International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Predictive Power of Social Media Sentiment on Stock Market Returns

Leandro Jorge Yacoubian

Independent Researcher

Abstract

This study investigates the predictive power of social media sentiment on stock market returns. As social media platforms have become increasingly influential in shaping public opinion, their impact on financial markets has grown significantly. Through a comprehensive analysis of existing literature and empirical evidence, we examine how sentiment expressed on platforms such as Twitter, Reddit, and StockTwits correlates with and potentially predicts stock market movements. The research employs a multi-methodological approach, combining sentiment analysis techniques with machine learning models to evaluate the relationship between social media discourse and market performance. Our findings indicate that social media sentiment demonstrates significant predictive power for stock market returns, particularly during periods of high market volatility and for specific sectors. The GameStop short squeeze of 2021 serves as a pivotal case study, illustrating how coordinated retail investor sentiment on social media can dramatically impact stock prices and challenge traditional market dynamics. This research contributes to the growing field of behavioral finance by highlighting the importance of incorporating social media sentiment analysis into investment strategies and market forecasting models. The implications extend to investors, financial institutions, and regulators seeking to understand and navigate the evolving landscape of information flow in financial markets.

Keywords: social media sentiment, stock market prediction, financial forecasting, investor behavior, sentiment analysis, market efficiency, behavioral finance, predictive analytics

1. Introduction

The relationship between information and financial markets has long been a central focus of financial research. According to the Efficient Market Hypothesis (EMH), stock prices fully reflect all available information, making it theoretically impossible to consistently outperform the market through stock selection or market timing (Fama 1970). However, the emergence of social media platforms has fundamentally reshaped the creation, dissemination, and consumption of information, challenging traditional perspectives on market efficiency and information processing.

Platforms such as Twitter, Reddit and StockTwits have become influential sources of financial information and sentiment. These channels offer real-time insights into investor opinions, market narratives and trading opportunities, often preceding traditional financial media or analyst reports. This democratization of financial discourse has amplified the role of retail investors and introduced new dynamics in market behavior, with the potential to influence asset prices through collective sentiment and coordinated action.

A defining moment in this paradigm shift was the GameStop (GME) short squeeze in early 2021. Orche-

strated largely through the Reddit forum r/WallStreetBets, a coalition of retail investors drove GME's stock price up by over 1700 percent within weeks, inflicting substantial losses on institutional investors who had shorted the stock. This episode underscored the disruptive influence of social media sentiment on financial markets and prompted renewed interest in the predictive relationship between online discourse and market outcomes.

This study seeks to examine the extent to which social media sentiment can predict stock market returns, combining a critical review of the existing literature with empirical investigation. The central research questions guiding this work are:

1. To what extent does sentiment expressed on social media correlate with subsequent stock market movements?
2. Which sentiment analysis techniques and machine learning models are most effective in capturing the link between social media discourse and financial performance?
3. How does the predictive power of social media sentiment vary across different market conditions, sectors and time horizons?
4. What are the implications of social media sentiment analysis for investment strategies, regulatory frameworks and financial theory?

Understanding the interplay between social media sentiment and stock market behavior has significant implications for a wide range of stakeholders. For investors, it offers valuable signals for sentiment-driven trading strategies. For financial institutions, it presents a tool for anticipating market shifts. For regulators, it highlights the growing importance of monitoring online discourse in safeguarding market integrity. For scholars, it provides a novel lens through which to reassess theories of market efficiency and behavioral finance.

2. Literature Review

2.1 Theoretical Framework

The relationship between social media sentiment and stock market returns can be interpreted through several theoretical lenses. The Efficient Market Hypothesis (EMH), introduced by Fama (1970), argues that stock prices fully incorporate all available information, making it impossible to consistently achieve above-average returns through stock selection or market timing. However, behavioral finance offers a contrasting view by emphasizing the influence of investor psychology and sentiment in driving market fluctuations.

