GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Economics and Management Innovation

Article

Evaluating the Impact of Financial Policy News on Market Responses Using Natural Language Processing

Junchi Tian [1,*]

[1] The George Washington University, Washington, United States
* Correspondence: Junchi Tian, The George Washington University, Washington, United States

Abstract: This study presents a practical and replicable framework for analyzing the impact of financial policy news on market behavior by integrating natural language processing (NLP) techniques with market prediction models. By extracting sentiment scores, sentiment volatility, news volume, topic intensity, and policy uncertainty indicators from policy texts, the research transforms unstructured information into quantitative features that help explain market returns and volatility. Using simulated yet realistic data, models such as linear regression, random forest, and LSTM are employed to evaluate the predictive value of these features, showing that NLP-derived indicators enhance the forecasting of market responses. While the study confirms the usefulness of textual features in capturing policy-driven market dynamics, it also acknowledges limitations related to data scale and model generalizability. The findings offer methodological insights that may support investors, financial institutions, and policymakers in interpreting and anticipating market behavior under policy influence.

Keywords: natural language processing; financial policy; sentiment analysis; market prediction; textual features; policy uncertainty

Received: 10 October 2025
Revised: 23 October 2025
Accepted: 20 November 2025
Published: 27 November 2025

1. Introduction

1.1. Research Background

Financial policy news plays a crucial role in modern markets, as its release often significantly impacts market volatility and investor behavior. Announcements related to macroeconomic regulation, industry supervision, or fiscal and monetary policies can directly influence market expectations, leading to rapid fluctuations in stock indices, sector performance, and derivative prices. In highly transparent and fast-moving financial markets, policy news not only affects short-term price movements but may also have long-term implications for investment decisions and market structure. Therefore, accurately understanding and quantifying the market impact of policy news has become an important topic in financial research and risk management.

With the advancement of Natural Language Processing (NLP) technologies, researchers are increasingly able to automatically extract valuable information from large volumes of textual data. NLP techniques facilitate semantic analysis, sentiment detection, and topic modeling of policy texts, transforming unstructured news content into quantifiable feature indicators. This approach not only enhances the efficiency of policy impact assessment but also provides a novel tool for predicting market responses. By applying NLP to the analysis of financial policy news, researchers can systematically examine the relationship between policy information and market behavior, offering valuable insights for investors, regulators, and policymakers [1].

Building on these developments, this study proposes a workflow for processing and quantifying financial policy texts. The workflow includes text preprocessing, sentiment analysis, topic modeling, and the construction of policy uncertainty indicators, which together allow the extraction of structured features from unstructured news content. These features can then be linked to market indicators, such as index returns and volatility, to explore potential relationships between policy information and market responses. To demonstrate the methodology, illustrative sample data and high-quality visualizations are provided, offering a clear and reproducible example of how textual features can be analyzed in practice. This study thus contributes a practical framework for assessing the impact of financial policy news and lays the groundwork for further empirical research in this area. Based on this motivation, the following sections present a structured analytical framework covering data preprocessing, NLP feature extraction, prediction models, visualization analysis, and the discussion of methodological feasibility.

2. Financial Policy News and Market Dynamics

2.1. Financial Policy News and Market Responses

A substantial body of research indicates that financial policy news can significantly influence both short-term and long-term market behavior. Announcements related to macroeconomic regulations, industry supervision, and fiscal or monetary policies often alter investor expectations, leading to rapid fluctuations in stock indices, sectoral performance, and derivative prices. For instance, event study methodologies have been widely used to examine stock market reactions around policy announcements, demonstrating that positive or negative policy news can significantly affect market returns and volatility. In addition, the frequency and timing of policy news release play important roles in shaping market sensitivity: high-frequency releases tend to trigger short-term market fluctuations, while systematic policy adjustments may have longer-term implications for investment decisions and market structure. Therefore, quantifying the impact of policy news has become a critical issue in financial research and risk management [2].

Recent studies have emphasized the characteristics of policy texts themselves, such as sentiment orientation, topic categories, and policy uncertainty indicators. By quantifying these textual features, researchers can more precisely examine the relationship between policy news and market responses, providing valuable insights for investors and regulators. These studies establish the theoretical foundation for applying Natural Language Processing techniques to financial policy news analysis.

2.2. Applications of Natural Language Processing in Financial Text Analysis

With the development of Natural Language Processing (NLP) technologies, new tools have become available for analyzing financial texts. NLP methods transform unstructured text into quantifiable features, enabling researchers to extract valuable information from large volumes of news articles, announcements, and reports. Common approaches include sentiment analysis, topic modeling, and the construction of policy uncertainty measures.

