Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD31028

Forecasting Financial Market Volatility Based on Social Media Sentiment

Tingxuan Ren

Jinan University – University of Birmingham Joint Institute, Jinan University, Guangzhou, China
TXR349@student.bham.ac.uk

Abstract. Predicated on a critical review of the Efficient Market Hypothesis within the modern financial landscape, this research supposes that social media functions work as a barometer of public sentiment, conveying emotional data that potentially possesses predictive utility for the financial market volatility. Research using the S&P 500 index and text-sensitive stocks as the study objects, this paper analyzes Twitter text data from 2021–2022. By consolidating econometric GARCH-type models with machine learning techniques (LSTM and XGBoost), the study performs sentiment quantification and forecasts market volatility. The study reveals that traditional econometric methods are unable to identify the predictive link between market sentiment and volatility. However, the LSTM model outperforms conventional GARCH (1,1) and GARCH-X models by capturing nonlinear relationships and effectively extracting predictive signals to create a robust forecasting model. Moreover, the results highlight an asymmetric effect: positive sentiment drives market volatility more intensely than negative sentiment. In conclusion, social media sentiment holds predictive potential for financial markets. Yet its relationship with volatility is complex and nonlinear—deep learning models are required to fully capture the relationship. Also, the effect of sentiment on volatility is asymmetric.

Keywords: Financial Market Volatility Forecasting, Social Media Sentiment Analysis, Machine Learning

1. Introduction

The Efficient Market Hypothesis (EMH) serves as a foundational pillar of traditional finance, stipulating that asset prices instantaneously and comprehensively incorporate all available information. Nevertheless, existing studies suggest that real-world markets frequently deviate from this theoretical state of perfect efficiency [1]. Asset prices exhibit a degree of predictability, a phenomenon typically stemming from market underreaction to incoming information [1]. Functioning as real-time barometers of public sentiment, platforms like Twitter and Reddit offer market participants an unprecedented stream of data [2]. The advent of such information streams constitutes a challenge to the strong and semi-strong forms of the EMH. If sentiment derived from social media possesses the capacity to systematically forecast market volatility, this suggests that there is publicly available information that is not immediately incorporated into market prices. Building upon this theoretical framework, the present study seeks to explore the capacity of social

DOI: 10.54254/2754-1169/2026.LD31028

media sentiment as an alternative data source to reflect market inefficiency and to assess its applicability in quantitative models for predicting financial market volatility. Preliminary Granger causality and GARCH analyses indicated that the daily sentiment index lacks a significant linear predictive nexus with the subsequent day's realized volatility (i.e., results were statistically insignificant). Consequently, to address potential nonlinearities, this research developed a stacked LSTM model. After hyperparameter optimization, the model demonstrated robust fitting capabilities, validating the hypothesis that a complex nonlinear relationship exists between sentiment and volatility.

Overall, the research validates social media sentiment as a critical nonlinear predictor for market volatility. It further demonstrates that deploying advanced machine learning models, notably LSTM, is essential for identifying such complex associations.

2. Methodology and results

As a preliminary step, Granger causality analysis is utilized to determine the predictive utility of one time series regarding another, testing primarily for predictive power rather than physical causation.

If time series X is a Granger cause of time series Y, then forecasting Y's future values using the historical values of X yields superior results compared to using Y's historical values alone. This implies that X contains incremental information useful for predicting Y that is beyond Y's own history.

In this study, this method is applied to examine the relationship between the daily social media sentiment index (S) and the next-day realized volatility (RV).

2.1. GARCH model

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model adopted in this study is the most classic model in financial econometrics for modeling and forecasting time series volatility [3].

The model is chiefly employed to account for the dynamic nature of financial asset volatility—specifically, that it is time-varying rather than static, and characterized by 'volatility clustering' [4]. This aligns precisely with the objective of this study, which seeks to analyze the predictive relationship between the social media sentiment index and financial market volatility.

2.2. Machine learning

Regarding the subsequent analysis, the study utilizes LSTM (Long Short-Term Memory) and XGBoost (eXtreme Gradient Boosting). These algorithms represent the state-of-the-art in contemporary machine learning and deep learning, recognized for their widespread adoption and superior performance in practical applications.