Baker and Wurgler (2006) developed an influential investor sentiment index, demonstrating that sentiment can significantly impact stock returns, especially for assets that are difficult to value or arbitrage. Similarly, De Long et al. (1990) showed that noise traders who act on sentiment rather than fundamentals can create persistent mispricing in financial markets. These perspectives provide a foundation for understanding how sentiment expressed on social media may influence asset prices.

Further contributing to this framework, Kim et al. (2023) introduced the concept of social informedness, which explains how information shared through social platforms can lead to collective behavior among investors. This concept is particularly relevant for understanding phenomena like the GameStop short squeeze, where coordinated retail investor actions led to dramatic price shifts.

2.2 Social Media and Financial Markets

In recent years, a growing body of literature has examined the relationship between social media and financial markets. One of the earliest contributions was made by Antweiler and Frank (2004), who

analyzed internet stock message boards and found that online discussions contained predictive signals related to market volatility.

Bollen et al. (2011) provided further evidence by showing that public mood states on Twitter correlated with movements in the Dow Jones Industrial Average. Analyzing nearly ten million tweets, they demonstrated that specific mood indicators could predict DJIA fluctuations with an accuracy of 87.6 percent. This pioneering work opened the door for sentiment analysis as a tool for financial forecasting.

Recent studies have extended these insights. Greyling and Rossouw (2022) examined sentiment and emotion extracted from stock-related tweets and found that such information could predict intraday price movements across both developed markets such as France, Germany, Japan, Spain, the United Kingdom and the United States, and emerging markets such as India and Poland. Their findings showed that tweet sentiment achieved predictive accuracy above 50 percent across all markets studied.

2.3 Methodological Approaches

Research in this area has employed a range of methodologies to assess the link between social media sentiment and stock returns. Sentiment analysis techniques range from traditional dictionary-based models to advanced machine learning and deep learning approaches.

Pagolu et al. (2016) applied supervised machine learning techniques to sentiment analysis of tweets, using Word2vec and N-gram models. Their findings revealed a strong relationship between public sentiment on Twitter and stock price movements.

In more recent developments, Peivandizadeh et al. (2024) proposed a two-stage framework combining sentiment analysis with time-series modeling. Their method used three dilated convolutional layers to extract sentiment features and applied an Off Policy Proximal Policy Optimization algorithm to manage unbalanced classification issues. They then utilized a Transductive Long Short Term Memory model to integrate sentiment data with historical price series for prediction.

Ferraro and Sperlì (2024) developed a multimodal forecasting system that combined social textual content with historical financial data. By feeding multimodal time series into deep learning architectures, they evaluated the predictive capacity of online social network data, particularly in the context of the GameStop phenomenon.

2.4 Empirical Evidence

Empirical research has generally supported the predictive potential of social media sentiment, though results vary depending on context and methodology. The GameStop short squeeze remains one of the most emblematic case studies in this field.

Kim et al. (2023) investigated social media activity during the GameStop episode and found that user posts on Reddit significantly influenced trading behavior and transaction volumes. Their findings confirmed that the valence and volume of posts were strong precursors to irrational trading behavior.

Ferraro and Sperlì (2024) also reinforced these findings by documenting GME's 1700 percent price increase by January 27, 2021, and the corresponding losses incurred by major institutional investors such as Melvin Capital, which lost 53 percent of its capital. Their study also underscored the differential impact of various social media platforms, emphasizing the need to consider platform-specific dynamics in predictive models.

2.5 Research Gaps

Despite significant progress, several gaps remain in the literature. Many studies focus exclusively on one platform, typically Twitter, without accounting for variations across platforms such as Reddit, StockTwits or Facebook. There is often a lack of integration between advanced data science

methodologies and core financial theories. Additionally, existing research is largely centered on developed markets, with limited focus on emerging economies or non-English language data.

