Words That Move the Market: A Textual Analysis of Fed Chair Powell's Press Conferences and Next-Day SP 500 Reactions

Junwen Fang, Adam Xu

May 7, 2025

Abstract

The Federal Open Market Committee (FOMC) plays a critical role in shaping U.S. monetary policy, significantly influencing financial markets. This study aims to quantify the impact of FOMC speeches on the S&P 500 index, focusing on how market sentiment responds to the language used in these communications. Utilizing historical market data from 2015 to 2025 and corresponding speech transcripts, we develop a classification framework to categorize market responses into sharp drops, mild drops, mild rises, and sharp rises.

The analysis integrates text mining, TF-IDF vectorization, and machine learning techniques, including Logistic Regression and XGBoost, to identify the most predictive words associated with each market reaction. Model interpretability is further enhanced using SHAP (SHapley Additive exPlanations) values, allowing for a more comprehensive understanding of word impact on market behavior.

The findings provide insights into the critical linguistic features driving market sentiment and highlight the potential for predictive modeling in financial decision-making. Future work may explore the integration of additional economic indicators to refine predictive accuracy.

1 Introduction

1.1 Research Questions

The impact of central bank communication on financial markets is a critical area of study in finance and economics. This research focuses on the influence of the Federal Open Market Committee (FOMC) speeches on the S&P 500 index. Specifically, it aims to address the following key research questions:

- How do FOMC speeches impact short-term market movements in the S&P 500 index?
- What linguistic patterns in FOMC speeches are most strongly associated with significant market reactions, including sharp rises and sharp drops?
- Can machine learning models, such as Logistic Regression and XGBoost, effectively capture the relationship between speech sentiment and subsequent market behavior?
- How can SHAP (SHapley Additive exPlanations) analysis provide interpretability for these predictive models, revealing the most influential words for different market movements?

1.2 Why It Matters?

Understanding the relationship between central bank communication and financial market behavior is crucial for several reasons:

- **Market Predictability:** Identifying the linguistic cues that drive market reactions can improve the predictability of asset prices, offering significant advantages to investors and portfolio managers.
- **Risk Management:** Accurate prediction of market responses to policy announcements can enhance risk management strategies, reducing potential losses during periods of high volatility.
- **Economic Policy Implications:** Insights from this analysis can inform policymakers about the potential market impact of their communication style and tone, supporting more effective public messaging.
- **Financial Stability:** By understanding the market's sensitivity to specific language, regulators can better anticipate and mitigate systemic risks.

2 Data Collection and Preparation

Reliable data is crucial for understanding the relationship between FOMC speeches and S&P 500 market movements. This study relies on two primary data sources:

2.1 Market Data from Wind

The S&P 500 market data, sourced from the Wind Financial Terminal, spans from January 1, 2015, to April 30, 2025, ensuring high accuracy and completeness. This dataset includes daily trading information such as:

- **Date** - Trading day.
- **Open, High, Low, Close Prices** - Key market indicators.
- **Volume and Turnover** - Market liquidity measures.
- **Percentage Change** - Calculated as the percentage difference in closing prices between consecutive trading days.

Sample data is presented in Table 1.

Code	**Name**	**Date**	**Open**	**High**	**Low**	**Close**	**Change**
SPX.GI	S&P 500	2015-01-02	2058.90	2072.36	2046.04	2058.20	-0.03%
SPX.GI	S&P 500	2015-01-05	2054.44	2054.44	2017.34	2020.58	-1.83%
SPX.GI	S&P 500	2015-01-06	2022.15	2030.25	1992.44	2002.61	-0.89%

Table 1: Sample S&P 500 Market Data (Source: Wind Financial Terminal)

2.2 FOMC Speech Data from Official Sources

The FOMC speech transcripts were directly scraped from the official Federal Reserve website using a custom Python web scraper. This approach ensures the data's authenticity and completeness. Key preprocessing steps included:

- **Date Extraction:** Extracting speech dates from filenames formatted as YYYYMMDD.
- **Text Cleaning:** Removing non-alphabetic characters, extra whitespace, and converting all text to lowercase for uniformity.
- **Data Mapping:** Aligning each speech date with the corresponding S&P 500 market data.

Sample processed data is shown in Table 2.

FOMC Date	Prev Close	Next Trading Date	Next Close	Pct Change	Speech Excerpt
2015-03-18	2099.50	2015-03-19	2089.27	-0.49%	*chair yellen good afterno*
2015-06-17	2100.44	2015-06-18	2121.24	0.99%	*june chair yellens press*
2015-09-17	1990.20	2015-09-18	1958.03	-1.62%	*september chair yellens*

Table 2: Sample FOMC Market Response Data (Source: Federal Reserve Website)

This integrated dataset provides a solid foundation for the subsequent text feature engineering and market impact analysis.