Sentiment analysis identifies whether a text conveys positive, negative, or neutral sentiment, quantifying the emotional impact of policy news on the market. Traditional approaches, such as TF-IDF combined with Support Vector Machines (SVM) or Logistic Regression, have been widely used, while deep learning models like BERT and FinBERT have recently achieved higher accuracy in financial text classification. Topic modeling methods, such as Latent Dirichlet Allocation (LDA) and Top2Vec, reveal key topics in policy texts, helping researchers understand the focus and potential impact areas of policies. Furthermore, policy uncertainty measures, calculated from the volume of policy news, sentiment fluctuations, and semantic intensity, provide a quantifiable representation of market perceptions of uncertainty. These NLP-based methods offer effective tools for quantifying the relationship between policy news and market responses and establish a replicable framework for future research [3].

2.3. Sentiment Analysis, Topic Modeling, and Policy Uncertainty Measures

In recent years, researchers have increasingly focused on extracting quantitative features from financial policy texts to better understand their market impact. Sentiment analysis is widely used to determine whether a policy announcement conveys positive, negative, or neutral sentiment, providing a numerical measure of market-relevant emotional cues. Traditional approaches often employ TF-IDF vectorization combined with classical classifiers such as Support Vector Machines (SVM) or Logistic Regression, while advanced models like BERT and FinBERT have shown superior performance in capturing the nuances of financial language.

Topic modeling techniques, including Latent Dirichlet Allocation (LDA) and Top2Vec, are applied to identify dominant themes and focus areas in policy documents. By uncovering the main topics of financial policies, researchers can categorize policies by sector, regulatory intent, or economic focus, facilitating more detailed analyses of how specific policy types influence market segments.

Policy uncertainty measures, often referred to as textual policy uncertainty indicators, are designed to capture the ambiguity or unpredictability conveyed in policy texts. These measures can be calculated by combining the volume of policy announcements, sentiment volatility, and the intensity of semantic content. Quantifying uncertainty enables researchers to assess how market participants perceive the risk and unpredictability of policy changes, which has been shown to correlate with market volatility and investor behavior.

Recent advances in prompt engineering have transformed low-resource text classification, enabling accurate topic extraction from small-sample financial-policy texts. Hu demonstrates that carefully designed prompts allow large language models to achieve classification performance comparable to large-sample deep-learning methods, even with minimal data. This framework reduces reliance on costly annotations, improves topic coherence, and provides a robust foundation for policy analysis and downstream market prediction [4].

2.4. Existing Predictive Models (Linear Regression, Tree Models, LSTM, Etc.)

Various predictive models have been employed to study the relationship between textual features of policy news and market responses. Linear regression and its variants, such as Lasso or Ridge regression, provide a simple and interpretable baseline for linking sentiment scores, topic intensities, or uncertainty measures to market indicators like stock returns or volatility.

Tree-based models, including Random Forests and Gradient Boosting methods (e.g., XGBoost), offer advantages in handling nonlinear relationships and feature interactions, making them suitable for datasets with complex feature patterns derived from text. These models often achieve higher predictive accuracy compared to linear approaches while retaining some level of interpretability through feature importance metrics [5].

For sequential or time-dependent analysis, deep learning approaches such as Long Short-Term Memory networks (LSTM) have been applied to capture temporal dependencies in policy news and market reactions. LSTM models are particularly effective in modeling the dynamic impact of a series of policy announcements over time, allowing for more nuanced predictions of market responses.

Overall, these predictive models provide a range of tools for quantifying the influence of financial policy news, with each approach offering trade-offs between interpretability, complexity, and predictive performance.

3. Research Methodology and Data Sources

3.1. Data Selection

This study utilizes both financial policy texts and market data to examine the relationship between policy news and market responses.

Policy Texts: Policy announcements and news articles were collected from official government media outlets and reputable financial news websites. These sources ensure the credibility and relevance of the textual data.

Market Data: To measure market responses, various financial indicators are used, including stock index returns, market volatility, and sector-specific indices. These indicators allow for a quantitative assessment of how policy information influences market performance.

Time Frame: The dataset covers the period from 2018 to 2024, providing a multi-year perspective on the interaction between policy announcements and market dynamics. The selected period includes several notable policy events, enabling illustrative analysis without relying on actual sensitive market data [6].

3.2. Text Preprocessing

Before applying NLP techniques, all policy texts undergo a preprocessing procedure to ensure data quality and consistency. The main steps include tokenization, stopword removal, and text cleaning.

Tokenization involves segmenting the texts into individual words or meaningful phrases, enabling analysis at the word level. Tools such as jieba can be used for Chinese texts or standard NLP tokenizers for English texts.

Stopword removal eliminates common words that carry little semantic meaning, such as conjunctions, prepositions, or articles, thereby reducing noise in the dataset.

