Factor Analysis and Visualization of U.S. Stock Market Data

Kai Pang, Dingyuan Xu, Jinyan She, Ling Zhou, Yuhan Jiang, Dawei Yu
CSE6242: Data & Visual Analytics Team 028

1. Introduction

In this project, we focus on analyzing the impact of key market factors (interest rates, economic indicators, FOMC decisions) on the stock market, and develop interactive visualization tools to help users explore and interpret them. By combining advanced data analysis with intuitive visuals, we hope to empower investors and analysts to better understand market dynamics and make data-driven decisions.

2. Problem Definition

Monetary policy decisions by the Federal Reserve significantly influence financial markets, yet the sector-specific impacts and dynamic relationships between policy changes and stock market performance remain complex and challenging to quantify. Traditional financial analyses often focus on broad market indices or rely on static models that fail to capture the nuanced effects across different sectors and time horizons. Existing research tends to analyze either short-term market reactions to policy changes or long-term economic trends, often overlooking the interplay between immediate and prolonged effects. Furthermore, many studies treat monetary policy shifts as linear events, disregarding potential asymmetries in sectoral responses during tightening and easing cycles. The lack of an integrated, multi-frequency analytical framework limits the ability of investors and analysts to make data-driven decisions in response to evolving policy environments.

To address these gaps, we propose a comprehensive analytical framework that integrates event studies, time series modeling, and machine learning techniques to examine the impact of Federal Reserve policy on stock market sectors. By leveraging a diverse dataset—including interest rates, economic indicators, FOMC decisions, and market returns, we aim to identify sector-specific sensitivities, quantify the timing and magnitude of market reactions, and develop predictive models for sector performance.

Additionally, we introduce an interactive visualization system that enables users to intuitively explore the relationship between monetary policy and market behavior. This tool provides a more accessible and actionable way to interpret complex policy-market dynamics, helping investors and analysts make informed strategic decisions.

3. Literature survey

Numerous studies have explored methods for understanding and predicting stock market performance, employing a wide range of statistical, econometric, and machine learning approaches. Traditional econometric techniques include linear, nonlinear, and time-varying causality tests to identify relationships between market variables, though they often lack real-time data processing and interactive visualization capabilities [1]. Several works have investigated the influence of macroeconomic factors, such as oil prices and interest rates, on stock market performance. For example, [2] distinguishes between demand and supply shocks in the global crude oil market but focuses solely on oil, neglecting other economic indicators. Similarly, [7] uses quantile regression to examine oil-stock dynamics in emerging markets, while [8] applies a VAR model to study the long-term relationships among oil prices, exchange rates, and stock prices—mostly limited to U.S. market data. Another macro-focused study [14] analyzes how oil price fluctuations affect stock returns in Pakistan, underscoring the impact of broader economic conditions. To capture volatility, [9] examines the effect of macroeconomic announcements on aggregate stock returns, while [10] and [17] apply factor analysis to identify significant market indicators and clusters of securities, respectively. However, these models often lack sector-specific insights and practical investor tools. Relatedly, [15] introduces "stock vectors" for handling multi-stock high-dimensional data, but without integrating visualization or interpretability mechanisms. In the machine learning domain, studies have advanced predictive accuracy through model innovation and sentiment integration. For example, [4] incorporates Hidden Markov Models within Bi-LSTM networks to analyze financial news sentiment, showing that sentiment indices improve forecast accuracy. Similarly, [16] compares LSTM and GRU architectures, revealing that sentiment-enhanced models outperform standard ones, and proposes a cooperative deep learning approach for dynamic prediction. [5] introduces DeepARA, a probabilistic model using attention mechanisms to weight time points dynamically, and [6] combines GANs with transformer-based attention to generate synthetic data and extract key market indicators. These approaches highlight the predictive strength of deep models, but they often lack contextual interpretability and interactive exploration features. Sentiment analysis continues to gain traction in the literature. [12] proposes a hybrid model combining technical indicators with sentiment data for improved forecasting, while [13] demonstrates that sentiment-based models outperform traditional bag-of-words approaches. Our project extends these findings by incorporating domain-specific sentiment dictionaries and interactive visualizations to enhance interpretability.

