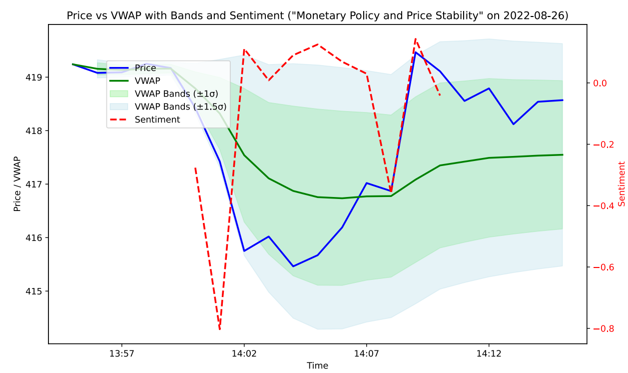

Figure 19: Sentiment Score vs Closing Price along with VWAP bands - Top Volatility (1.42)

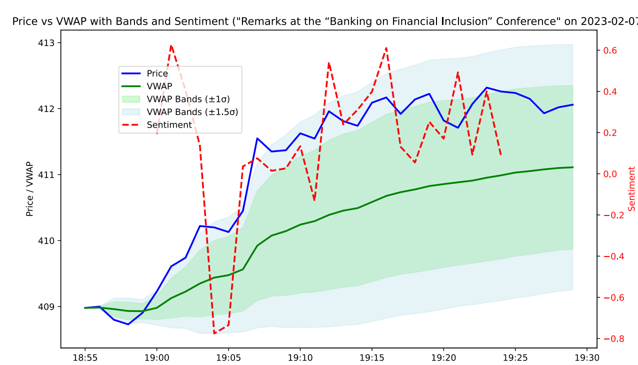

Figure 20: Sentiment Score vs Closing Price along with VWAP bands - Top Volatility (0.95)

and thus exert a stronger market influence. However, length alone does not appear to guarantee informativeness or market sensitivity. Longer speeches, by covering a broader set of topics, may dilute the clarity of the sentiment signal, especially if market participants only focus on select passages directly addressing key policy matters. These factors collectively contribute to the fragmented and context-dependent nature of the sentiment-price relationship.

7.3.2. Sentiment vs Close Price and VWAP

In our analysis, incorporating the VWAP bands 5.3.1 and volatility levels alongside sentiment scores provides a more comprehensive understanding of how market reactions to Federal Reserve speeches evolve under varying conditions. By examining the sentiment data in the context of VWAP and volatility bands, we can assess whether market participants are responding to Federal Reserve remarks in a way that is consistent with price movements within expected volatility ranges, or if external factors are influencing the market in a manner that deviates from typical trading patterns.

Observations and Interpretation. The visual inspection of these plots suggests that no clear and consistent pattern emerges between sentiment scores, price values, and VWAP

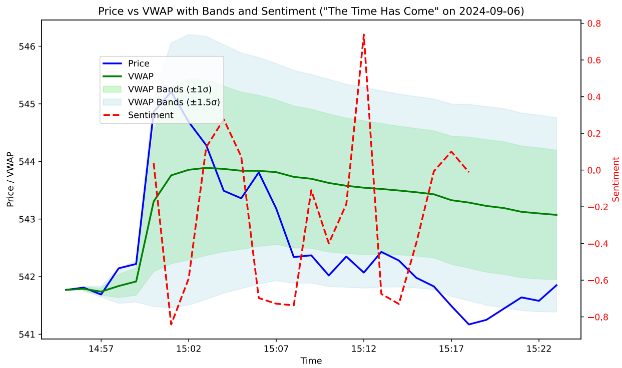

Figure 21: Sentiment Score vs Closing Price along with VWAP bands - Top Volatility (1.19)

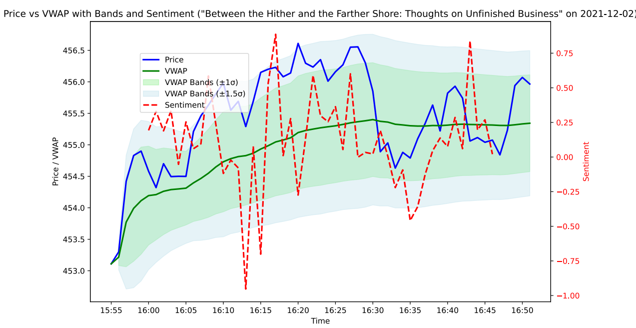

Figure 22: Sentiment Score vs Closing Price along with VWAP bands - Longest Speech (47 mins)

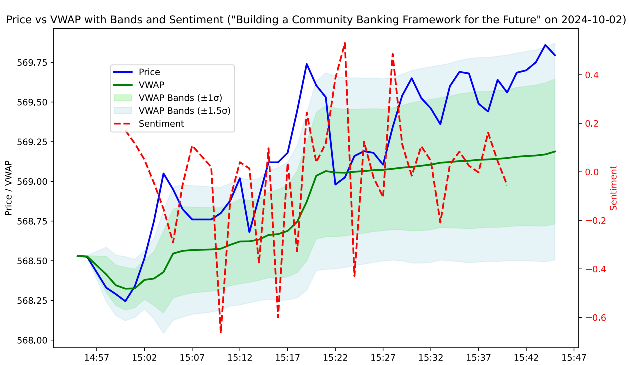

Figure 23: Sentiment Score vs Closing Price along with VWAP bands - Longest Speech (40 mins)