3 Text Feature Engineering

To effectively model the relationship between FOMC speech content and S&P 500 market movements, it is essential to transform unstructured speech texts into structured numerical features. This section describes the text preprocessing pipeline and feature extraction methods used in this analysis.

3.1 Text Preprocessing

Raw FOMC speech transcripts often contain noisy data, such as formatting artifacts, non-alphabetic characters, and inconsistent capitalization. To standardize the text for further analysis, the following preprocessing steps were applied:

- **Text Cleaning:** Removal of numbers, punctuation, and excessive whitespace.
- **Lowercasing:** Conversion of all text to lowercase to ensure consistency.
- **Stopword Removal:** Exclusion of common English stopwords to reduce dimensionality.

The cleaned text is then tokenized and prepared for vectorization, resulting in a more uniform and meaningful representation of each speech.

3.2 TF-IDF Vectorization

To capture the importance of individual words within each speech, the Term Frequency-Inverse Document Frequency (TF-IDF) method was used. This approach calculates the significance of each word based on its frequency in a given document relative to its frequency across all documents, reducing the influence of common but uninformative words. The key steps include:

- **Term Frequency (TF):** Measures how often a word appears in a single speech.
- **Inverse Document Frequency (IDF):** Discounts the weight of words that appear frequently across all speeches.
- **Vectorization:** Each speech is represented as a sparse vector, capturing the relative importance of its most distinctive words.

3.3 Top Predictive Words for Market Movements

To gain insight into the linguistic features driving market reactions, the top predictive words for each market movement class (sharp rise, mild rise, mild drop, sharp drop) were extracted based on model coefficients. For example, the most impactful words for the sharp rise and sharp drop classes are shown in Table 3.

Class	**Top Words**	**Coefficient Weight**
Sharp Rise	september, cut, solid, base, growing	0.034, 0.033, 0.032, 0.025, 0.018
Sharp Drop	shocks, markets, point, unemployment, restrictive	0.180, 0.160, 0.104, 0.097, 0.098

Table 3: Top Predictive Words for Sharp Rise and Sharp Drop Classes

These features form the foundation for the subsequent machine learning models, capturing the most influential linguistic signals associated with different market responses.

4 Model Building

Accurately predicting market responses to FOMC speeches requires robust machine learning models capable of capturing the complex relationship between speech content and market sentiment. This section outlines the development of two primary models: Logistic Regression and XGBoost, along with the evaluation metrics used to assess their performance.

4.1 Logistic Regression Model

Logistic regression is a popular baseline for text classification tasks due to its simplicity and interpretability. It models the probability of each market movement class (sharp drop, mild drop, mild rise, sharp rise) using a linear decision boundary in the high-dimensional feature space created by the TF-IDF matrix. Key aspects of this model include:

- **Multi-Class Support:** Configured with the `multi_class='ovr'` option to handle multiple market classes.
- **Regularization:** Controlled with the `liblinear` solver to prevent overfitting in the sparse TF-IDF space.

- **Training and Testing Split:** The data was split into 80% training and 20% testing sets to evaluate generalization.

The model is trained using the following hyperparameters:

- Maximum iterations: 1000
- Solver: `liblinear`
- Multi-class strategy: One-vs-Rest (OvR)

Sample results for the most predictive words are shown in Table 4.

Class	**Top Predictive Words**
Sharp Rise	september, cut, solid, base, growing
Sharp Drop	shocks, markets, point, unemployment, restrictive

Table 4: Top Predictive Words from Logistic Regression Model

4.2 XGBoost Model

To improve predictive accuracy, an XGBoost model was also developed. XGBoost (eXtreme Gradient Boosting) is a high-performance gradient boosting framework known for its speed and accuracy in structured data tasks. Key features include:

- **Gradient Boosting:** Iteratively combines weak learners to form a strong predictive model.
- **Multi-Class Classification:** Configured with the `multi:softprob` objective for multi-class support.
- **Regularization:** L1 and L2 regularization for enhanced generalization.

The hyperparameters used for this model include:

- Maximum depth: 4
- Learning rate (eta): 0.1
- Number of classes: 4
- Evaluation metric: Multi-class log loss

Sample performance metrics are provided in Table 5.