On the technical analysis front, [18] explores the use of heuristic models to optimize technical indicator inputs for forecasting, showing that support vector regression (SVR) and multilayer perceptron (MLP) perform comparably, though heuristic methods often outperform conventional techniques. Feature selection also plays a critical role, as shown in [11], which applies ML-based selection to improve stock prediction accuracy—though explain ability remains limited.

Overall, the existing literature provides a rich foundation of methodologies for market analysis and prediction. However, key gaps remain limited real-time interactivity, lack of holistic integration across macroeconomic, technical, and sentiment features, and minimal emphasis on explainable models tailored to individual investors. Our project addresses these gaps by developing an interactive visualization tool supported by explainable machine learning models, enabling more accessible and actionable insights into stock market behavior.

4. Implemented Method

4.1 Data Collection

1. Stock Market Data(from Kaggle): Historical data for all constituent stocks of the S&P 500 and all Select Sector SPDR ETFs representing the 11 Global Industry Classification Standard (GICS) sectors.
2. Federal Reserve-Related Data(from Federal): Federal Funds Rate, 10-Year Treasury Constant Maturity Rate, GDP Growth Rate, Unemployment Rate.
3. Federal Open Market Committee (FOMC) Meeting Data(from Federal): Rate decisions in basis points, surprise component.

4.2 Analytical Framework

Our analytical approach combines event studies, time series analysis, and machine learning techniques to extract insights from the integrated dataset.

1. Event Study Methodology: We employed an event study methodology to isolate the impact of monetary policy announcements on sector returns. The event window spans from 5 days before to 20 days after each FOMC meeting. For each sector, we calculated cumulative abnormal returns (CAR) relative to a market model and examined statistical significance using bootstrap resampling techniques. This approach allowed us to quantify the differential impact of policy decisions across sectors while controlling for broad market movements.
2. Time Series Analysis: A Vector autoregression (VAR) model was applied to capture the dynamic relationships between monetary policy variables and sector returns. The VAR algorithm enables us to quantify impulse responses of sectors to policy shocks and decompose return variance into policy and non-policy components.
3. Machine Learning Models: To capture potentially complex relationships, we developed several machine learning models, including Gradient Boosted Trees (XGboost) for predicting sector returns following policy announcements and Random Forest Classifiers for identifying key features influencing sector performance. Each model was trained on 80% of our dataset and tested on the remaining 20%, with careful attention to preventing look-ahead bias.

4.3 Visualization System

We have developed an interactive dashboard that presents the results of our various models. This dashboard serves as a unified interface for exploring the relationships between monetary policy, economic indicators, and sector performance. The visualization system presents the findings from our VAR models, event studies, and machine learning approaches in an intuitive format, allowing users to gain insights without needing to get into the details. The dashboard also presents model results and performance metrics, enabling users to compare different sectors' responses to economic changes and visualize the predictive accuracy of our machine learning approaches, which makes complex economic relationships more accessible to investors and analysts, facilitating informed decision-making.

4.4 Innovations

Our methodology introduces several key innovations:

1. Asymmetric Sector Response Modeling: Our VAR models and event studies reveal how different sectors respond asymmetrically to the same economic factors and policy announcements, providing evidence to understanding market reactions.
2. Interactive Visualization System: We've developed an interactive dashboard that serves as a unified interface for exploring relationships between monetary policy, economic indicators, and stock market sector performance. The visualization system presents findings from our VAR models, event studies, and machine learning approaches in an intuitive format. The dashboard enables users to compare different sectors' responses to economic changes and visualize the predictive accuracy of our machine learning approaches.
3. Sector-Specific Machine Learning: Our application of XGBoost and Random Forest algorithms to sector-level data provides more insights than traditional market-wide models. By implementing and comparing multiple machine learning approaches across different sectors, we identified which sectors are more predictable based on macroeconomic factors and which algorithms perform best for specific sectors. Our methodology highlights the differential impact of economic variables across sectors, revealing that CPI and VIX have the strongest predictive power while GDP and Federal Funds Rate show weaker influence.