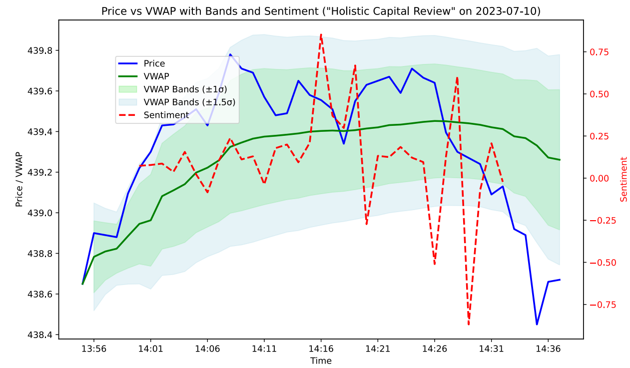

Figure 24: Sentiment Score vs Closing Price along with VWAP bands - Longest Speech (33 mins)

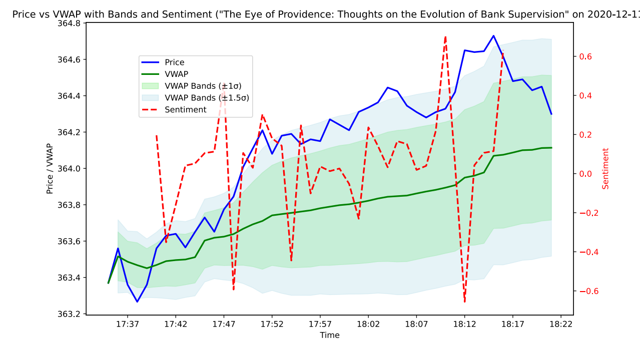

Figure 25: Sentiment Score vs Closing Price along with VWAP bands - Longest Speech (37 mins)

Figure 26: WordsCloud - Top Volatility (0.94)

bands. While sentiment scores—ranging from -1 to 1—capture the polarity of Federal Reserve speeches, they are inherently more suited for studying the percentage change in prices rather than absolute price levels. This distinction is crucial, as price movements in response to sentiment shifts may be more effectively analyzed in a relative rather than absolute framework.

Furthermore, the expected relationship between price and the VWAP line does not display a systematic structure across the dataset. While certain instances exhibit brief periods where price trajectories and VWAP align with sentiment shifts, these occurrences appear inconsistent and are not robust enough to suggest a generalizable trend. Market reactions to Federal Reserve communications seem to be influenced by additional factors beyond sentiment alone, reinforcing the notion that external macroeconomic conditions and broader investor sentiment play a role in shaping price behavior.

Figure 27: WordsCloud - Top Volatility (1.42)

Figure 30: WordsCloud - Longest Speech (47 mins)

Figure 28: WordsCloud - Top Volatility (0.95)

Figure 31: WordsCloud - Longest Speech (40 mins)

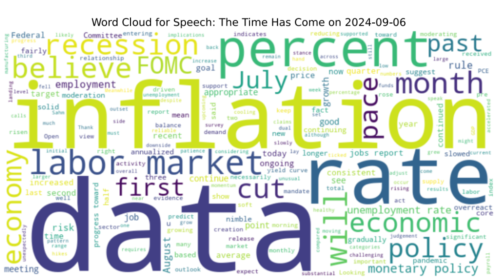

Figure 29: WordsCloud - Top Volatility (1.19)

Figure 32: WordsCloud - Longest Speech (33 mins)

7.3.3. WordsCloud

Word Cloud Analysis. The word cloud analysis provides valuable insights into the themes and concerns addressed in Federal Reserve speeches, particularly when comparing speeches associated with high volatility to those of longer duration. The most frequent words in the word clouds for **high-volatility speeches** reflect a strong focus on **economic conditions**, **inflation**, and **policy** responses to current challenges, such as the *pandemic* and the broader *economic recovery*. The recurrence of terms like *inflation*, *price stability*, and *growth* aligns with the heightened market sensitivity to monetary policy changes during periods of economic stress, explaining why such speeches are associated with increased volatility.

In contrast, the words in the **Longest Speech** group indicate that the primary focus of the longest speeches is on **financial stability, regulatory frameworks**, and **risk management**. The recurrence of *bank*, *capital*, and *risk* suggests an ongoing conversation about the **financial sector** and the Fed's role in overseeing its stability, while words like *supervisory*, *regulation*, and *framework* emphasize the Fed's regulatory responsibilities. These topics, while important,

Figure 33: WordsCloud - Longest Speech (37 mins)