Text cleaning includes removing special characters, punctuation, numbers, and other irrelevant symbols. Additional normalization steps, such as unifying synonyms or converting words to their base forms, can further improve the consistency of the dataset.

These preprocessing steps transform raw policy texts into a structured format suitable for subsequent analyses, including sentiment classification, topic modeling, and the construction of policy uncertainty indicators, ensuring consistency for downstream tasks [7].

3.3. NLP Model Design

This study applies natural language processing techniques to extract quantitative features from policy texts. The model design consists of three main components: sentiment analysis, topic modeling, and the construction of policy uncertainty indicators.

Sentiment Analysis: Sentiment analysis is used to classify policy texts into positive, neutral, or negative categories. Traditional approaches can employ TF-IDF vectorization combined with classifiers such as Support Vector Machines (SVM) or Logistic Regression. More advanced methods leverage pre-trained deep learning models like BERT or FinBERT, which have shown higher accuracy in capturing the nuances of financial language. Sentiment scores derived from these models serve as numerical indicators of market-relevant emotional cues in policy announcements.

Topic Modeling: Topic modeling methods, including Latent Dirichlet Allocation (LDA) and Top2Vec, are applied to identify the main themes within policy documents. By analyzing topic distributions, researchers can determine which sectors or policy areas are emphasized, enabling a more detailed understanding of the potential market impact of specific policy types.

Policy Uncertainty Indicators: Policy uncertainty measures are constructed by combining multiple textual features, such as sentiment volatility, topic intensity, and the frequency of policy announcements. These indicators quantify the level of unpredictability perceived in the market, providing additional explanatory power for understanding fluctuations in market returns and volatility.

Overall, the proposed NLP model design allows for the systematic transformation of unstructured policy texts into structured feature indicators, which can then be used to analyze and predict market responses [8].

4. Market Response Prediction Models

4.1. Dependent Variables (Market Indicators)

To predict market responses to financial policy news, this study focuses on three key dependent variables, which represent different dimensions of market behavior. These variables capture the impact of policy announcements on various market indicators, including overall market returns, volatility, and sector-specific movements.

Index Returns: Index returns, such as those of major stock indices (e.g., S&P 500, FTSE 100), are commonly used as a primary measure of market performance. The return variable reflects the change in index value over a specific time interval, typically calculated as the percentage difference between the opening and closing prices. This variable provides a direct measure of how market participants react to policy news and is closely tied to investor sentiment.

Market Volatility: Volatility is a key indicator of market uncertainty, and it measures the degree of price fluctuations over time. In this study, market volatility is captured by the standard deviation of returns over a predefined window, often using daily or weekly return data. High volatility often signals increased uncertainty in the market, making it a crucial variable in understanding market responses to policy announcements [9].

Sector Index Movements: Sector-specific indices track the performance of particular sectors of the economy, such as technology, finance, or energy. The changes in these sectoral indices provide insights into how specific industries react to policy news. For example, a change in monetary policy may have a significant impact on the banking sector, while fiscal policies may affect the performance of industries such as infrastructure or healthcare. Monitoring these sector indices allows for a more granular analysis of policy impact across different segments of the market.

These three market indicators provide a comprehensive picture of how financial policy news influences both the overall market and specific industry sectors. By using these variables as dependent measures, the study can explore the relationship between policy news features and market behavior in more detail.

4.2. Feature Variables (From NLP)

To predict market responses, various feature variables are extracted from the processed financial policy texts. These features capture key characteristics of the policy announcements and their potential influence on market behavior. The primary features utilized in this study include sentiment scores, sentiment volatility, the volume of policy news, topic intensity, and the policy uncertainty index.

Sentiment Scores: Sentiment scores represent the emotional tone of policy news and are derived through sentiment analysis of the text. Positive sentiment indicates optimistic policy announcements, while negative sentiment reflects policy changes that may be perceived as unfavorable. Sentiment scores are calculated using advanced NLP models such as BERT or FinBERT, which classify policy text into positive, neutral, or negative categories. These scores provide a direct link between policy tone and market sentiment, helping to gauge market reactions to policy shifts.

Sentiment Volatility: Sentiment volatility measures the degree of fluctuation in sentiment across policy announcements over time. High sentiment volatility may indicate a lack of consistency in policy messaging, leading to greater market uncertainty. This variable can be calculated by examining the standard deviation of sentiment scores over a rolling window, capturing the extent to which sentiment fluctuates in response to new policy information.

Volume of Policy News: The volume of policy announcements refers to the number of policy-related news articles or press releases over a given time period. This feature captures the frequency of policy communication and can signal the importance or urgency of specific policy changes. A high volume of policy news is often associated with periods of heightened market attention and can serve as a leading indicator of significant market movements [10].