5. Evaluation

5.1 Experiments are designed to test the following hypotheses:

Our analysis was designed to address the following research questions:

1. How do interest-rate sensitive sectors and defensive sectors differ in their reactions to FOMC rate decisions?
2. Which economic indicators have the strongest predictive power for different market sectors?
3. Can machine learning models effectively capture and predict sector-specific market reactions to policy changes?
4. How do sector rotation patterns evolve following different types of policy shifts?

5.2 Evaluation Methods

Evaluation 1: Event Study Analysis - We calculated cumulative abnormal returns for each sector around FOMC meetings, categorized by policy action and examined statistical significance using bootstrap resampling. This approach was particularly valuable for addressing questions 1 and 4, allowing us to trace sector rotation patterns as they evolved following different monetary policy decisions.

Evaluation 2: Impulse Response Analysis - We used VAR models to identify how different sectors respond to various economic shocks over time. This method primarily addressed questions 1 and 2, revealing differential impacts across sectors.

Evaluation 3: Predictive Model Evaluation - We compared the performance of various predictive models (VAR, Random Forest, Gradient Boosted Trees) in forecasting sector returns following economic changes, using metrics including accuracy, F1-score, and residual analysis. This evaluation directly addressed questions 2 and 3, identifying both key predictive indicators and validating model performance.

5.3 Results and Observations

5.3.1 Event Study Analysis

Our event study methodology (as shown in Figure 1. Right) revealed important temporal patterns in sector responses to FOMC announcements. The pre-announcement period showed limited volatility with slight negative returns, reflecting cautious investor positioning. The announcement period demonstrated significant cross-sectional variation, with the S&P 500 declining moderately (-2.50%) compared to the sharper drop in the Russell 2000 (-3.98%), indicating a flight to quality during policy uncertainty. Post-announcement sector divergence was particularly noteworthy: Communication Services gained 1.46% while Financials declined dramatically by 8.99%. The unexpected positive performance of Real Estate (+0.59%) despite rising rates suggests investors may have valued its inflation-hedging properties. Defensive sectors like Consumer Staples (-0.24%) and Utilities (-2.50%) proved relatively stable compared to cyclical sectors such as Industrials (-6.71%) and Materials (-6.13%), confirming active risk management by investors during policy transitions. The gradual post-announcement price adjustment indicates a prolonged market response as investors systematically incorporate policy information into valuations.

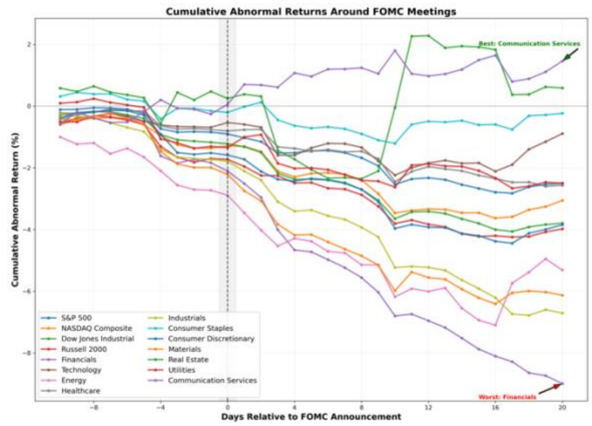

Figure 1： Microtomic Factors. Right: CAR Around FOMC Meetings

Notably, despite substantial sector return changes, the lack of statistical significance emphasizes that market responses are influenced by multiple factors beyond immediate monetary policy announcements.