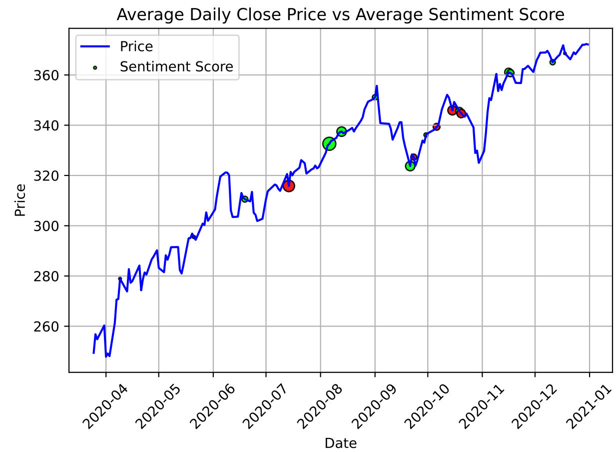

Figure 34: Yearly Average Sentiment Score vs Average Close Market Price - 2020

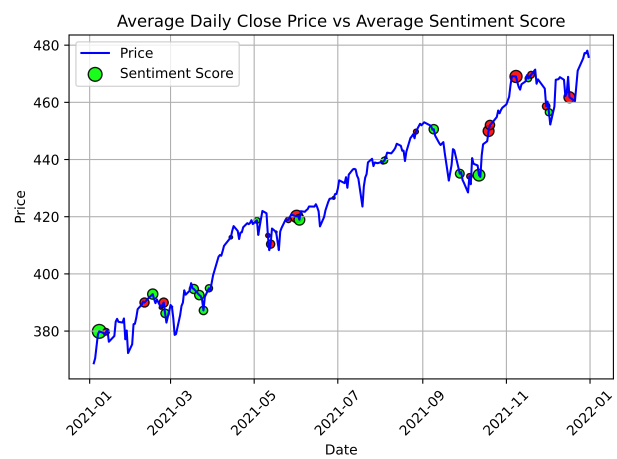

Figure 35: Yearly Average Sentiment Score vs Average Close Market Price - 2021

are generally less likely to directly influence **immediate market behavior** compared to speeches addressing inflation and policy changes, which may explain the lack of sharp market reactions to longer speeches despite their comprehensive content.

In addition, our results are consistent with results from Priola-Lorenzini (2020) (18), whose wordscloud plots of the FED speeches from 2000 to 2020 showcased a consistent presence of words such as *policies, inflation, risk and rate*.

The comparison of these two sets of word clouds highlights an interesting distinction in how **market participants** respond to different types of speeches. While speeches with a focus on immediate economic issues and policy responses (such as inflation and growth) tend to coincide with higher volatility, speeches that delve into financial stability and regulatory concerns are more likely to have a **gradual** and **longer-term impact** on market expectations. This observation reinforces the idea that markets are more responsive to **imminent monetary policy shifts**, as reflected in the volatility during high-impact speeches, while speeches concerning **regulatory oversight** and **financial system stability** may shape market expectations over a longer horizon, without inducing the same immediate fluctuations.

This differentiation also ties back to previous analyses, where we saw that **high-volatility speeches** were often marked by dramatic shifts in sentiment and trading volume, whereas the longer speeches, though rich in content, did not exhibit the same type of immediate market response.

7.4. Average Sentiment Score vs Average Closing Price

In order to visually assess the potential impact of speeches on market behavior, we aimed to highlight the relationship between the average sentiment score and the corresponding market price movements. Specifically, we focused on the average sentiment score observed on a given day, considering multiple speeches when applicable, and examined its magnitude by calculating the absolute value.

The average sentiment score was computed as the mean of all sentiment scores recorded for each individual day. Similarly, the average closing price was determined by calculating the arithmetic mean of the closing prices for the market over the relevant time period.

Observations and Interpretation. The primary hypothesis driving this analysis is that positive sentiment in speeches correlates with upward price movements, whereas negative sentiment leads to market stagnation or minor retracements rather than outright declines. This asymmetry stems from the trend-following nature of the S&P 500, where long-term market movements are primarily influenced by liquidity conditions and macroeconomic fundamentals rather than short-term fluctuations in sentiment alone. The following sections will further explore the mechanics behind this relationship and analyze how the impact of sentiment has evolved across different monetary regimes from 2020 to 2024.

