# Sector-specific forecasting of volatility from Twitter sentiment

Nicole Liang* sl5332@barnard.edu

Jowe Sham* jas2533@barnard.edu

Yiming Song* ys3635@columbia.edu

July 2023

**Abstract**

In this paper, we examine the usefulness of Twitter sentiment data when applied to the prediction of volatility across 11 different market sectors. To do so, we acquire information about both public sentiment and stock volatility. With the rise of social media usage globally, we believe that platforms such as Twitter are influential sources of public opinion. Volatility is a highly useful metric that informs investors about risk and potential returns. We first compute the Pearson correlation and test for Granger causality and find statistically significant relationships between market sentiment and volatility for certain sectors, notably the technology and consumer discretionary sectors. We then apply these findings to four different machine learning models. We compare cross-model performance and find that a long short-term memory (LSTM) model outperforms all other models (linear regression, decision tree, XGBoost) when predicting future volatility, with an order-of-magnitude lower mean absolute percentage error, both with and without sentiment data input. The LSTM model is also the only model that has a statistically significant improvement upon the inclusion of sentiment data. This suggests that the true underlying relationship between sentiment and future volatility is highly complex and difficult to capture with conventional machine learning models that rely on regression alone.

*Equal contribution

# 1 Introduction

Whether stock prices are predictable or not has been a long-standing question in mathematical modeling. Stock pricing can be considered a stochastic process, and can be extremely susceptible to news, events, and future information. The measure of how much stock prices can change is volatility, which provides an estimation of price fluctuations, allowing investment risk to be quantified even if the prices of the stocks themselves are inherently difficult to predict.

While there are traditional statistical methods of predicting volatility that rely purely on past values, investors today have access to immediate textual information about stocks, from news and social media– a major factor influencing the decisions of investors. Advances in natural language processing (NLP) have also allowed for financial market sentiments to be extracted automatically from such sources. We want to explore the impact of such data on specific market sectors. To do so, we will explore how changes in public sentiment can affect the magnitude of the volatility of different stock market sectors. Applying sentiment data to predicting financial market data has been attempted successfully in the past with works such as [BNA+21], [LXC+14], and [JWW21], although these authors applied sentiment data to predict changes in stock price, not volatility.

In this paper, we will find a way to model changes in the volatility of different market sectors based on sentiment data and determine whether sentiment-driven volatility is more prominent in certain market sectors. Our findings will be helpful in informing investors about hedging strategies and risk management techniques. Our analysis will enable us to compare the effect that public sentiment change has on volatility across different market sectors, allowing us to determine the robustness of investments made in these various sectors. Having a good sense of volatility outlook helps investors with risk assessment, diversification, and return expectations. Investors that are more risk-averse can decide to stay out of more volatile sectors when excessive sentiment is detected, and vice versa.

Our paper is structured as follows: section two contains a review of relevant literature surrounding sentiment data extraction and processing as well as stock market prediction methods; in section three, we discuss our approach to acquiring and processing data, running statistical tests and building machine learning models; in section four and five, we discuss our results and present a conclusion and future avenues.

# 2 Literature Review

[JWW21] propose a model which uses both technical indicators (momentum index, moving average, etc.) and sentiment analysis to predict stock market prices. For their sentiment classification, they used Word2vec, a 2013 Google deep learning tool developed to obtain a vector representation of words from large datasets. [GZ23] aggregates public sentiment using FinBERT, a variation of BERT that is further trained on financial data. BERT addresses issues that vectorization-based models can have, including information loss when out-of-vocabulary words are encountered, and a lack of readily available data. It is designed specifically for financial sentiment classification and outperforms all previous models [YUH20].

For closing price prediction, [JWW21] use the long short-term memory (LSTM) neural network model for one-day ahead closing price prediction. LSTMs have been used frequently in financial market forecasting: [dOCCdS21] use it to predict market prices based on inputs from multiple time series. Similarly, we will be using LSTM to predict volatility using sentiment.