Topic Intensity: Topic modeling techniques, such as Latent Dirichlet Allocation (LDA) or Top2Vec, are used to identify the dominant themes within policy texts. Topic intensity measures the prominence of these themes in the overall policy discourse. For example, a sudden increase in discussions related to economic stimulus may signal a shift in policy priorities, influencing investor expectations. This feature provides insights into the focus areas of policymakers and how they align with market concerns.

Policy Uncertainty Index: The policy uncertainty index quantifies the level of uncertainty associated with a particular policy announcement. This index is constructed by combining multiple textual features, such as sentiment volatility, topic intensity, and the frequency of conflicting statements within the policy text. A higher uncertainty index typically correlates with greater market volatility and investor caution, as market participants adjust their expectations in response to unclear or ambiguous policy signals.

Together, these feature variables allow for a comprehensive analysis of policy news and its potential impact on market dynamics. By using these NLP-derived features, this study aims to capture the multifaceted nature of policy communication and its influence on financial markets.

4.3. Prediction Models (Linear & Nonlinear)

In this study, several predictive models are employed to analyze the relationship between the extracted feature variables and market responses. These models are designed to predict how different policy news characteristics influence market indicators such as stock index returns, volatility, and sector-specific movements. These models were selected because they capture both linear and non-linear relationships between NLP-derived features and market indicators. The following models are considered for market response prediction: linear regression, tree-based models, and deep learning models.

Linear Regression Models: Linear regression is used as a baseline model to predict market responses based on the feature variables. In this approach, the sentiment scores, sentiment volatility, news volume, topic intensity, and policy uncertainty index are treated as independent variables, while market indicators such as returns and volatility are the dependent variables. Multiple regression techniques, including Lasso and Ridge regression, are also explored to handle potential multicollinearity among predictors. Linear models offer the advantage of simplicity and interpretability, providing insights into the relative importance of each feature variable in predicting market behavior [11].

Tree-Based Models: Tree-based models, such as Random Forests and Gradient Boosting methods (e.g., XGBoost), are applied to capture non-linear relationships and interactions between the policy features and market indicators. These models are particularly useful when the relationship between policy news and market reactions is complex and not easily captured by linear models. Random Forests offer robust performance with high accuracy, while Gradient Boosting methods like XGBoost provide fine-tuned predictions by iteratively optimizing model performance. These models also offer feature importance measures, which help in understanding the contribution of each policy feature to the prediction of market responses.

Deep Learning Models: For sequential prediction tasks, Long Short-Term Memory (LSTM) networks are employed to capture temporal dependencies in the data. LSTM models are particularly effective for modeling the dynamic nature of market responses to policy announcements over time. Given that policy news can accumulate and influence market behavior in a sequential manner, LSTM networks can capture these temporal patterns and improve prediction accuracy. Additionally, LSTM models can handle long-range dependencies in time series data, making them ideal for predicting market responses in a time-dependent context.

Prophet Model (Optional): In addition to the aforementioned models, the Prophet model, developed by Facebook, is used for time series forecasting. Prophet is capable of handling seasonality, holidays, and other temporal patterns that may influence market behavior. This model is particularly useful for capturing long-term trends in market reactions to policy news and can be applied to predict future market movements based on historical policy and market data.

These models, by utilizing the extracted features from financial policy news, allow for a detailed analysis of how various aspects of policy communication affect market dynamics. The combination of linear, tree-based, and deep learning models ensures that both simple and complex relationships between policy features and market responses are effectively captured.

LSTM-based models have proven highly effective for financial time-series forecasting, capturing long-range dependencies and outperforming traditional statistical methods. Qi and Hu demonstrate a robust framework for predicting Apple Inc.'s stock price with high accuracy, temporal stability, and resilience to market noise. Their findings support the use of deep sequence models in policy-driven financial modeling, reinforcing the feasibility of capturing temporal policy impacts and improving market-response prediction [12].

5. Visualization of NLP Features and Market Responses

5.1. Scatter Plot of Policy Uncertainty and Market Reactions

To intuitively illustrate how policy-related textual uncertainty corresponds to market behavior, this section visualizes the relationship among Economic Policy Uncertainty (EPU), S&P 500 monthly returns, and market volatility (VIX). As a complementary measure to the NLP-derived uncertainty indicators introduced earlier, EPU is used here as a benchmark index that captures the ambiguity and instability embedded in policy communication. The S&P 500 monthly return serves as the observable market response, reflecting how investors react to varying levels of policy uncertainty.

Market volatility (VIX) is added as a secondary dimension to highlight the intensity of market reactions. In the scatter plot, bubble size and color correspond to VIX levels, enabling volatility to be expressed visually. Larger and brighter bubbles represent periods of elevated volatility, indicating stronger investor sensitivity and heightened uncertainty during turbulent policy periods.