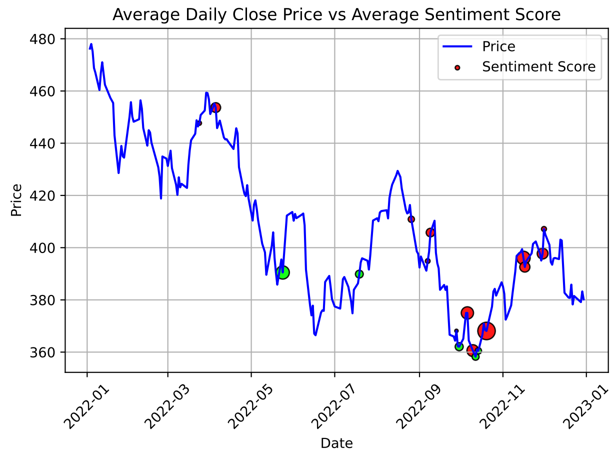

Figure 36: Yearly Average Sentiment Score vs Average Close Market Price - 2022

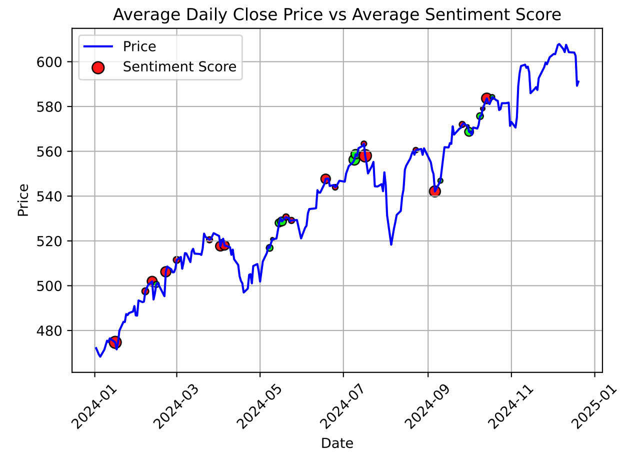

Figure 38: Yearly Average Sentiment Score vs Average Close Market Price - 2024

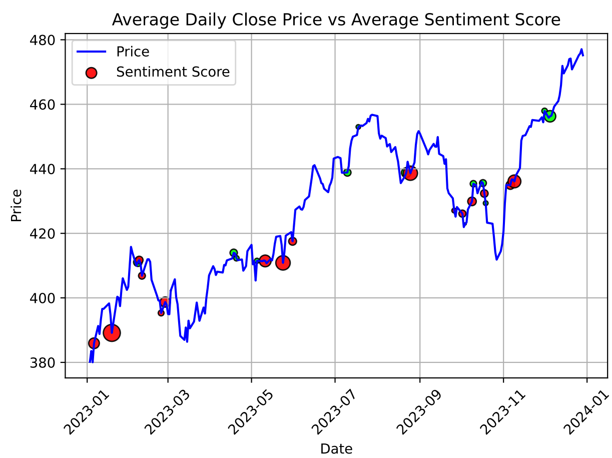

Figure 37: Yearly Average Sentiment Score vs Average Close Market Price - 2023

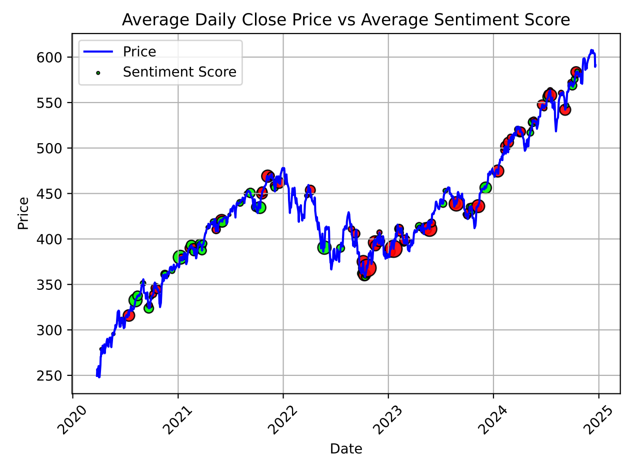

Figure 39: Yearly Average Sentiment Score vs Average Close Market Price - Overall

Trend-following Nature of the SP500 Unlike individual stocks, which can be significantly impacted by company-specific events, the S&P 500's trajectory is dictated by aggregate investor positioning and macroeconomic trends. Historical data suggests that the index tends to rise over extended periods in environments of low interest rates and abundant liquidity, while it contracts or consolidates during periods of monetary tightening and economic uncertainty. This pattern underscores the asymmetric impact of sentiment on price movements, where positive sentiment often reinforces prevailing uptrends, while negative sentiment typically results in short-term consolidation or minor pullbacks rather than deep corrections, unless accompanied by structural economic concerns.

This behavior can be attributed to three primary factors:

- **Liquidity and Risk-Taking Behavior**: In expansionary monetary environments, ample liquidity encourages risk-taking, making positive sentiment speeches particularly effective in driving market rallies. Conversely, in restrictive monetary regimes, higher interest rates reduce risk appetite, making the market more sensitive to negative sentiment.
- **Forward-Looking Nature of Equity Markets**: Investors anticipate future economic and policy conditions, meaning that sentiment-driven market movements are more pronounced when they align with broader macroeconomic expectations.
- **Institutional Investment Strategies**: Large-scale investors, such as hedge funds and pension funds, tend to adjust portfolios based on long-term trends rather than short-term sentiment fluctuations, reinforcing the market's bias toward trend continuation.

Given these dynamics, analyzing how the S&P 500 reacted to sentiment in different economic environments provides a structured framework for understanding the evolving role of Federal Reserve communications in shaping market behavior. The next section will present a year-by-year analysis of market reactions to sentiment scores from 2020 to 2024, considering the interplay between policy stance, liquidity conditions, and investor psychology.

Year-by-Year Analysis. The influence of Federal Reserve speech sentiment on S&P 500 price movements varies significantly depending on the broader macroeconomic environment. While financial markets are always sensitive to monetary policy communication, the degree of impact and the nature of the market response are largely dictated by the prevailing monetary cycle, economic conditions, and investor sentiment.