The effect that sentiment change can have on stock volatility is also documented. [BMG20] observe the effects that COVID has on stock market volatility. They obtained daily financial information from Federal Reserve databases and performed a regression with COVID proxies. They found a highly statistically significant link between COVID indicators and volatility across all market sectors. They also found that changes in market volatility are more sensitive to COVID news than economic indicators. We will use comparable regression and statistical significance methods to detect the correlation between our time series, which is sentiment indicators and volatility. [RAC+15] apply the Pearson correlation test and Granger causality to see the effect that sentiment has on volatility over a 15-month time period. They take data from Twitter and the Dow Jones Industrial Average Index. They find that tweet volume Granger-causes volatility for a third of the companies included, while sentiment is not useful in predicting price, which corroborates the findings of the following authors.

[DGPP22] also uses Granger causality to verify the accuracy of using market sentiment to predict FTSE100 index movements. Using the causality test, they find that the market sentiment has no predictive power. However, results from a Pearson correlation test indicate that there is a statistically significant relationship between time-lagged sentiment and market volatility. Notably, negative sentiment has a greater impact on volatility than positive sentiment, suggesting that consumers react to negative sentiment more extremely. Their results imply that news headlines are lagging indicators that do not have strong predicted power on stock price returns, and suggest that it is much more feasible to analyze the effect of sentiment on volatility than market returns.

# 3 Methodology

Figure 1 below shows an outline of our methodology, which has three main sections: data processing, statistical tests, and modeling.

## 3.1 Data Acquisition and Processing

### 3.1.1 Sentiment

From online blogs, we collect over one hundred and thirty finance-related Twitter usernames to extract sentiment from. We searched for the most followed

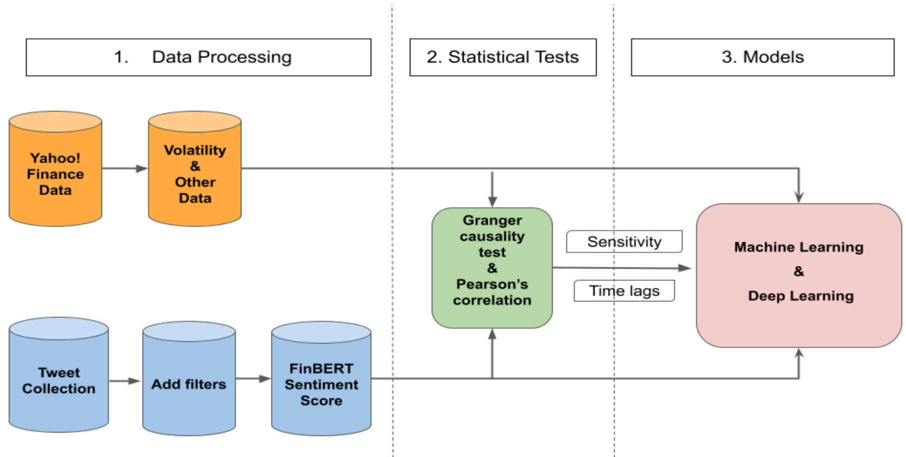



Figure 1: Methodology overview.

finance-related accounts, which include individuals, such as economics professors at prestigious universities, company executives, and esteemed market analyzers, as well as institutions, such as leading news companies, asset management firms, and investment banks. We believe that these users are more representative of market sentiment compared to the more general userbase of Twitter as the public would perceive them as more knowledgeable and experts of the industry. We perform this additional step of user selection in order to avoid downloading a high volume of irrelevant data. We scrape all tweets from these users since Jan 1, 2022, and we filter for tweets with a minimum of 20 likes and 5 retweets to account for popular influence. We obtain a total of 100,658 tweets across this time period.

Then, to further ensure that the tweets are relevant to the financial market and general sentiment, we filter according to a list of financial terms, including names of corporations, market-related terminology, terms that signal movements such as decline, rise, and boost, as well as words related to emotion such as "feel", "happy", etc. We end up with 85,799 tweets.