2.2.1. LSTM model

The LSTM model is a specialized variant of Recurrent Neural Networks (RNNs), specifically designed to address the vanishing and exploding gradient problems encountered during the training of long sequential data. It is capable of capturing causal relationships across distant time intervals and, as a deep learning architecture, can effectively fit extremely complex nonlinear relationships.

DOI: 10.54254/2754-1169/2026.LD31028

2.2.2. XGBoost model

XGBoost, a decision-tree-based ensemble algorithm, excels at handling tabular data, typically surpassing deep learning models in this domain. Its advantages include fast training, strong robustness, and interpretability via feature importance metrics. Consequently, it is utilized here to assess the comparative influence of positive and negative sentiment on volatility.

2.2.3. Data set & variables definition

The data employed in this study is sourced from the Kaggle dataset titled 'Stock Tweets for Sentiment Analysis and Prediction.' The core value of this dataset lies in its direct association of raw tweets with corresponding dates, stock prices, and trading volume data.

The study applies VADER and FinBERT for sentiment analysis. It is noted that traditional lexicon-based methods often lack the capacity to capture financial context and nuances. In contrast, FinBERT—a pre-trained NLP model engineered for finance—surpasses these traditional models by effectively parsing the semantics of financial text [5]. In the quantification process, it is posited that sentiment coinciding with high trading volumes signals robust investor belief and thus influences volatility more profoundly [6].

Accordingly, the study calculates the mean sentiment for each trading day, weighting this average using the natural logarithm of trading volume, to obtain the final Daily Social Media Sentiment Index (S).

Financial market volatility is measured by the rolling standard deviation of stock market returns, referred to herein as the next-day realized volatility (RV).

2.3. Research process & results

Initially, to examine the relationship between the daily social media sentiment index and realized volatility, Granger causality analysis is conducted. The null hypothesis (H0) states that the sentiment index is not the Granger cause of volatility.

2.3.1. Prior research

Initially, to examine the relationship between the daily social media sentiment index and realized volatility, Granger causality analysis is conducted. The null hypothesis (H0) states that the sentiment index is not the Granger cause of volatility.

In order to broaden the applicability of the findings, lag orders 1 through 5 were tested. At lag 1, the results ($F=3.8046$, $p=0.0519$) failed to surpass the 0.05 significance threshold. Thus, the null hypothesis could not be rejected. As the F-value declined with higher lags (2–5), the study concludes that the relationship achieved only marginal significance at lag 1.

Given that Granger causality analysis in detecting linear relationships is still limited, the aforementioned results suggest that the linear relationship between the daily social media sentiment index and realized volatility is not significant in statistics.

The study assumes a secondary relationship between the daily social media sentiment index and realized volatility. This hypothesis is grounded in existing research indicating that extreme sentiment leads to an increase in future financial market volatility. No matter the positive or negative sentiment [7]. As this pattern approximates a secondary function, the adjusted equation is presented as follows:

DOI: 10.54254/2754-1169/2026.LD31028

$$Y_t = c_2 + \sum_{i=1}^{p} \left(\gamma_i Y_{t-i} + \delta_i Y_{t-i}^2\right) + \eta_t \tag{1}$$

With F=0.798123 and p=0.332040, the results indicate an inability to reject zero hypothesis.

Consequently, it is concluded that neither a linear nor a quadratic relationship exists between the daily sentiment index and realized volatility in a statistically significant sense. Nonetheless, the application of GARCH analysis is still needed.

To validate the data for modeling, the Augmented Dickey-Fuller (ADF) test was used to assess the stationarity of the stock market log returns. The obtained ADF statistic (-15.655289) falls well below the critical threshold (-2.862003). Thus, the null hypothesis of a unit root is rejected, indicating that the log return series is stationary at the 5% significance level.