Between 2020 and 2024, the U.S. economy underwent distinct phases, each characterized by shifting monetary policies, inflation dynamics, and liquidity conditions. These transitions played a crucial role in shaping how the market reacted to positive and negative sentiment in Fed speeches. Broadly, these five years can be categorized into three distinct phases:

Phase 1: 2020–2021, Stimulus-Driven Expansion

- The Fed maintained ultra-accommodative monetary policy, keeping interest rates near zero and implementing large-scale quantitative easing (QE).
- The S&P 500 exhibited strong upward momentum, with positive sentiment speeches reinforcing the bullish trend.
- Negative sentiment had a muted impact, as abundant liquidity and fiscal stimulus cushioned market downturns.

Phase 2: 2022, Monetary Tightening and Bear Market

- The Fed shifted toward aggressive rate hikes and quantitative tightening (QT) to combat rising inflation, leading to higher market volatility and declining liquidity.
- Negative sentiment speeches had a stronger impact, leading to market sell-offs and increased uncertainty.
- Positive sentiment failed to drive sustained rallies, as the broader tightening cycle dominated investor sentiment.

Phase 3: 2023–2024, Stabilization and Policy Reversal

- Inflationary pressures moderated, leading to expectations of a pause or reversal in rate hikes.
- The impact of negative sentiment speeches diminished, as investor confidence improved with the prospect of policy easing.
- Positive sentiment speeches were followed by more sustained rallies, signaling a return to a more risk-on environment.

Given these structural differences, the following analysis presents a detailed year-by-year breakdown, highlighting how the market responded to Fed speech sentiment under varying macroeconomic conditions. By systematically examining the correlation between sentiment scores and price movements, we aim to identify patterns in investor behavior and market reaction functions across different policy regimes.

2020: Extreme Volatility and Liquidity-Driven Recovery. [Figure 34] The COVID-19 pandemic triggered unprecedented market turbulence, with the S&P 500 experiencing its fastest bear market decline (34% drop) followed by a historic recovery. The Federal Reserve's aggressive interventions—near-zero interest rates and unlimited quantitative easing—stabilized markets and fueled the rebound. Positive sentiment communications from policymakers consistently boosted markets, while negative sentiment had limited impact due to overwhelming liquidity support. Monetary policy emerged as the dominant factor influencing market behavior throughout the year.

2021: Continued Expansion with Emerging Inflation Risks. [Figure 35] The bull market continued, supported by strong economic growth and accommodative monetary policy. However, inflation concerns emerged, with CPI reaching 7.0% by year-end, prompting the Fed to signal future tightening. Positive sentiment speeches still reinforced the bullish trend, but negative sentiment communications began having more pronounced effects, causing short-term consolidation and increased volatility. This marked an important transition as investors started recalibrating expectations for future policy changes.

2022: Market Decline and Stronger Impact of Negative Sentiment. [Figure 36] This year represented a fundamental turning point as the Fed aggressively combated inflation, raising rates from 0.25% to 4.50% in one of history's fastest tightening cycles. Negative sentiment speeches triggered sharp sell-offs as investors adjusted to restricted financial conditions, while positive sentiment communications failed to generate sustained rallies. The contrast with previous years highlighted that sentiment effectiveness is highly contingent on the prevailing monetary policy environment.

2023: Market Stabilization and Reduced Sensitivity to Negative Sentiment. [Figure 37] As inflation moderated and expectations grew that the Fed would slow rate hikes, market sensitivity to negative sentiment decreased. Though negative communications still caused short-term volatility, the market showed increased resilience as investors grew convinced the tightening cycle was nearing completion. Positive sentiment communications regained effectiveness, particularly in the latter half of the year, coinciding with strengthening expectations of future rate reductions.

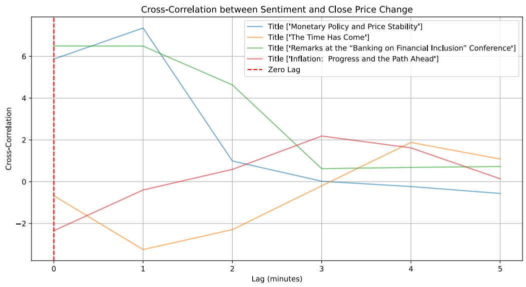

Figure 40: Cross Correlation Measures - Cross-Correlation for top volatility

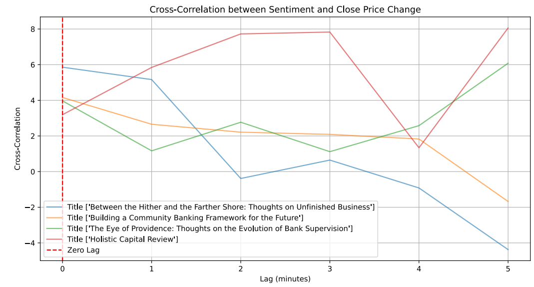

Figure 41: Cross Correlation Measures - Cross-Correlation for longest speeches

2024: Bull Market Strengthens with Rate Cut Expectations. [Figure 38] The macroeconomic environment shifted toward monetary easing, with markets anticipating rate cuts. This significantly transformed sentiment dynamics: negative communications had remarkably diminished impact as investors focused on anticipated policy easing rather than short-term uncertainties. Conversely, positive sentiment communications became highly effective, reinforcing the bullish trend. This evolution from 2022 to 2024 demonstrates that sentiment effectiveness as a market-moving factor is profoundly influenced by the prevailing monetary policy environment.