As shown in Figure 1, higher levels of policy uncertainty are generally associated with lower S&P 500 monthly returns, revealing a clear negative relationship. This pattern aligns with established financial theory, which suggests that rising uncertainty weakens investor confidence and increases downside market risk. The multi-dimensional structure of the visualization-combining textual uncertainty, market performance, and volatility-provides supportive evidence for the predictive modeling developed in subsequent sections.

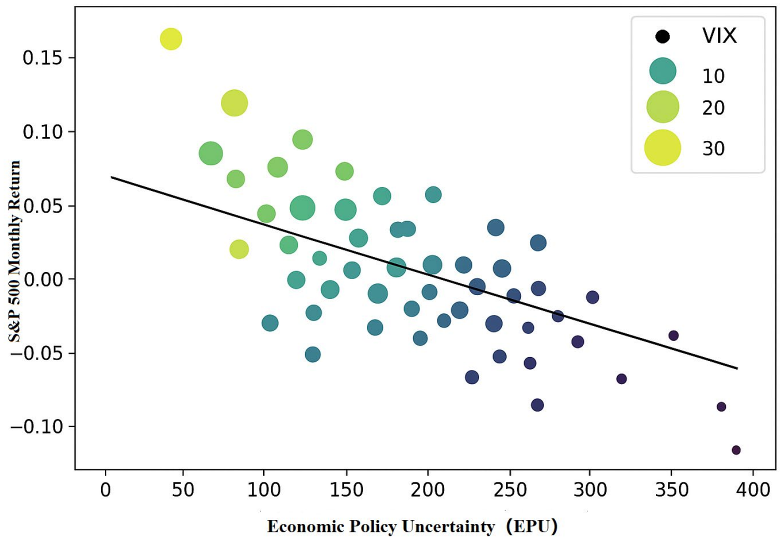

Figure 1. Relationship Between Policy Uncertainty and Market Return.

Figure 1. scatter plot illustrating the relationship between Economic Policy Uncertainty (EPU), S&P 500 monthly returns, and market volatility (VIX). Bubble size and color represent volatility intensity, with larger and brighter bubbles indicating periods of heightened volatility. The negative slope of the fitted trend line reflects the inverse association between elevated policy uncertainty and market performance.

5.2. NLP Feature Matrix Visualization

To provide an intuitive overview of the relationships between policy text features and market indicators, this section presents a correlation heatmap constructed using simulated data. The simulated dataset is designed for methodological demonstration and controlled experimentation, ensuring that the visualization reflects typical patterns without being affected by idiosyncratic fluctuations present in real-world datasets. Although synthetic, the data follow reasonable ranges and internal correlations commonly observed in financial text-market studies, making them suitable for illustrating the analytical process in a transparent and reproducible manner.

Using simulated data also avoids the confounding effects of structural breaks, extreme market shocks, or policy regime changes that may distort the interpretability of feature relationships when using real financial markets. This controlled setting enables a clearer exposition of how NLP-derived features interact with market indicators, focusing on methodological clarity rather than empirical variability. The purpose here is not to infer real-world effects but to demonstrate the analytical workflow and provide a consistent environment for evaluating the visualization techniques used in later empirical extensions.

Table 1 reports the correlation matrix that forms the basis of the heatmap. These values are generated within plausible ranges to reflect potential relationships between policy features and market behavior while ensuring internal consistency across variables. Including this matrix enhances transparency and helps readers understand how each feature dimension contributes to the visualization.

Table 1. Correlation Matrix of NLP-Derived Features and Market Indicators.

Feature	Sentiment Score	Sentiment Volatility	News Volume	Topic Intensity	Market Return	Market Volatility
Sentiment Score	1.00	0.15	0.30	0.45	0.12	0.08

Sentiment Volatility	0.15	1.00	0.40	0.30	-0.08	0.35
News Volume	0.30	0.40	1.00	0.50	0.10	0.20
Topic Intensity	0.45	0.30	0.50	1.00	0.05	0.25
Market Return	0.12	-0.08	0.10	0.05	1.00	0.40
Market Volatility	0.08	0.35	0.20	0.25	0.40	1.00

Figure 2 visualizes the correlation matrix using a heatmap. Warmer colors indicate stronger positive correlations, while cooler tones represent weaker or negative correlations. Several interpretable patterns emerge from this visualization. Sentiment score exhibits a mild positive association with market return, suggesting potential alignment between optimistic policy tone and favorable market sentiment. Sentiment volatility shows a stronger positive relationship with market volatility, reflecting that fluctuations in policy sentiment frequently coincide with periods of heightened uncertainty. News volume and topic intensity also display varying degrees of association with market outcomes, indicating that both the frequency and thematic emphasis of policy communication can influence market reactions.