To test for conditional heteroskedasticity (ARCH effects) in the log returns, an ARCH-LM test was conducted. The results indicate a significant LM statistic of 122.507322 ($p < 0.001$). The residuals display a mean of μ= -0.001495 (effectively zero) and a standard deviation of 0.032215. Based on the previous analysis results of data characteristics, the series shows an extremely high kurtosis (11.457819), significantly exceeding the normal distribution's benchmark of 3. This is a typical characteristic of volatility clustering—a phenomenon where financial market volatility tends to appear in clusters, with periods of similar volatility persisting over time, thereby demonstrating inertia [8]. Both tests yielded p-values below the 0.05 significance level, indicating the presence of significant ARCH effects within the dataset. Therefore, the application of the GARCH model is appropriate.

2.3.2. GARCH model results

The aforementioned analysis confirms the suitability of the GARCH model for this dataset. Initially, the standard GARCH (1,1) model was applied, resulting the following ARCH term coefficient: α=0.1000, GARCH term coefficient:β=0.8800. In this model, α measures the impact of past market shocks on current volatility, and β measures the impact of past volatility on current volatility. Analysis of the ARCH effect reveals that shocks from advance returns contribute approximately 1% to current volatility. In contrast, the GARCH effect indicates that past volatility accounts for 88% of the influence. For persistence, the sum of coefficients ($\alpha+\beta$ = 0.98) is very close to 1, demonstrating a high degree of volatility persistence. This implies that market volatility shocks tend to decay slowly and last for an extended period.

To address the limitations of the standard model in capturing comprehensive volatility dynamics, the framework was augmented by introducing S_{t-1}(the lagged daily sentiment index) as an exogenous regressor in the variance equation [4]. The specific conditional variance equation is given by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma S_{t-1} \tag{2}$$

In detail, the estimated coefficient for the sentiment index is γ = -0.000116 (p=0.703), which is not statistically significant. In response, the model was re-estimated using the Student's t-distribution and the Generalized Error Distribution (GED) to account for the leptokurtic (fat-tailed) nature of the data. These findings are summarized in Table 1, with corresponding graphical representations in Figure 1.

DOI: 10.54254/2754-1169/2026.LD31028

Table 1. Statistical significance of GARCH-X results under different distributions

Dist. / par.	Loglikelihood	Γ coefficient	Change	significance
Normal dist.	836.8574	-0.0016	-	×
T dist.	833.0227	-0.000061	+47.25%	×
GED dist.	970.3587	-0.000122	-5.6%	×

Thus, we reject the hypothesis of a significant statistical link between S and RV. In specific terms, empirical evidence suggests that no linear statistical relationship exists between these variables.

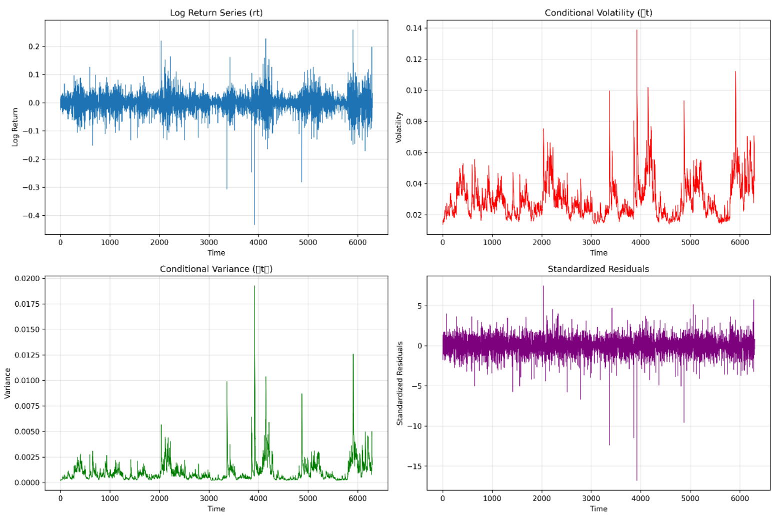

Figure 1. GARCH-X result

2.3.3. LSTM & XGBoost results

Considering that the aforementioned econometric methods failed to gain sturdy conclusions, this study proceeds to employ two advanced machine learning techniques—XGBoost and LSTM to investigate the predictive relationships and evaluate the feature importance of various parameters in forecasting. Initially, XGBoost analysis was conducted to quantify the predictive weight of sentiment features. The model results indicate that the aggregate feature importance of sentiment is merely 0.76%, whereas the historical values of RV itself account for 99.24% of the predictive contribution. However, existing literature posits that S, acting as an agent variable for market irrationality and emotions, is a significant forecasting estimator among the non-price features. It reflects the collective psychological state of investors and their tendencies toward irrational behavior, which are known drivers of market volatility [9,10]. The fact that the subsequent LSTM analysis achieved a model with superior goodness-of-fit suggests that the predictive relationship between RV and S is basically nonlinear (and thus better captured by deep learning).