Using FinBERT, we classify each of these tweets according to sentiment. The model provides a label (positive, negative, or neutral), and a certainty value (between 0 and 1) for each text input. A higher magnitude value means the model is more certain of the tweet's positivity or negativity. For example, the tweet `"Business has gotten stupid slow, and we estimate having many days of just a few hours' work due to low volume."` is assigned a `Negative` label with a score of `0.99`, indicating that the model is very certain that the tweet is indicative of negative sentiment. We compute the mean sentiment scores of all tweets from each day, and figure 2 shows the distribution of our average daily sentiment. All the average sentiment values are negative. This indicates that online content, from financial news sources or experts in the field, mostly expresses pessimism, and that such news receives more engagement from the Twitter userbase. Past literature suggests that this is indeed the case, with negative news sparking more engagement and clicks [RPS+23].

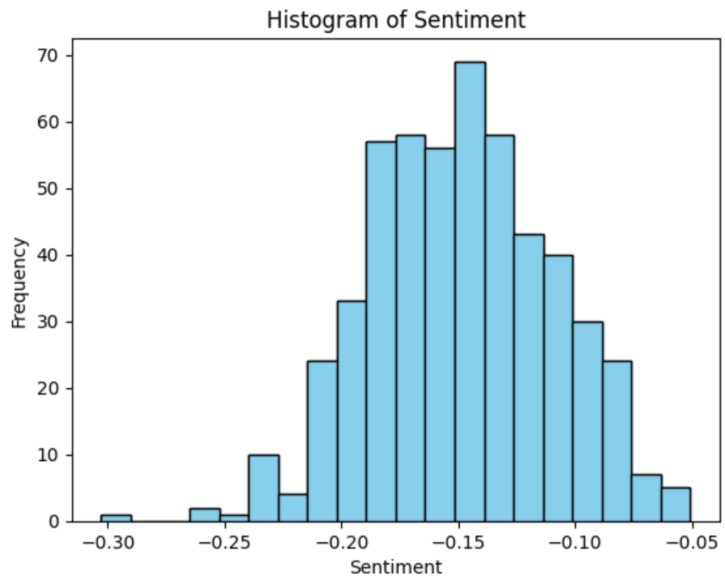



Figure 2: Distribution of average daily sentiment scores.

#### 3.1.2 Volatility

We download US exchange-traded fund (ETF) data from all eleven major sectors in the same time frame as a proxy for whole-sector performance. ETFs are fund indices that are aggregates of other assets, in our case, other individual stocks; thus, sector-specific ETFs provide a reference for how the entire sector is generally performing. We choose eleven ETFs from Vanguard for consistency across all market sectors. Vanguard Group is one of the largest registered investment advisors and the second-largest ETF provider after BlackRock. Each index is composed of from 100 to 500 company stocks and we believe this to be a sufficient representation of the volatility and performance of these sectors. These data were obtained from Yahoo! Finance. The sectors, abbreviations, and Vanguard symbols are shown in Table 1.

Because we do not have access to daily volatility data, we choose to use three commonly accepted range-based volatility estimators to calculate daily volatility of the ETFs: the Parkinson, the Garman-Klass, and the Rogers-Satchell estimators. Each of these estimators is built on a variety of statistical assumptions about the true underlying volatility, which is explored in [SZ06]. They are calculated as follows:

The Parkinson method takes into account only the daily high and low prices. It is given by the following formula:

| **Sector** | **Abbreviation** | **Symbol** |
|---|---|---|
| Healthcare | Health. | VHT |
| Technology | Tech. | VGT |
| Real Estate | R.E. | VNQ |
| Financials | Fins. | VFH |
| Consumer Discretionary | C.D. | VCR |
| Materials | Mats. | VAW |
| Energy | Energy | VDE |
| Industrials | Inds. | VIS |
| Utilities | Utils. | VPU |
| Communications | Comms. | VOX |
| Consumer Staples | C.S. | VDC |

Table 1: Sectors with corresponding abbreviations and index symbols.

$$V^P = \sqrt{\frac{1}{4\ln(2)} \sum_{i=1}^{n} \frac{1}{n} \ln\left(\frac{H_i}{L_i}\right)^2} \tag{1}$$

where $n$ is the number of trading days in the sample period, and $H_i$ and $L_i$ are the high and low prices of the stock on a given day $i$.