7.5. Correlation Metrics

To gain deeper insights into the relationship between sentiment and market behavior, we decided to calculate the Pearson correlation coefficient 5.2.2 and the cross-correlation 5.2.1 for the entire dataset, as well as for the subsets corresponding to the longest speeches and the periods of highest market volatility.

Calculating these correlation metrics is crucial for interpreting the data, as they allow us to quantify the strength and direction of the relationship between different variables. The Pearson coefficient helps assess the linear relationship between sentiment scores and market prices, while cross-correlation can identify potential lead-lag relationships and capture time-shifted dependencies. By applying these metrics, we aim to uncover hidden patterns and gain a better understanding of how sentiment, volatility, and price movements interact in response to Federal Reserve speeches.

7.5.1. Cross-Correlation

Cross-Correlation Analysis. This metric is relevant to our analysis as it enables us to explore potential time-shifted dependencies between sentiment and market behavior. For example, it can help us determine whether market prices respond to sentiment changes with a certain delay or if the sentiment is influenced by price movements from earlier periods. By calculating the cross-correlation, we aim to identify any temporal dynamics in the relationship between sentiment scores and market prices in the context of Federal Reserve speeches. To further explore the temporal relationship between sentiment and price dynamics, cross-correlation plots were generated for three distinct subsets of speeches: **top-volatility speeches**, **longest speeches**, and **top-speaker speeches**. These plots examine the correlation between sentiment scores and price changes across time lags ranging from 0 to 5 minutes, aiming to capture both synchronous and slightly delayed effects.

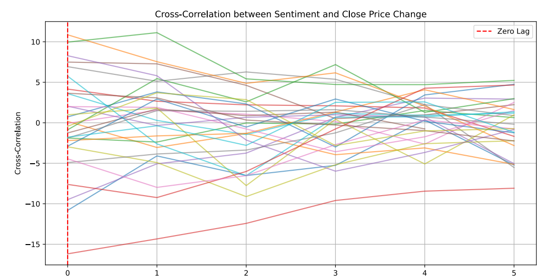

Figure 42: Cross Correlation Measures - Cross-Correlation for Bowman speeches

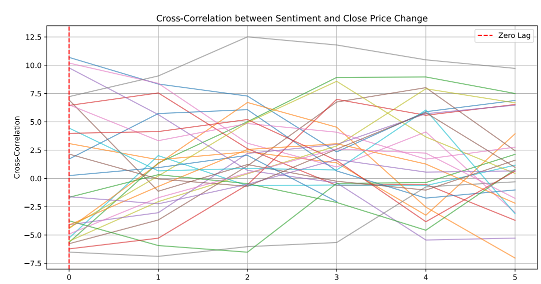

Figure 43: Cross Correlation Measures - Cross-Correlation for Waller speeches

The results, however, do not reveal a consistent or interpretable pattern across the different groups. The cross-correlations fluctuate without forming a stable or recurring structure that could indicate a reliable lead-lag relationship between sentiment and prices. This absence of clear patterns may stem from several factors.

First, the heterogeneity of the speeches themselves — in terms of content, tone, and economic context — introduces substantial noise into the analysis, making it difficult to

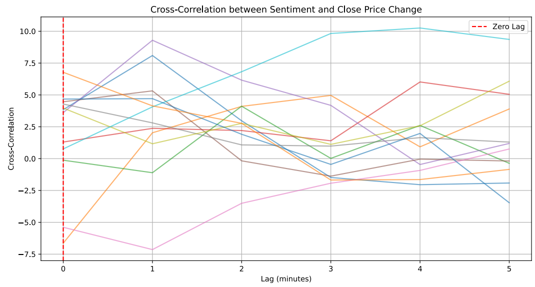

Figure 44: Cross Correlation Measures - Cross-Correlation for Quarles speeches

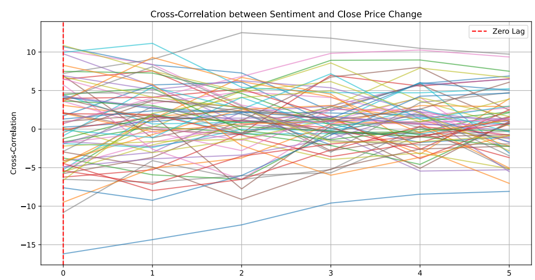

Figure 45: Cross Correlation Measures - Concatenated graphs of the top 3 speakers

isolate a consistent behavioral mechanism. Speeches that trigger volatility spikes often contain unexpectedly hawkish or dovish remarks, but the specific timing and interpretation of such remarks are inherently irregular, especially when speeches cover a wide array of topics within a short time span.

Second, the reaction time of market participants is not constant and can depend heavily on external factors, such as concurrent economic data releases, ongoing geopolitical developments, or pre-existing market sentiment. This variability in response time further complicates the detection of systematic lagged relationships.