Subsequently, to derive the predictive model, construct an LSTM framework. The training pipeline proceeded as follows: four features were extracted from the dataset, and lag-p terms were generated for each feature. The data underwent cleaning, with missing values handled by forward fill imputation. All features were standardized using Z-score normalization. The dataset was then partitioned into training, testing, and validation sets in a 70:15:15 ratio. The model framework consists of a dual-layer LSTM with Dropout and Dense layers, compiled using the Adam optimizer

DOI: 10.54254/2754-1169/2026.LD31028

and Mean Squared Error (MSE) loss function. The final model performance is illustrated in Figure 2.

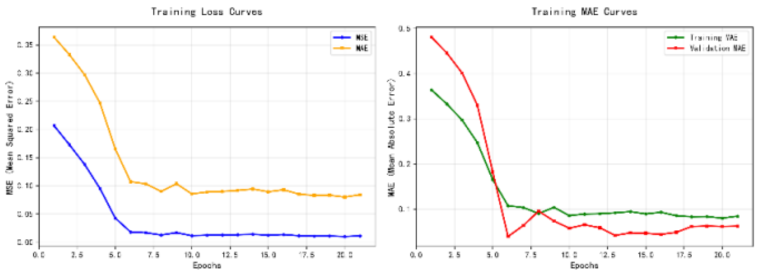

Figure 2. LSTM training results

Conclusion: While the influence of sentiment on volatility is not characterized by simple linearity, advanced machine learning models like LSTM—distinguished by their ability to model long-term dependencies and complex nonlinearities—demonstrate efficacy in extracting the underlying predictive signals.

To investigate the distinct impacts of positive and negative sentiment on financial market volatility, this study employs both econometric approaches (GARCH) and machine learning methods. Initially, the previous GARCH model was modified by adding the lagged terms of the positive and negative sentiment indices to the variance equation. The revised variance equation is specified as follows:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma_{pos}S_{pos,t-1} + \gamma_{neg}S_{neg,t-1} \tag{3}$$

Here, $\boldsymbol{S_{pos,t-1}}$ and $\boldsymbol{S_{neg,t-1}}$ define the lagged positive and lagged negative sentiment indices, respectively, with $\boldsymbol{\gamma}$ indicating their associated coefficients. These parameters are analyzed to determine the extent of their predictive influence. The estimation results reveal a coefficient of -0.000172 for positive sentiment and 0.000090 for negative sentiment. Consequently, the GARCH analysis (Figure 3) demonstrates that elevated positive sentiment tends to mitigate market volatility, while heightened negative sentiment exacerbates (increases) it.

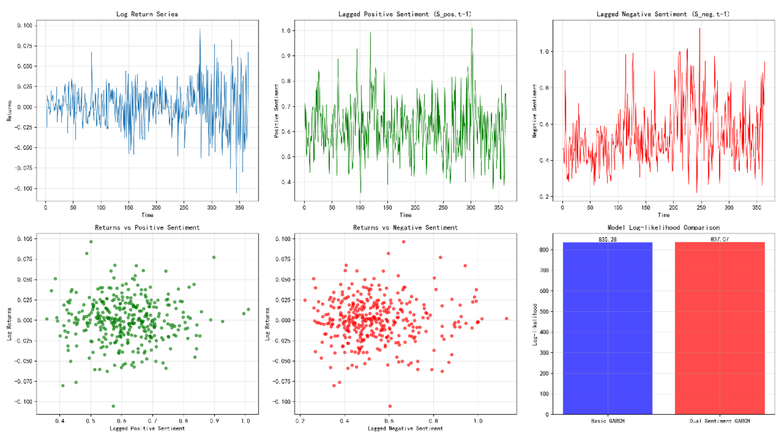

Figure 3. GARCH-X results of positive & negative sentiment

Subsequently, an XGBoost analysis was conducted to rank the feature importance of the positive and negative sentiment evidence. The results are as follows: Market Returns (36.92%) emerged as

DOI: 10.54254/2754-1169/2026.LD31028

the most critical feature, followed by Trading Volume (21.68%), the Positive Sentiment Index (21.56%), and the Negative Sentiment Index (19.84%).