The Garman-Klass method builds on Parkinson's foundation by incorporating the closing and open prices of stock. It is given by the following formula:

$$V^{GK} = \sqrt{\sum_{i=1}^{n} \frac{1}{2n} \ln\left(\frac{H_i}{Oi}\right)^2 - \frac{(2\ln(2)-1)}{n} \ln\left(\frac{C_i}{Oi}\right)^2} \tag{2}$$

where $O_i$ and $C_i$ are the open and close prices of the stock on a given day $i$, and the other variables are the same as above.

Finally, the Rogers-Satchell estimator also uses all four values to estimate volatility. It is given by the following formula [Res19]:

$$V^{RS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{H_i}{Oi}\right) \ln\left(\frac{H_i}{Ci}\right) + \ln\left(\frac{L_i}{Oi}\right) \ln\left(\frac{L_i}{Ci}\right)} \tag{3}$$

Since our sentiment data includes weekends and other holidays while our volatility data does not, we fill in the latter values to aid our analysis. To do so, we adapt an approximation strategy discussed in [AM12]. If the volatility on a certain day is missing, we take the two nearest existing values and take an arithmetic mean. For example, when the volatility value on Friday is $x$ and the volatility value on Monday is $y$, the Saturday value will be given by $z = \frac{x+y}{2}$, and the Sunday value will be $\frac{z+y}{2}$. We apply this to all of our volatility data. In total, our processed sentiment and volatility data spanned 521 days, from January 3, 2022, to June 8, 2023.

## 3.2 Preliminary Analysis

### 3.2.1 Pearson Correlation

To quantitatively determine the relationship between sentiment and various values across the 11 market sectors (closing price, percentage change, and volatility calculated using the three different estimators), we calculate the Pearson correlation coefficient. It is important to note that the Pearson test is not designed to be used for time series data, and is simply a measure of how linearly related two datasets are. Thus, we will use these results only to see whether a relationship exists, and then we will futher quantify it in the time domain.

It is calculated as follows. It returns a number between -1 and 1; if it is positive, it implies a positive correlation, and vice versa. Values closer to the ends imply stronger correlations. Given two time series $X_t$ and $Y_t$, $n$ total observations, the correlation coefficient $\rho$ is calculated as follows[SC+02]:

$$\rho(X,Y) = \frac{\sum_{t=1}^{n} \left(X_t - \bar{X}\right)\left(Y_t - \bar{Y}\right)}{\sqrt{\sum_{t=1}^{n} \left[\left(X_t - \bar{X}\right)^2 \left(Y_t - \bar{Y}\right)^2\right]}} \tag{4}$$

### 3.2.2 Granger Causation

To determine whether market sentiment and volatility are further linked in the time domain, we perform a Granger causality test. It is a method commonly used in econometrics to analyze unidirectional or bidirectional causation. Granger causality is not true causation and only checks whether one time series is useful in forecasting another. Thus, we refer to significant results as one time series and "Granger-causing" another. We perform the Granger causality test from two different directions to see whether sentiment Granger causes volatility and vice versa, across different sectors. The Granger causality test between two time series is performed as follows [Gra69]. Given two time series $X$ and $Y$, we say that $X$ Granger-causes $Y$ if predictions of $Y$ based on its own past values *and* past values of $X$ are significantly better than predictions of $Y$ solely based on its own past values.