Methodological challenges also persist. One notable issue is unbalanced sentiment classification, where certain sentiment categories are overrepresented in the dataset, complicating model training and interpretation. Furthermore, there is insufficient research on how predictive power varies across different market phases, sectors and investment horizons.

This study seeks to address these gaps by conducting a comprehensive analysis of the predictive value of social media sentiment on stock returns, incorporating multiple platforms, methodological innovations and a broader range of market contexts.

3. Methodology

3.1 Research Design

This study adopts a multi-methodological approach to explore the predictive power of social media sentiment on stock market returns. The research design combines a critical review of existing literature with an empirical analysis of real-world case studies, with a particular emphasis on the GameStop short squeeze of early 2021 as a key instance of social media influence on financial markets.

The study follows a structured process that includes:

1. Identification of relevant theoretical frameworks and prior research.
2. Examination of methodological approaches for sentiment analysis and market prediction.
3. Review and interpretation of empirical case studies.
4. Synthesis of findings in relation to the research questions posed.

3.2 Data Sources and Collection

The analysis is informed by diverse and complementary data sources, allowing for a well-rounded understanding of the relationship between social media sentiment and stock market behavior:

1. Academic Literature: A broad review of peer-reviewed publications from disciplines including finance, economics, computer science, and digital communication was conducted. The focus was placed on studies published within the last decade that explore the linkage between social media sentiment and financial market dynamics.
2. Case Studies: The GameStop short squeeze is examined in detail, drawing on data related to stock price fluctuations, trading volume, and social media activity during key time periods.
3. Empirical Studies: Existing quantitative research providing statistical evidence on the correlation between sentiment metrics and stock returns was analyzed, with attention to different geographic markets, platforms, and timeframes.

3.3 Analytical Framework

To investigate the predictive capacity of social media sentiment, this study integrates theoretical insights and computational techniques across three main dimensions:

1. Sentiment Analysis Techniques: The study evaluates various methods for extracting sentiment from social media posts, including lexicon-based models, machine learning classifiers, and deep learning architectures.
2. Predictive Modeling: A range of models used for forecasting stock movements based on sentiment data are reviewed, from linear regression and support vector machines to more advanced tools such as Transductive Long Short-Term Memory (TLSTM) networks.
3. Platform-Based Differences: Differences in sentiment expression and predictive utility across platfor

ms are analyzed, comparing formats such as Twitter's short-form, real-time posting with Reddit's more detailed, community-oriented discussions.

3.4 Evaluation Metrics

The effectiveness of social media sentiment in forecasting market returns is assessed through several commonly used performance metrics:

1. Prediction Accuracy: The proportion of correct directional predictions (e.g., price increase or decrease).
2. Correlation Coefficients: Statistical measures indicating the strength and direction of the relationship between sentiment indicators and stock returns.
3. Root Mean Square Error (RMSE): A metric for evaluating the deviation between predicted and actual stock values.
4. Trading Performance: Simulated returns generated by trading strategies informed by sentiment analysis, compared against benchmark or passive strategies.

Together, these metrics provide a robust framework for assessing the viability of social media sentiment as a tool for predicting stock market movements.

4. Results

4.1 Correlation Between Social Media Sentiment and Stock Market Returns

Our analysis of the literature and empirical evidence reveals a significant correlation between social media sentiment and subsequent stock market returns. Multiple studies have documented this relationship across various markets, time periods, and social media platforms.

Bollen et al. (2011) found that specific Twitter mood states could predict movements in the Dow Jones Industrial Average with an accuracy of 87.6 percent. Greyling and Rossouw (2022) demonstrated that sentiment analysis of tweets could predict stock market movements with accuracy exceeding 50 percent in all markets analyzed, including both developed and emerging economies.

The correlation is especially strong during periods of heightened market volatility and for stocks characterized by high levels of retail investor activity. For example, during the GameStop short squeeze, Kim et al. (2023) observed that both the valence and the volume of posts on Reddit's r/WallStreetBets had a significant impact on GME's intraday transaction volumes.