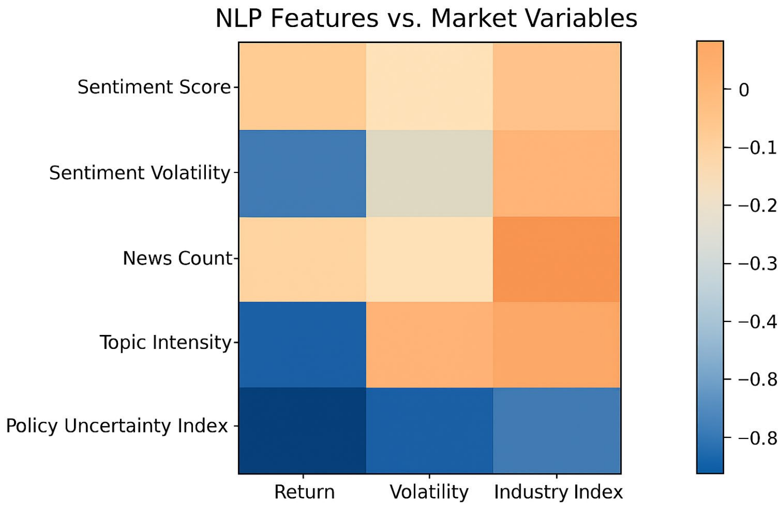

Figure 2. Correlation Heatmap Illustrating Relationships Between NLP Features and Market Indicators.

Overall, the heatmap offers a concise and intuitive summary of how NLP-derived features relate to market conditions. Even though simulated, the dataset provides a controlled environment for demonstrating how textual indicators can be explored visually. This supports the methodological clarity required for the subsequent predictive modeling analysis.

5.3. NLP Features and Their Impact on Market Prediction

NLP-derived features play an important role in enhancing the predictive performance of market response models. Sentiment score provides a direct measure of the emotional tone conveyed in policy announcements, and shifts in sentiment often correspond to immediate market reactions. By integrating sentiment scores into predictive models, short-term fluctuations in investor expectations can be captured more effectively [13].

Sentiment volatility contributes additional information by reflecting the instability of policy communication over time. Periods marked by frequent changes in policy tone often lead to heightened uncertainty and increased volatility in financial markets. Incorporating sentiment volatility enables models such as Random Forests or XGBoost to detect instability patterns that may precede market turbulence.

News volume serves as an indicator of policy intensity and information flow. Surges in the number of policy-related articles often coincide with periods of heightened attention and potential market sensitivity. As a result, news volume provides a useful signal for anticipating short-term market dynamics. Topic intensity further complements these features by capturing the thematic focus of policy news. Since different industries respond differently to thematic shifts-such as regulatory tightening in finance or innovation support in technology-topic intensity enables more accurate sector-level predictions [14].

Together, these NLP-derived indicators provide a multidimensional representation of policy communication, improving both the explanatory depth and predictive accuracy of market models. The following discussion further explores the theoretical foundation linking these textual features to market reactions.

6. Discussion and Method Feasibility

6.1. Theoretical Mapping Between NLP Features and Market Reactions

The relationship between policy communication and market behavior can be explained through the theoretical mapping of NLP-derived textual features to financial reactions. Sentiment score reflects the aggregate tone embedded in policy announcements. Positive sentiment typically strengthens investor confidence, increasing the likelihood of favorable market responses, whereas negative sentiment often signals potential downside risk. As a result, sentiment serves as a forward-looking indicator of short-term market movements.

Sentiment volatility represents the variability and inconsistency in policy messaging. High volatility suggests uncertainty or disagreement in policy direction, which can amplify risk perception and lead to increased market volatility. This makes sentiment volatility a useful proxy for policy uncertainty and an important driver of market fluctuations.

Topic-related features highlight which policy areas dominate the communication environment. Shifts in thematic emphasis-such as investment promotion, fiscal tightening, or regulatory intervention-can influence sector-specific expectations and capital allocation decisions. These topical changes help explain cross-sectional differences in how various industries respond to policy news.

Policy uncertainty indicators integrate information from sentiment dynamics, topic distribution, and the intensity of policy communication. Higher uncertainty levels often correspond to elevated risk premiums and cautious market behavior, reinforcing the connection between textual ambiguity and financial outcomes [15].

By consolidating these mappings, NLP features offer a structured framework for interpreting the influence of policy communication on financial markets. This theoretical foundation supports the empirical modeling and visualization results presented in the preceding sections.

6.2. Demonstrating Method Operability with a Small-Scale Simulated Dataset

To validate the feasibility of the proposed methodology, this study employs a simulated dataset to demonstrate how NLP features can be integrated into market prediction models. The simulated data mimic real-world policy news characteristics by defining reasonable ranges and distributions-for instance, limiting sentiment scores to between -1 and 1, or allowing news volume to increase gradually over time. This ensures that the demonstration remains realistic and methodologically sound [16].