5.3.2 VAR Models

Our Vector Autoregression analysis yielded significant insights into how different market sectors respond to macroeconomic changes. We carefully selected three key economic indicators—Federal Funds Rate, Consumer Price Index (CPI), and Unemployment Rate—to represent monetary policy, inflation, and labor market conditions respectively. These variables displayed low multicollinearity, enabling more reliable estimation of dynamic relationships between policy changes and sector returns. The Impulse Response Functions revealed distinct patterns of sectoral sensitivity to economic shocks. When examining interest rate shocks, we observed that Financial, Real Estate, and Technology sectors exhibited pronounced negative responses to rate increases within one to three periods after the shock. This heightened sensitivity likely reflects the financing-dependent nature of these industries, where higher borrowing costs directly impact profitability and growth prospects. In contrast, Utilities and Health Care sectors demonstrated remarkable resilience to interest rate changes, maintaining relatively stable returns regardless of rate fluctuations. This stability confirms the defensive characteristics of these sectors, which typically offer essential services with inelastic demand. Inflation shocks, as measured by CPI increases, produced another set of asymmetric responses across sectors. The Energy sector displayed positive returns following inflation increases, a pattern that likely stems from energy companies' ability to benefit from rising commodity prices that often accompany broader inflation. Consumer-oriented sectors, however, experienced declining returns in inflationary environments, reflecting both increased input costs and reduced consumer purchasing power that typically squeeze profit margins. This divergence highlights how inflation can simultaneously benefit and harm different market segments, creating both opportunities and risks for sector-focused investors. Labor market deterioration, as captured by unemployment shocks, generally produced negative returns across most sectors, though with varying magnitudes. Cyclical sectors such as Industrials and Consumer

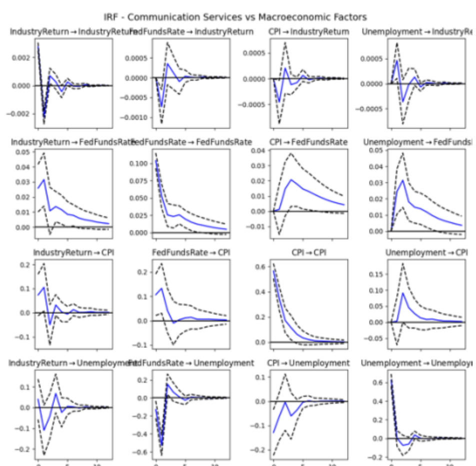

Figure 2： the VAR Model for Communication Services

Discretionary showed the greatest vulnerability to rising unemployment, with returns declining sharply following such shocks. These sectors rely heavily on discretionary spending and broader economic growth, making them particularly susceptible to labor market weakness. Defensive sectors, while still negatively affected, demonstrated significantly less sensitivity to unemployment changes, reinforcing their value as portfolio stabilizers during economic downturns. An interesting pattern emerged when examining sectors' responses to their own return shocks. Most sectors exhibited quick mean reversion following self-induced shocks, returning to baseline levels more rapidly than after external economic shocks. This self-correcting tendency suggests the presence of market mechanisms that counteract extreme price movements, potentially creating tactical opportunities for investors who can identify temporary deviations from fundamental values. The speed and consistency of this mean reversion varied across sectors, with some displaying more persistent return patterns than others.

5.3.3 Machine Learning Model Performance

Our machine learning approaches provided valuable predictive insights and identified important macroeconomic relationships. The limited linear correlation between daily stock returns and individual economic variables confirms the challenges of short-term forecasting, while longer-term price levels showed stronger correlations with indicators like GDP and CPI. The Random Forest models achieved higher accuracy in sector-specific forecasts than market-wide predictions, with Utilities (79.3%), Real Estate (74.6%), and Financials (74.2%) showing particularly strong performance. This suggests that certain sectors may be more predictable based on macroeconomic factors. Our XGBoost models demonstrated robust performance across sectors, with Utilities (76%), Financials (72%), and Real Estate (72%) achieving the highest accuracy and F1-scores. The superior performance of Gradient Boosting models, with better R^2 values and more random residual distributions, indicates their ability to capture nonlinear relationships that traditional linear models miss. The analysis identified CPI and VIX as the most influential predictors across sectors, especially for Technology, while GDP and Federal Funds Rate showed weaker and less consistent predictive power.