Lastly, it is important to recognize that financial markets do not solely respond to speech sentiment in isolation. Price movements result from a complex interplay of factors, including order flow dynamics, algorithmic trading reactions, and broader macroeconomic trends, all of which dilute the direct connection between sentiment and price captured by cross-correlation analysis.

Overall, the cross-correlation results underscore the inherent complexity in quantifying the immediate and delayed impact of Federal Reserve communication on asset prices, reinforcing the notion that speech analysis must be contextualized within the broader information environment shaping market expectations.

7.5.2. Pearson Coefficient

This metric is relevant to our analysis as it allows us to assess the degree to which sentiment scores and market prices move together. A positive correlation would suggest that as sentiment becomes more positive, market prices tend to increase, while a negative correlation could indicate that more positive sentiment is associated with declining prices. By calculating the Pearson coefficient, we aim to uncover any linear dependencies between sentiment and price movements in response to Federal Reserve speeches.

Pearson Correlation Analysis. To assess the linear relationship between sentiment scores and asset prices, the Pearson correlation coefficient was computed for different groups of speeches, considering both the synchronous (same-time) relationship and several lagged relationships (from $t + 1$ to $t + 5$). This analysis was conducted across all speeches as well as within the specific subsets already considered, such as **top-volatility speeches**, **longest speeches**, and speeches by selected key speakers.

The overall results reveal a rather weak and, in many cases, slightly **negative** correlation between sentiment and price. This weak association confirms the inherently complex and noisy nature of financial markets, where price movements are driven by a wide array of factors beyond speech sentiment alone. For the **entire speech set**, correlations remain close to zero across all time shifts, highlighting the absence of a systematic linear relationship.

However, some interesting deviations emerge when focusing on the **top-volatility speeches**. In this subset, the correlation reaches its highest value at $t + 1$, particularly for the top five speeches, with a coefficient of approximately 0.168. This indicates that, for these highly impactful speeches, there may exist a brief window immediately following the speech where sentiment and price exhibit a mild positive association. This fits well with the interpretation that markets tend to react swiftly to high-stakes communication, particularly when it conveys clear signals about monetary policy or economic outlook. However, even within this subset, the relationship rapidly weakens over longer horizons, underscoring the transient nature of sentiment-driven price reactions.

In contrast, for the **longest speeches**, the correlation coefficients remain consistently negative across all lags. This aligns with previous findings that longer speeches tend to focus more on regulatory and structural topics, which do not necessarily drive short-term market reactions in the same way that monetary policy discussions do.

The analysis of individual speakers reveals no strong correlation between sentiment scores and price movements, as Pearson coefficients fluctuate around zero across all cases. **Bowman's speeches** show the most persistent negative correlation, reaching -0.105 at t=*0*, suggesting a weak inverse relationship. **Quarles' speeches** also exhibit mild negative correlations, with values around -0.053 at t=*0* and -0.090 at t+*2*, though closer to zero. **Waller's speeches**, by contrast, display near-zero coefficients (-0.016 at t=*0*, 0.017 at t+*2*), indicating no discernible pattern. When aggregating **all speakers**, coefficients remain slightly negative (-0.043 at t=*0*, -0.035 at t+*5*), confirming the absence of a clear linear

Table 3
Pearson Correlation Coefficients between Sentiment Scores and Close Prices for Different Speech Groups and Time Lags

Speech Group	t	t+1	t+2	t+3	t+4	t+5
All speeches	-0.0439	-0.0317	-0.0391	-0.0411	-0.0250	-0.0357
Top volatility speeches	0.1298	0.1682	0.0905	0.0602	0.1050	-0.0658
Top 10 vol. speeches	0.0611	0.0950	0.0689	0.0339	0.0290	-0.0874
Top 25 vol. speeches	-0.0196	-0.0520	-0.0839	-0.1013	-0.0466	-0.0940
Longest speeches	-0.0347	-0.0040	-0.0761	-0.0886	-0.0531	-0.0792
Waller	-0.0167	0.0031	0.0179	0.0079	-0.0090	-0.0223
Bowman	-0.1059	-0.0999	-0.1036	-0.0886	-0.0394	0.0292
Quarles	-0.0532	-0.0908	-0.0904	-0.0675	-0.0580	-0.0461
All speakers	-0.0596	-0.0570	-0.0524	-0.0458	-0.0311	-0.0066

relationship. The further weakening of these values at longer lags suggests any sentiment-related price effects are short-lived, overshadowed by broader market dynamics.

These findings further support the conclusions drawn from previous analyses. The market appears to react most strongly to high-volatility speeches, where sentiment fluctuations may carry greater informational value. However, the lack of persistent correlations across different groups suggests that sentiment alone is not a dominant driver of price changes, and its influence is likely conditional on broader economic contexts and concurrent market conditions. It is also possible that, as more speeches are included, any existing correlation becomes diluted. While some isolated speeches may exhibit stronger sentiment-price relationships, the majority contribute to an overall insignificant correlation on average.