4.2 Effectiveness of Different Methodological Approaches

Our review indicates that more sophisticated methodological approaches generally yield better predictive performance. While dictionary-based sentiment analysis offers simplicity and interpretability, machine learning and deep learning models tend to deliver superior accuracy and robustness.

Pagolu et al. (2016) demonstrated the effectiveness of combining Word2vec and N-gram models with supervised machine learning techniques for analyzing tweet sentiments. Their methodology revealed a strong correlation between Twitter-based sentiment and subsequent stock price fluctuations.

Even more advanced techniques have shown notable improvements in predictive accuracy. Peivandizadeh et al. (2024) introduced a two-stage methodology that combines sentiment analysis with Transductive Long Short-Term Memory (TLSTM) networks, enabling enhanced sensitivity to subtle variations in data. Similarly, Ferraro and Sperlì (2024) developed a multi-modal framework that integrates social media textual content with historical financial data. Their model achieved promising results in forecasting stock market trends during the GameStop event, illustrating the potential of combining diverse data sources within deep learning architectures.

Table 1. Prediction accuracy of sentiment-based stock market forecasting models from selected studies.

Study	Platform	Method	Prediction Accuracy (%)
Bollen et al. (2011)	Twitter	Mood analysis (correlation)	87.6
Greyling & Rossouw (2022)	Twitter	Sentiment + emotion classification	50.0
Pagolu et al. (2016)	Twitter	Word2Vec + N-Gram + ML	70.2
Peivandizadeh et al. (2024)	Reddit	TLSTM + PPO (deep learning)	75.4
Ferraro & Sperlì (2024)	Reddit + Twitter	Multi-modal deep learning	79.3

Source: compiled by the author based on Bollen et al. (2011), Greyling & Rossouw (2022), Pagolu et al. (2016), Peivandizadeh et al. (2024), and Ferraro & Sperlì (2024).

4.3 Platform Differences

Our analysis reveals significant differences in the predictive power of sentiment across various social media platforms. While Twitter has been the most extensively studied due to its real-time nature and wide accessibility, Reddit has emerged as particularly influential in specific contexts, especially for certain stocks and high-profile market events.

The GameStop case study underscores Reddit's unique role, particularly through the r/WallStreetBets subreddit, in organizing and amplifying retail investor activity. Kim et al. (2023) found that Reddit's platform structure, which permits longer-form posts and threaded discussions, enabled more detailed analysis and coordination among users when compared to Twitter's character-limited format.

Ferraro and Sperlì (2024) further emphasized that the influence of social media on stock market behavior varies notably by platform. Their findings reinforce the importance of incorporating multiple social media sources in predictive models, rather than relying on a single dominant platform.

4.4 Case Study: GameStop Short Squeeze

The GameStop short squeeze of early 2021 offers a vivid illustration of the predictive and catalytic power of social media sentiment on stock market dynamics. Our analysis of this event reveals several key findings:

1. Price Impact: GameStop (GME) shares experienced an extraordinary price surge of approximately 1,700 percent. The stock rose from $17.25 on January 4, 2021, to $347.51 on January 21, ultimately peaking at an intraday high of $483 on January 28 (Kim et al., 2023).
2. Social Media Coordination: This price escalation was fueled by collective action among retail investors, who coordinated their efforts through Reddit's r/WallStreetBets community. This demonstrates the potential of social media platforms to enable large-scale investor mobilization (Ferraro and Sperlì, 2024).
3. Institutional Impact: The short squeeze inflicted significant losses on institutional investors holding short positions in GME, most notably Melvin Capital, which reported a 53 percent loss on its investments (Ferraro and Sperlì, 2024).
4. Sentiment Indicators: The volume and emotional valence of user submissions on r/WallStreetBets were found to be strong predictors of GME's intraday trading activity, providing empirical support for the role of sentiment as a leading indicator (Kim et al., 2023).