The cumulative contribution of sentiment metrics stands at 41.40%. In general, the positive sentiment index demonstrates higher importance than its negative counterpart (see Figure 4). The fact that sentiment indicators comprise 41.40% of the model's predictive power validates the substantial impact of social media sentiment. Nevertheless, traditional market variables (returns and trading volume) retain a dominant role, with positive sentiment identified as a significantly stronger predictor than negative sentiment.

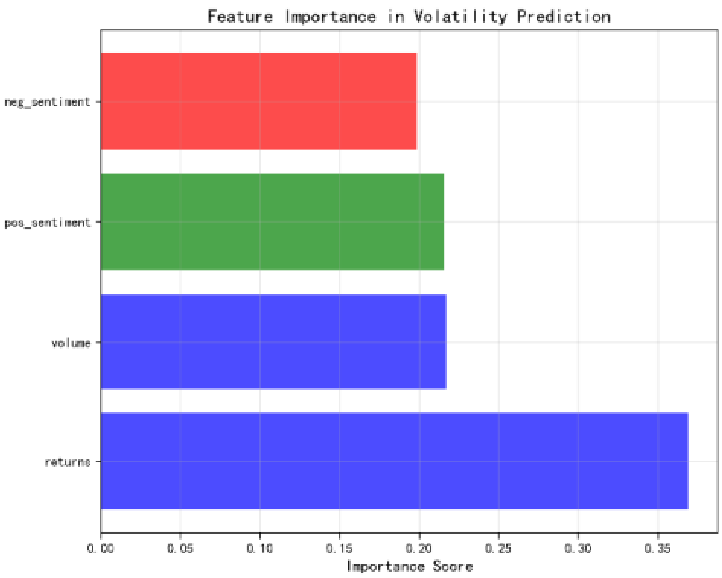

Figure 4. Importance of positive and negative sentiments

Synthesizing the findings, Empirical evidence derived from the GARCH and XGBoost analyses demonstrates that positive sentiment exerts a significantly stronger influence on market volatility than negative sentiment.

3. Conclusion

This study systematically investigates the feasibility of forecasting financial market volatility based on social media sentiment. In the early stage, Granger causality analysis was used to rule out the possibility of simple linear or quadratic predictive relationships between sentiment and volatility. Granger causality analysis was employed to rule out the possibility of simple linear or quadratic predictive relationships between sentiment and volatility. The construction of a benchmark GARCH(1,1) model confirmed the presence of significant volatility clustering in the market return series.

However, after introducing the sentiment index as an exogenous variable into the GARCH-X model, no significant linear explanatory power could be observed. To capture potential nonlinear dynamics, this study turns to machine learning methodologies. The XGBoost analysis reveals that although the aggregate feature importance of sentiment (0.76%) is significantly lower than that of historical market data (99.24%), sentiment remains a critical proxy variable external to the market data itself. Furthermore, the constructed dual-layer LSTM model, following parameter optimization, achieved a low Root Mean Square Error (RMSE) of 0.066712 on the test set. This result

DOI: 10.54254/2754-1169/2026.LD31028

demonstrates the model's capacity to effectively capture the nonlinear relationships and long-term dependencies between sentiment and volatility.

Lastly, to investigate the asymmetric effects of sentiment, the research uses GARCH-X and XGBoost to evaluate positive and negative sentiment evidence. The findings uniformly reveal that positive sentiment drives volatility more significantly than negative sentiment. In conclusion, this experiment validates social media sentiment, especially positive sentiment, as a significant nonlinear predictor of market volatility. It highlights that advanced machine learning architectures are pivotal for extracting these complex associations.