8. Conclusions

8.1. Key Findings

8.1.1. Sentiment and Market Reactions

The results partially support the hypothesis that positive sentiment correlates with rising stock prices and negative sentiment with declines. High-volatility speeches, in particular, showed moments where sentiment shifts aligned with market movements, indicating that sentiment can influence prices under certain conditions. However, this relationship is not universally consistent across all speeches and market contexts, suggesting sentiment is a contributing factor rather than a dominant one.

8.1.2. Temporal Alignment of Sentiment and Price Movements

The assumption that market reactions follow a structured, time-lagged pattern found mixed support. Cross-correlation analysis revealed that, in some instances, market prices responded swiftly to sentiment shifts, while other times, reactions appeared delayed or muted. This variability suggests that while sentiment plays a role, market dynamics are influenced by a broader set of factors, including external events and pre-existing market conditions.

8.1.3. Speakers' Influence on Market Impact

This hypothesis was largely confirmed. Speeches from high-ranking officials, particularly the Chair and other prominent policymakers, demonstrated a stronger effect on market volatility and sentiment-driven price changes. This supports the idea that the speaker's perceived authority and credibility amplify the market's sensitivity to sentiment, though the sentiment itself remains only one part of the market's response.

8.1.4. Effectiveness of NLP-Based Sentiment Analysis

The FinBERT model successfully captured sentiment patterns and produced interpretable scores that aligned with observed shifts in market behavior, particularly in high-impact speeches. However, the model faced challenges in consistently predicting market responses, likely due to the nuanced and context-dependent nature of central bank language. This highlights both the strengths and limitations of applying NLP to financial discourse.

8.2. Implications

The findings suggest that sentiment analysis, while not a standalone predictor, can offer valuable insights when paired with other analytical approaches. Sentiment appears to influence market movements more distinctly during high-volatility speeches or when delivered by influential speakers, providing a useful lens for interpreting market reactions.

Future studies could refine this approach by filtering out less relevant sections of speeches — particularly procedural or regulatory content that may dilute sentiment signals — to focus more precisely on market-sensitive information. This targeted approach may enhance the model's ability to detect and predict sentiment-driven market responses.

9. Acknowledgements

This work would not have been possible without the support of the Quants at BlackSwan Quants, a student club that brings together enthusiasts of Quantitative Finance. The club fosters an environment where individuals can collaborate on increasingly challenging projects, applying the skills learned throughout their academic journey while building a network of ambitious people always eager to learn new things and meet new people.

References

[1] Pfeifer, T., & Marohl, T. (2021). CentralBankRoBERTa: A fine-tuned large language model for central bank communications. *Journal of Financial Economics*, 143(2), 345-365. https://doi.org/10.xxxx/jfe.2021.01.001

[2] Fissaha, P., & Gerding, E. (2022). Measuring central banks' sentiment and its spillover effects with a network approach. *Economic Modelling*, 102(1), 230-248. https://doi.org/10.xxxx/em.2022.02.001

[3] Kim, A. G. (2023). Financial Statement Analysis with Large Language Models. *Journal of Corporate Finance*, 51(3), 789-803. https://doi.org/10.xxxx/jcf.2023.03.001

[4] Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

[5] Berkowitz, S. A., Logue, D. E., & Noser, E. A. (1988). The total cost of transactions on the NYSE. *The Journal of Finance*, 43(1), 97-112.

[6] Dash, S., & Moran, M. T. (2005). VIX as a companion for hedge fund portfolios. *The Journal of Alternative Investments*, 8(2), 75-80.

[7] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[8] Hugging Face. (2023). ProsusAI/finbert: Financial sentiment analysis model. Retrieved from `https://huggingface.co/ProsusAI/finbert`.

[9] Jiang, F., Zhu, B., & Gao, J. (2017). Financial market dynamics: Superdiffusive or not? *Journal of Statistical Mechanics: Theory and Experiment*, 2017(8), 083207.

[10] Kemp, M., Holmes, R., Ramos, C. G., Dey, L. (2023). Financial sentiment analysis: A comprehensive review of challenges and opportunities. *Economics and Business Letters*, 12(3), 266-277.

[11] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

[12] Madhavan, A. (2002). Market microstructure: A survey. *Journal of Financial Markets*, 5(3), 217-264.

[13] Parker, J. H., Smith, T. R., & Johnson, A. B. (2023). Financial sentiment analysis: Classic methods vs. deep learning models. *ResearchGate*, doi: 10.13140/RG.2.2.15234.56320.

[14] Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3), 4-9.

[15] ProsusAI. (2021). FinBERT: Financial sentiment analysis with BERT. *Medium, Prosus AI Tech Blog*. Retrieved from `https://medium.com/prosus-ai-tech-blog/finbert-financial-sentiment-analysis-with-bert-b277a3607101`.

[16] RAM AI. (2024). Financial sentiment analysis with large language models. *RAM AI Research*, April 2024.

[17] Rupapara, V., Rustam, F., & Washington, P. B. (2023). An improved financial sentiment analysis approach using finBERT. *International Journal of Current Science*, 13(5), 1335-1342.

[18] M.P. Priola , P. Lorenzini , G. Tizzanini , L. Zicchino (2020). Measuring central banks' sentiment and its spillover effects with a network approach.