This case study illustrates that social media sentiment can do more than forecast market movements—it can directly shape them, especially when it fosters coordinated behavior among a critical mass of retail investors.

5. Discussion

5.1 Implications for Market Efficiency

The predictive power of social media sentiment on stock market returns has significant implications for the Efficient Market Hypothesis (EMH). Traditional formulations of the EMH assert that stock prices reflect all available information, making it impossible to consistently earn above-average returns through stock selection or market timing (Fama 1970).

However, the findings of this study suggest that social media sentiment may constitute a form of information that is not immediately incorporated into stock prices, creating opportunities for prediction and arbitrage. This perspective aligns with behavioral finance theories that emphasize the psychological and emotional dimensions of investor behavior (Baker and Wurgler 2006).

The GameStop case study challenges the strong form of market efficiency by illustrating how coordinated retail investor sentiment on social media significantly influenced stock prices, in ways not immediately captured by conventional indicators. This suggests that social media may introduce new forms of information asymmetry and collective dynamics that traditional efficiency models fail to address.

5.2 Practical Applications for Investors

The results of this research have several practical applications for investors interested in incorporating social media sentiment into their decision-making processes:

1. Sentiment Monitoring: Investors can implement monitoring systems that track sentiment across various social media platforms, especially those with high levels of discussion surrounding specific stocks or sectors.
2. Signal Generation: Sentiment data can be integrated into quantitative trading models to generate buy or sell signals, thereby enhancing existing technical or fundamental analysis strategies.
3. Risk Management: Detecting sudden shifts in online sentiment can help investors anticipate potential volatility or disruption in financial markets, enabling more proactive risk mitigation.
4. Platform Diversification: Given the differing levels of predictive power among platforms, it is advisable to monitor sentiment across multiple social media sources rather than relying on a single channel.

Nonetheless, investors should remain mindful of the challenges associated with sentiment-based strategies, including the volatility of online discourse, platform-specific biases, and the fluid nature of digital communities.

5.3 Regulatory Considerations

The growing influence of social media sentiment on financial markets also raises important regulatory questions. The GameStop short squeeze sparked global debate over the potential for market manipulation through social platforms and highlighted the limitations of existing regulatory frameworks. Key areas of concern for regulators include:

1. Monitoring Capabilities: Enhancing surveillance tools to detect coordinated actions and sentiment-driven disruptions in real time.
2. Disclosure Requirements: Reevaluating whether current disclosure regulations adequately account for the impact of social media on investment behavior.
3. Platform Responsibilities: Defining the obligations of social media companies in curbing misinformation or manipulative behavior that could affect market integrity.
4. Retail Investor Protection: Balancing the democratization of market participation enabled by social

media with the need to safeguard less-experienced investors from undue risk.

As noted by Kim et al. (2023), more robust monitoring of social platforms may be necessary to mitigate systemic risk while maintaining the benefits of broader access to information and participation.

5.4 Limitations and Future Research Directions

While this study contributes to the growing body of knowledge on social media sentiment and financial markets, several limitations must be acknowledged, along with avenues for future inquiry:

1. Causality vs. Correlation: Determining the direction of causality remains complex, as sentiment may both influence and be influenced by market movements, or be driven by third variables.
2. Sentiment Classification Challenges: Accurately interpreting sentiment remains difficult due to factors such as sarcasm, contextual variation, and platform-specific language patterns.
3. Market Adaptation: As awareness of social media's influence increases, market participants may adjust their behavior, altering the strength or nature of sentiment-driven effects over time.

To advance research in this area, future studies should consider the following:

1. Development of Advanced Models: Employing multi-source data integration and cutting-edge machine learning algorithms to improve sentiment classification and forecasting.
2. Cross-Platform Dynamics: Investigating how sentiment propagates across different platforms and how these interactions influence investor behavior.
3. Longitudinal Analyses: Conducting time-series research to evaluate how the sentiment–market relationship evolves over different periods and under varying market conditions.
4. Sector-Specific Investigations: Exploring whether the impact of sentiment varies across industries or types of assets.
5. International Expansion: Extending research to non-English platforms and emerging markets to identify broader global trends and regional differences.
6. These lines of inquiry are essential to deepen understanding of how social media sentiment shapes modern financial markets and to refine the theoretical and practical tools available to investors, analysts and regulators.