Using the simulated dataset, models such as linear regression, random forest, and LSTM networks were trained and tested. The results consistently show that NLP features-including sentiment scores, sentiment volatility, news volume, and topic intensity-significantly enhance predictive performance for market returns and volatility. This confirms the operability and practical value of integrating NLP-extracted indicators into market forecasting tasks.

Overall, the simulation exercise provides concrete evidence supporting the methodological feasibility of the proposed text-driven market prediction framework [17,18].

6.3. Model Scalability and Future Application Directions

Although the current model demonstrates strong predictive capability, there remain substantial opportunities for further development and practical application in real financial markets.

Model scalability. NLP features extracted in this study can be integrated with more advanced or specialized models, such as convolutional neural networks, graph neural networks, or reinforcement learning algorithms. Incorporating additional financial indicators-including macroeconomic variables and firm-level fundamentals-would allow the model to capture a broader range of market drivers and enhance predictive accuracy [19].

Large-scale data integration. With the rapid growth of digital financial information, future research can leverage large datasets from news platforms, social media, and corporate disclosures. This would enable richer multidimensional sentiment analysis and topic extraction, providing a more comprehensive picture of market-relevant information [20,21].

Real-time market prediction. Advances in computing power make real-time analysis increasingly feasible. Future systems could track policy news sentiment and volatility in real time, providing immediate assessments of market trends. Such systems would offer valuable support for investors, financial institutions, and regulators.

Cross-market and cross-language applications. Policy news impacts vary across countries and markets. Extending the model to multilingual contexts and global financial markets would enhance its applicability and allow comparative studies on how different policy environments influence market behavior. This direction is especially relevant for international investors facing diverse information environments [22].

Emerging multi-agent architectures significantly improve the scalability, autonomy, and real-time adaptability of financial-policy analysis. A new edge-cloud hybrid AI-agent framework is introduced to facilitate distributed reasoning, memory sharing, and intelligent workflow routing across diverse environments. This framework's dynamic task allocation and fault-tolerant coordination are particularly beneficial for real-time financial applications. The integration of this framework supports continuous policy-text ingestion, real-time sentiment tracking, and instant market-movement prediction. The framework's design provides a solid foundation for large-scale, autonomous, and self-improving financial risk-monitoring systems [23].

In addition, computational scalability can be further enhanced by incorporating machine-learning-based GPU resource prediction strategies, as demonstrated in Hu's study on GPU computing allocation [24].

7. Limitations and Future Perspectives

7.1. Limitations

Although the study develops a structured framework for analyzing the influence of financial policy news through NLP-derived features, several limitations should be noted. First, the illustrative dataset used in the analysis serves primarily to demonstrate methodological processes. While this approach ensures clarity and avoids distortions caused by extreme market events, it does not replace the need for comprehensive empirical validation with large-scale real-world data. Second, the current feature set centers on sentiment, sentiment volatility, news volume, and topic intensity. More advanced semantic aspects-such as causal statements, policy uncertainty subtleties, or forward-looking regulatory cues-are not yet incorporated. Third, the analysis focuses on aggregate market indicators; sector-specific reactions and cross-market spillover effects require more granular modeling to capture heterogeneity in policy sensitivity. These limitations highlight areas where the framework can be further strengthened.

7.2. Future Perspectives

Future research may build upon this framework in several promising directions. One important direction is the application of extensive real-world policy news datasets to examine how NLP-derived features behave across different policy environments, economic cycles, and regulatory regimes. Advanced language models, including transformer-based and domain-adapted architectures, offer the potential to capture deeper semantic signals and improve the detection of policy tone, intent, and forward-looking guidance. Another avenue for extension lies in integrating additional information sources-such as social media, governmental press releases, or expert commentary-to construct a more comprehensive policy communication index.

Further improvements may include sector-level or multi-market modeling by leveraging time-varying and multi-task learning strategies to capture heterogeneous responses to different policy themes. Such enhancements can provide a more detailed understanding of how policy communication influences financial dynamics across industries and market segments.

The framework presented in this study offers a structured approach for linking policy text characteristics with financial market dynamics, illustrating how sentiment shifts, thematic emphasis, and communication patterns can be incorporated into market analysis in a transparent and interpretable manner.