This study provides valuable implications for future research in this domain, mainly by promoting a methodological pattern shift in volatility forecasting from linear econometrics to nonlinear intelligent computing. The significant performance difference observed between the GARCH and LSTM models suggests that the rigid linear structure of GARCH is ill-equipped to accommodate social media sentiment, which is a variable characterized as high-noise and unstructured. Hence, this finding suggests that future research should no longer be limited to examining the significance of the coefficient of sentiment factors by traditional econometrics. Scholars should shift towards using deep learning to capture the long-term reliance and nonlinear coupling relationships that exist between S and RV.

While the predictive contribution of sentiment appears numerically negligible compared to historical market data, it works as the first agent variable independent of endogenous trading metrics. So, it provides a 'marginal gain' in predictive value that is non-negligible. It is important to acknowledge the study's limitations. While the experimental design included tests for quadratic Granger causality, significant associations were not identified. It suggests that the interplay between the variables involves intricate nonlinearities beyond simple quadratic structures. To address this, future work is advised to adopt information theory-based methods (e.g., Transfer Entropy) for a more robust assessment of the predictive relationship. Despite the robust forecasting capabilities demonstrated by the machine learning approaches, the inherent black-box characteristic of LSTM limits the transparency regarding the underlying mechanisms of sentiment-induced volatility. Furthermore, while XGBoost offers insight into feature importance, it lacks the capacity to elucidate the sign of the relationships (positive versus negative) as visually as the coefficients in the GARCH framework.

Additionally, the study relies on a single channel for data collection, as sentiment is predominantly sourced from social media. Future work should aim to fuse multi-source information, including news articles and analyst reports, to develop a holistic sentiment index characterized by a superior signal-to-noise ratio. This approach would yield more robust forecasting results. In light of the study's findings and limitations, the following ways for future research are suggested:

Firstly, the Multimodal Sentiment Analysis. Integrate sentiment information from heterogeneous data sources, such as fusing news articles with professional analyst reports, to construct a more comprehensive sentiment indicator. Then, High-Frequency Data Research. Apply the proposed research framework to daily high-frequency data to catch the impact of sentiment across a shorter time range. Moreover, the Advanced Sentiment Index Construction. Develop sophisticated aggregation algorithms that assess user credibility—based on historical discourse and social network relationships—to serve as weighting mechanisms. Lastly, Quantitative Trading Strategy Development: Leverage the derived predictive models to develop actionable quantitative trading strategy tools for practical application.

DOI: 10.54254/2754-1169/2026.LD31028

References

[1] Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. arXiv preprint arXiv: 2304.07619.
[2] Carter, E. J., Dupont, R. M., & Johnson, J. (2025). Analyzing the Impact of Reddit and Twitter Sentiment on Short-Term Stock Volatility. ResearchGate.
[3] Philippas, D., Ranta, M., & Vasileiou, E. (2024). Media sentiment and financial markets: A survey of the literature. International Review of Financial Analysis, 95, 103390.
[4] Zhang, Y., & Wang, J. (2021). The predictive power of investor sentiment for stock market volatility: Evidence from a GARCH-MIDAS model. The North American Journal of Economics and Finance, 58, 101515.
[5] Dogu, M., Semeno, K., & Teti, E. (2024). The impact of sentiment on stock returns: Evidence from FinBERT. International Review of Financial Analysis, 93, 103188.
[6] Zhang, Y., Song, W., & Shen, D. (2022). The roles of public attention and sentiment in the stock market: Evidence from China. Finance Research Letters, 47, 102668.
[7] Umar, Z., Jareño, F., & González, M. O. (2021). The impact of media sentiment on the volatility of the US stock market sectors. International Review of Economics & Finance, 76, 1162-1176.
[8] Fakhfekh, M., Jeribi, A., & Ghorbel, A. (2021). Volatility dynamics and persistence of the Islamic and conventional stock markets. Quantitative Finance and Economics, 5(2), 329-353.
[9] Debata, B., & Mahakud, J. (2021). Investor sentiment and stock market volatility: Evidence from emerging markets. Journal of Asia-Pacific Business, 22(3), 205-227.
[10] Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. International Journal of Forecasting, 36(2), 334-357.