7. Conclusions

This study investigated the predictive power of social media sentiment on stock market returns through a comprehensive review of existing literature and empirical evidence. The findings indicate that social media sentiment possesses significant predictive capacity, especially during periods of heightened market volatility and for stocks characterized by strong retail investor participation.

The GameStop short squeeze of early 2021 serves as a pivotal case study, illustrating how coordinated retail investor sentiment on social media can dramatically impact stock prices and challenge traditional market dynamics. This event highlighted the potential for social media platforms, particularly Reddit, to facilitate collective investor behavior with significant market implications.

Our analysis reveals that more sophisticated methodological approaches, such as those combining advanced sentiment analysis techniques with machine learning models, tend to yield better predictive performance. Additionally, we find important differences in the predictive power of sentiment across different social media platforms, emphasizing the value of considering multiple platforms when analyzing social media sentiment for stock market prediction.

The predictive relationship between social media sentiment and stock market returns has significant implications for market efficiency theories, investment strategies, and regulatory frameworks. It suggests

that social media platforms may create new information dynamics that are not fully accounted for in traditional market efficiency theories, offering potential opportunities for prediction and arbitrage while also raising concerns about market manipulation and investor protection.

As social media continues to evolve and its influence on financial markets grows, understanding the predictive power of social media sentiment on stock market returns will become increasingly important for investors, financial institutions, regulators, and academics. Future research should focus on developing more sophisticated models, exploring cross-platform dynamics, conducting longitudinal studies, investigating sector-specific effects, and extending analysis to international markets.

In conclusion, social media sentiment represents a valuable source of information for predicting stock market returns, offering insights that complement traditional financial analysis and potentially enhancing investment decision-making in an increasingly connected and information-rich market environment.

References

1. Antweiler W., Frank M.Z., “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”, The Journal of Finance, 2004, 59 (3), 1259–1294.
2. Baker M., Wurgler J., “Investor Sentiment and the Cross-Section of Stock Returns”, The Journal of Finance, 2006, 61 (4), 1645–1680.
3. Bollen J., Mao H., Zeng X., “Twitter's Mood Predicts the Stock Market”, Journal of Computational Science, 2011, 2 (1), 1–8.
4. De Long J.B., Shleifer A., Summers L.H., Waldmann R.J., “Positive Feedback Investment Strategies and Destabilising Rational Speculation”, The Journal of Finance, 1990, 45 (2), 379–395.
5. Fama E.F., “Efficient Capital Markets: A Review of Theory and Empirical Work”, The Journal of Finance, 1970, 25 (2), 383–417.
6. Ferraro A., Sperlì G., “How Does User-Generated Content on Social Media Affect Stock Predictions? A Case Study on GameStop”, Online Social Networks and Media, 2024, 43–44, 100293.
7. Greyling T., Rossouw S., “The Prediction of Intraday Stock Market Movements in Developed and Emerging Markets Using Sentiment and Emotions from Twitter”, Finance India, 2022, 36 (3), 907–939.
8. Kim K., Lee S.Y.T., Kauffman R.J., “Social Informedness and Investor Sentiment in the GameStop Short Squeeze”, Electronic Markets, 2023, 33 (1), 23.
9. Pagolu V.S., Reddy K.N., Panda G., Majhi B., “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements”, 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016, 1345–1350
10. Peivandizadeh A., Hatami S., Nakhjavani A., Khoshsima L., Chalak Qazani M.R., Haleem M., “Stock Market Prediction With Transductive Long Short-Term Memory and Social Media Sentiment Analysis”, IEEE Access, 2024, 12, 1–1.