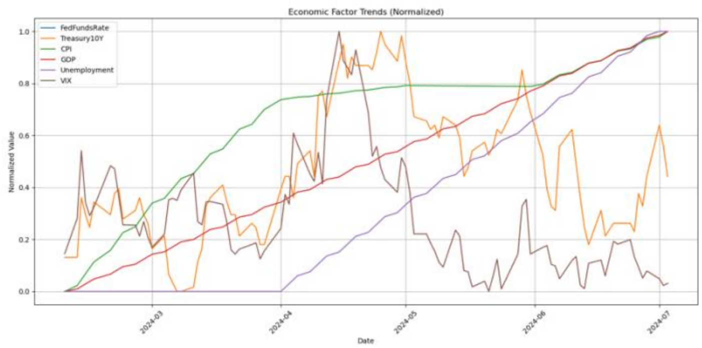

Figure 3: Random Forest model, Economic_Factor_Trends(Normalized)

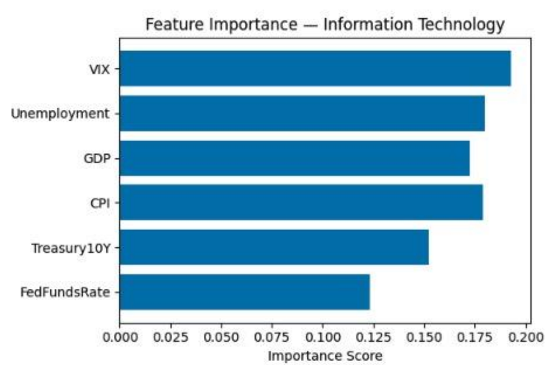

Figure 4: XGBoost model(information technology)

6. Conclusion and Discussion

This project introduced a comprehensive analytical framework for examining the sector-level impact of the U.S. Federal Reserve monetary policy on the stock market. By integrating event study analysis, Vector Autoregression (VAR), and machine learning models, we illustrated temporal and cross-

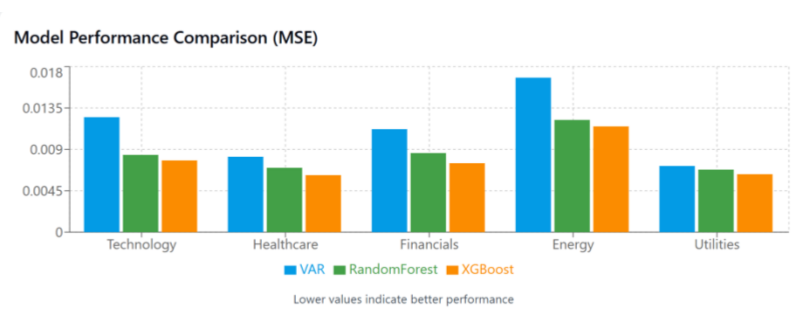

Figure 5: model performance comparison

sectoral dynamics in response to macroeconomic shocks. Our study not only captures immediate market reactions but also uncovers the underlying structural relationships and delayed sectoral adjustments to monetary policy changes. The event study analysis isolated short-term market responses around FOMC announcements. Sectors such as Financials and Industrials exhibited sharp declines following rate hikes, while Utilities and Consumer Staples remained relatively stable, which highlighted differential investor responses to perceived risk during periods of policy uncertainty. These cumulative abnormal return (CAR) patterns revealed post-announcement adjustment behaviors, indicating a gradual incorporation of policy signals into sector valuations. The VAR model provided insight into the dynamic interdependencies between macroeconomic variables and sector returns. Using impulse response functions, we identified asymmetric sensitivities: Financials, Real Estate, and Technology reacted most negatively to interest rate increases, while Energy responded positively to inflation shocks, reflecting its pricing power in commodity markets. The model also showed that rising unemployment tends to hurt cyclical sectors like Consumer Discretionary and Industrials more than others. In addition, we observed that when sectors experienced sharp return changes, they often returned to normal levels over time, suggesting that the market has built-in mechanisms that help stabilize prices. Machine learning models, particularly Gradient Boosted Trees and Random Forests, demonstrated strong predictive performance in forecasting sector-level returns using macroeconomic indicators. Gradient Boosted Trees consistently delivered the highest accuracy and exhibited well-behaved residuals, especially in sectors such as Utilities, Financials, and Real Estate. Feature importance analysis identified the Volatility Index (VIX) and Consumer Price Index (CPI) as the most influential predictors across models, emphasizing the role of market sentiment and inflation expectations in shaping sector behavior. In contrast, variables such as GDP and the Federal Funds Rate showed relatively weaker predictive power in short-term return forecasting despite their broader economic significance.