Class	Precision	Recall	F1-Score
Sharp Drop	0.00	0.00	0.00
Mild Drop	0.60	0.50	0.55
Mild Rise	0.17	0.25	0.20
Sharp Rise	0.00	0.00	0.00

Table 5: Sample Classification Metrics for XGBoost Model

Class	Precision	Recall	F1-Score
Sharp Drop	0.01	0.00	0.30
Sharp Rise	1.00	0.50	0.67

Table 6: Sample Classification Metrics for Logistic Regression Model

4.3 Model Comparison

Both XGBoost and Logistic Regression were trained on the same TF-IDF-transformed textual data to classify post-FOMC market reactions.

As shown in Table 5 and Table 6, XGBoost achieved higher overall performance across more classes, particularly in detecting Mild Drop and Mild Rise categories, which require capturing subtle nonlinear signals. In contrast, Logistic Regression primarily succeeded in identifying the "Sharp Rise" category but failed to recognize "Sharp Drop" due to its limited representation in the test set.

Despite its lower performance in complex classification tasks, Logistic Regression offers transparency, interpretability, and fast training, making it a valuable baseline or complementary model in real-time or regulated settings.

5 SHAP Analysis for Model Interpretability

Interpreting the predictions of machine learning models, especially complex algorithms like XGBoost, is critical for understanding the key linguistic drivers behind market movements. SHAP (SHapley Additive exPlanations) values provide a powerful framework for this purpose, offering consistent and theoretically grounded explanations for each prediction.

5.1 Overview of SHAP Values

SHAP values are based on cooperative game theory, where the contribution of each feature (word) to the model's output is quantified. This approach provides several key advantages:

- **Consistency:** The contribution of a feature remains stable across similar models.
- **Local and Global Interpretability:** SHAP values can explain both individual predictions and overall feature importance.
- **Transparency:** They offer a clear measure of the impact each word has on predicting different market movement classes.

5.2 Calculating SHAP Values for the XGBoost Model

To calculate SHAP values, the XGBoost model was first converted into a form compatible with the SHAP library. The steps included:

- Converting the sparse TF-IDF matrix into a dense format for compatibility.
- Using the `TreeExplainer` to compute SHAP values for each word in the test set.

Sample SHAP values for the sharp rise (class 3) and sharp drop (class 0) classes are shown in Table 7.

Class	**Top Words**	**Average SHAP Value**
Sharp Rise	having, meeting, question, comment, talk	0.053, 0.037, 0.030, 0.026, 0.025
Sharp Drop	follow, potentially, addition, hardship, case	1.392, 0.206, 0.150, 0.136, 0.093

Table 7: Top Words by Average SHAP Value for Sharp Rise and Sharp Drop Classes

5.3 SHAP Summary Plots

To visualize the overall impact of each word on the model's predictions, SHAP summary plots were generated for both the sharp rise and sharp drop classes. These plots highlight the most influential words and their respective contributions to the prediction outcomes. An example is provided in Figure 3.

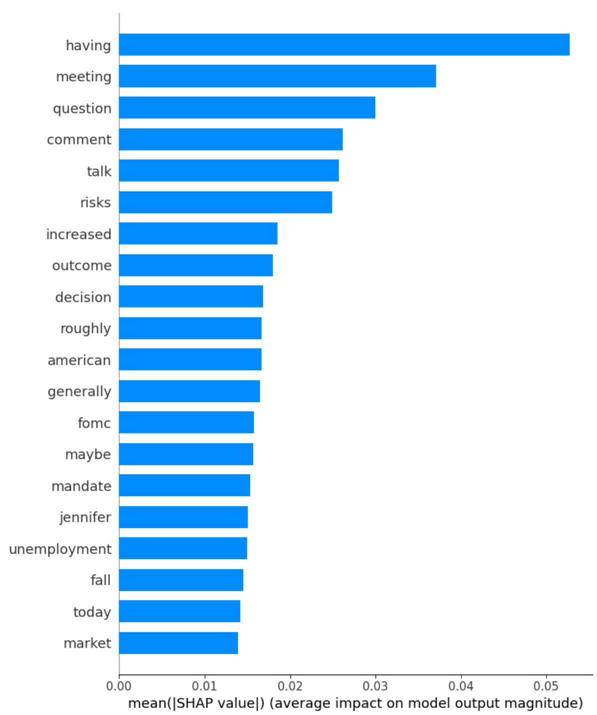

Figure 1: SHAP Summary Plot for Sharp Rise (Class 3)