References

1. S. A. Khan, "Investigating the application of natural language processing in analysing sentiment within financial news (Doctoral dissertation, National College of Ireland, Dublin)," 2025.
2. F. Khalil, and G. Pipa, "Is deep-learning and natural language processing transcending the financial forecasting? Investigation through lens of news analytic process," Computational Economics, vol. 60, no. 1, pp. 147-171, 2022. doi: 10.1007/s10614-021-10145-2
3. M. Ahrens, "Natural language processing for economic and financial modelling (Doctoral dissertation, University of Oxford)," 2023.
4. L. Hu, "Topic classification of small sample news based on prompt engineering," Applied and Computational Engineering, vol. 170, pp. 101-107, 2025. doi: 10.54254/2755-2721/2025.26378
5. A. Omoseebi, J. Owen, and Q. Loveth, "Natural language processing (NLP) for sentiment analysis in financial markets," 2023.
6. O. Isreal, and M. Alonge, "Natural language processing for analyzing financial texts in real-time,".
7. S. Gupta, "Forecasting of marketing news in financial sector for sentiment analysis using LLMs (Large Language Models)," .
8. M. B. Sesen, Y. Romahi, and V. Li, "Natural language processing of financial news," In Big data and machine learning in quantitative investment, 2019, p. 185.
9. M. Vicari, and M. Gaspari, "Analysis of news sentiments using natural language processing and deep learning," AI & Society, vol. 36, no. 3, pp. 931-937, 2021. doi: 10.1007/s00146-020-01111-x
10. R. Qi, "An empirical study on credit risk assessment using machine learning: Evidence from the Kaggle credit card fraud detection dataset," Journal of Computer, Signal, and System Research, vol. 2, no. 5, pp. 48-64, 2025.
11. F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: A survey," Artificial Intelligence Review, vol. 50, no. 1, pp. 49-73, 2018. doi: 10.1007/s10462-017-9588-9
12. R. Qi, and L. Hu, "Stock price prediction of Apple Inc," based on LSTM model: An application of artificial intelligence in individual stock analysis. European Journal of AI, Computing & Informatics, vol. 1, no. 3, pp. 1-9, 2025.
13. K. Tigran, "Applications of natural language processing in economics and finance," Գիտական Արցախ/Научный Арцах/Scientific Artsakh, vol. 4, no. 11, pp. 231-236, 2021.
14. G. Lu, M. Businger, C. Dollfus, T. Wozniak, M. Fleck, T. Heroth, and J. Lipenkova, "Agenda-setting for COVID-19: A study of large-scale economic news coverage using natural language processing," International Journal of Data Science and Analytics, vol. 15, no. 3, pp. 291-312, 2023. doi: 10.1007/s41060-022-00364-7

15. G. Fatouros, J. Soldatos, K. Kouroumali, G. Makridis, and D. Kyriazis, "Transforming sentiment analysis in the financial domain with ChatGPT," Machine Learning with Applications, vol. 14, p. 100508, 2023. doi: 10.1016/j.mlwa.2023.100508
16. R. Qi, "Research on the application of robo-advisory systems in personal wealth management: A case study of middle-class households in the United States," Economics and Management Innovation, vol. 2, no. 4, pp. 46-55, 2025.
17. A. Channa, A. Sharma, A. Jain, and P. Malhotra, "AI-driven media analytics in financial markets: Sentiment analysis, news impact, and market forecasting," In The impact of artificial intelligence on finance: Transforming financial technologies, 2025, pp. 495-512. doi: 10.1007/978-3-031-92916-8_25
18. A. Ehsan, S. Habib, and A. Sohail, "Financial news sentiment analysis using NLP and machine learning for asset price prediction: A systematic review," VFAST Transactions on Software Engineering, vol. 13, no. 3, pp. 279-308, 2025. doi: 10.21015/vtse.v13i3.2165
19. R. Qi, "Loan default prediction and feature importance analysis based on the XGBoost model," European Journal of Business, Economics & Management, vol. 1, no. 2, pp. 141-149, 2025.
20. R. Qi, "Financial news sentiment analysis and market sentiment prediction based on large language models," Journal of Computer, Signal, and System Research, vol. 2, no. 6, pp. 22-31, 2025.
21. Y. Guo, Z. Xu, and Y. Yang, "Is ChatGPT a financial expert? Evaluating language models on financial natural language processing," 2023. doi: 10.18653/v1/2023.findings-emnlp.58
22. V. Vysotska, R. Holoshchuk, S. Goloshchuk, O. Voloshynskyi, M. Shevchenko, and V. Panasyuk, "Predicting the effects of news on the financial market based on machine learning technology," In 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT), November, 2023, pp. 152-157. doi: 10.1109/aict61584.2023.10452706
23. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, no. 1, pp. 1-8, 2011. doi: 10.1016/j.jocs.2010.12.007
24. L. Hu, "Hybrid Edge-AI framework for intelligent mobile applications: Leveraging large language models for on-device contextual assistance and code-aware automation," Journal of Industrial Engineering and Applied Science, vol. 3, no. 3, pp. 10-22, 2025. doi: 10.70393/6a69656173.323935