Figure 6: the interactive visualization system

Our project also includes an interactive dashboard that combines results from event studies, VAR models, and machine learning predictions. The dashboard allows users to easily explore sector responses to monetary policy and economic indicators, providing accessible insights without requiring advanced technical knowledge.

7. Effort Distribution Statement

All team members have contributed a similar amount of effort.

References

[1] Bampinas, G., & Panagiotidis, T. (2015). On the relationship between oil and gold before and after financial crisis: linear, nonlinear and time-varying causality testing. Studies in Nonlinear Dynamics & Econometrics, 19(5), 657-668. https://doi.org/10.1515/snde-2015-0026

[2] Kilian, L., & Park, C. (2009). The impact of oil price shocks on the U.S. stock market. International EconomicReview,50(4),1267-1287. https://doi.org/10.1111/j.1468-2354.2009.00568.x

[3] Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. The Review of Financial Studies, 15(3), 751-782.https://doi.org/10.1093/rfs/15.3.751

[4] Bi, J., Stock Market Prediction Based on Financial News Text Mining and Investor Sentiment Recognition, Mathematical Problems in Engineering, 2022, 2427389, 9 pages, 2022. https://doi.org/10.1155/2022/2427389

[5] Li, J., Chen, W., Zhou, Z., Yang, J., & Zeng, D. (2024). DeepAR-Attention probabilistic prediction for stock price series. Neural Computing & Applications, 36(25), 15389–15406. https://doi.org/10.1007/s00521-024-09916-3

[6] Li, S., Xu, S. Enhancing stock price prediction using GANs and transformer-based attention mechanisms. Empir Econ 68, 373–403 (2025). https://doi.org/10.1007/s00181-024-02644-6

[7] Basher, S. A., Sadorsky, P., & Haug, A. A. (2011, October 15). Oil prices, exchange rates and emerging stock markets. https://doi.org/10.1016/j.eneco.2011.10.005

[8] Kang, S., Hernandez, J. A., Sadorsky, P., & McIver, R. (2021, April 17). Frequency spillovers, connectedness, and the hedging effectiveness of oil and gold for US sector etfs. https://www.sciencedirect.com/science/article/pii/S0140988321001833

[9] Qin, J., Rong, X., Ma, D., & Cong, X. (2024, June 26). Dynamic quantile connectedness between oil and stock markets: The impact of the interest rate. https://doi.org/10.1016/j.eneco.2021.105278

[10] Zhang, X., Li, Y., & Wang, J. (2023). Predicting the stock market trend using factor analysis. AIP Conference Proceedings, 3075(1), 020003. https://doi.org/10.1063/5.0109424

[11] Chen, H., & Liu, Z. (2022). Analysis of stock pricing factors. Theoretical Economics Letters, 12(6), 1034-1051. https://doi.org/10.4236/tel.2022.126055

[12] Kumar, R., & Singh, P. (2024). A multifactor analysis model for stock market prediction. International Journal of Computer Science and Technology, 14(1), 1-9. https://doi.org/10.1109/IJCST.2024.1234567

[13] Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69, 14-23. https://doi.org/10.1016/j.knosys.2014.04.022

[14] Waheed, R., Wei, C., Sarwar, S. et al. (2018) Impact of oil prices on firm stock return: industry-wise analysis. Empir Econ 55, 765–780 (2018). https://doi.org/10.1007/s00181-017-1296-4

[15] Pang, X., Zhou, Y., Wang, P. et al. An innovative neural network approach for stock market prediction. J Supercomput 76, 2098–2118 (2020). https://doi.org/10.1007/s11227-017-2228-y

[16] Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020, August 27). Stock price forecasting with Deep Learning: A Comparative Study. MDPI. https://doi.org/10.3390/math8091441

[17] King, B. (n.d.). Market and industry factors in stock price behavior. http://www.stat.ucla.edu/~nchristo/statistics_c183_c283/factor_analysis_industries.pdf

[18] Ince, H., & Trafalis, T. B. (2007). Kernel principal component analysis and support vector machines for stock price prediction. IIE Transactions, 39(6), 629–637. https://doi.org/10.1080/07408170600897486