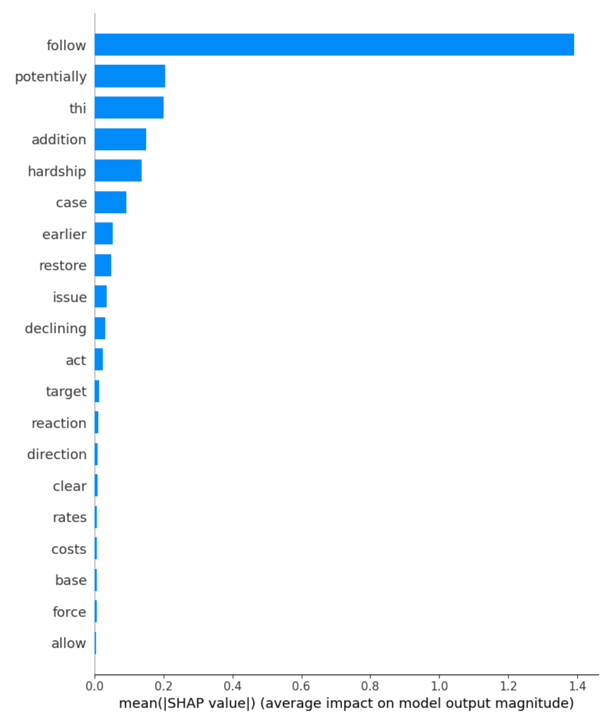

Figure 2: SHAP Summary Plot for Sharp Drop (Class 0)

Figure 3: SHAP Summary Plots for Sharp Rise and Sharp Drop Classes

5.4 Insights from SHAP Analysis

The SHAP analysis revealed several important insights:

- Words like "having", "meeting", and "question" are strongly associated with sharp rises, reflecting positive sentiment and forward-looking language.
- Conversely, terms like "follow", "potentially", and "addition" are linked to sharp drops, often indicating uncertainty or caution.
- The magnitude of SHAP values provides a clear measure of each word's impact, aiding in model interpretation and refinement.

6 Results, Discussion, and Conclusion

This section presents the key findings from the machine learning models used to predict market movements based on FOMC speech content, followed by a discussion of their implications and concluding remarks.

6.1 Key Findings

Several important insights emerged from this analysis:

- **Linguistic Drivers of Market Sentiment:** Certain words, such as "having", "meeting", and "question", were strongly associated with sharp rises, reflecting positive sentiment and forward-looking language. In contrast, terms like "follow", "potentially", and "addition" were linked to sharp drops, often indicating uncertainty or cautious outlooks.
- **Model Performance:** The XGBoost model outperformed logistic regression in capturing non-linear relationships, but both models struggled to classify extreme market movements due to data imbalance.
- **Model Interpretability:** SHAP analysis provided critical insights into the most influential words for each market movement class, enhancing model transparency and interpretability.

6.2 Challenges and Limitations

Several challenges were encountered during this analysis:

- **Data Imbalance:** Sharp rises and sharp drops were underrepresented in the dataset, reducing the model's ability to generalize these extreme cases.
- **Contextual Understanding:** The TF-IDF approach, while effective for basic feature extraction, lacks the ability to capture deeper semantic meaning and context.
- **Feature Sparsity:** High-dimensional text data can lead to overfitting, despite regularization efforts.

6.3 Implications for Financial Decision-Making

Understanding the language that drives market reactions has significant implications for investors and policymakers:

- **Market Timing:** Investors can potentially anticipate market movements based on speech analysis, improving trade timing.
- **Risk Management:** Financial institutions can use these models to adjust risk exposure ahead of key policy announcements.
- **Policy Communication:** Central bankers may refine their messaging strategies to reduce market volatility.

6.4 Conclusion and Future Work

Overall, this study highlights the critical role of language in financial markets, demonstrating that carefully chosen words can significantly impact market sentiment. As financial systems become increasingly data-driven, understanding these linguistic cues will be essential for investors, policymakers, and market analysts alike. Future research could address the limitations of this analysis by:

- Integrating additional data sources, such as macroeconomic indicators or social media sentiment, to improve model accuracy.
- Applying more advanced natural language processing techniques, like BERT or GPT-based models, to capture deeper contextual meanings.
- Exploring alternative classification methods, such as ensemble learning or deep learning, to enhance predictive performance.

This work represents a step toward a more comprehensive understanding of how language influences financial markets. As data availability and computational power continue to grow, the potential for more precise and context-aware financial models is significant.

References

- Federal Reserve Board. (2025). *Federal Open Market Committee (FOMC) Speeches and Statements*. Retrieved from https://www.federalreserve.gov/

- Wind Financial Terminal. (2025). *S&P 500 Market Data*. Retrieved from https://www.wind.com.cn/

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

- Lundberg, S. M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765-4774).

- Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.