To do so, we first perform an autoregression with the values of $Y$ only:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_m y_{t-m} + err_t \tag{5}$$

Next, we include the values of $X$, time-lagged by a certain period:

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + ... + b_m y_{t-m} + c_p x_{t-p} + ... + c_q x_{t-q} + err_t \tag{6}$$

We then perform a t-test on each of the individual values of $X$, keeping the ones that are significant, with the condition that they add explanatory power to the regression (under an F-test). Then, we fail to reject the null hypothesis (that $X$ adds no explanatory power) if any new values of $X$ are included in the regression.

According to [Owe19], a news cycle in 2018 would typically lasts a week before losing relevance in the public mind. Thus, we decide to run the Granger causality test on our data with time lags of up to 14 days, including an additional seven days to observe potential trends and patterns.

The Granger causality test also requires the time series to be stationary, meaning that they must have a constant joint probability distribution over time, such that statistics like the mean and variance are constant over time [HDC05]. To test for stationarity, we use the Augmented Dickey-Fuller (ADF) test, outlined in [DF79]. We find that all of our time series were stationary except for closing price. To keep using the closing price in our Granger causality test, we differentiate the data by subtracting the current value of the time series from the previous one [HRHK19]. In doing so, we obtain a differenced closing price which is just the change in closing price from the previous day. It is referred to in our table as $DC$. The differenced closing price data is stationary.

## 3.3 Model Building

### 3.3.1 Regression Models

We use the Python sklearn platform to implement our regression-based models, and we choose three models - linear regression, decision tree, and XGBoost - to try and predict volatility. We split our data into testing and training sets as follows for cross validation. With $k$ data splits, the model will be trained on the first $k$ sets and tested on the $k+1$th set, ensuring that no future values are used to forecast past values and to preserve the time dependency between data points [PVG$^+$11].

For model training, we include all the technical indicators from the previous day (open, close, high, low) as well as sentiment values time-lagged by the significant intervals we find in our Granger causality test, in order to predict the Garman-Klass volatility of the current day. For example, to predict future consumer discretionary Garman-Klass volatility, we include the sentiment, time-lagged by 1, 10, and 11 days. We then compare the accuracy of our models with and without these additional sentiment features.

### 3.3.2 Long Short-Term Memory

For a more complex model, we also implement a long short-term memory (LSTM) model built with Keras, a deep learning platform of TensorFlow, to predict the Garman-Klass volatility of the various sectors. LSTMs are a type of deep learning based recurrent neural network that is optimized for time series prediction as it can selectively remember and forget data from long periods of time as data passes through the memory cells and three gates: forget, input and output gates [HZL$^+$19]. An LSTM model also eliminates the problem of vanishing and exploding gradients for long term time series data [JWW21].

Our LSTM model is defined with a sentiment layer that processes the sentiment input to predict volatility. Because the LSTM model allows for more flexible treatment of input data, we use time steps of five for the input, similar to [YL18]. This means that the technical indicators (same as above) and sentiment from the previous five days are used to forecast the volatility on the sixth day. We set the number of epochs - iterations in one cycle of training a set of data - to 100, and train the model both with and without sentiment data.

#### 3.3.3 Evaluation

We calculate the mean absolute percentage error (MAPE) according to [GÖBD16] and [JWW21] for each of the four machine learning models. MAPE is a statistical method widely used for measuring forecasting prediction accuracy. It is given by the following formula:

$$MAPE = \frac{1}{n}\Sigma_{t=1}^{n}\frac{|y_{true} - y_{pred}|}{y_{true}} \tag{7}$$

# 4 Results and Discussion

## 4.1 Pearson Correlation

The relationship between market sentiment and percentage change of stock price (in a few sectors such as technology) shows a weak positive correlation, a result that resembles relevant literature [ACG22]. A weak negative correlation was observed across sentiment and market volatility, where positive sentiment correlates with lower volatility, resembling the results found in [ZGWY21]. From these results, we can infer that a stable market and better performance occur at the same time as optimistic public sentiment, in most sectors.

It is important to consider that the Pearson correlation test only gives information about same-day correlation; thus, we cannot infer from it whether one of the variables causes the other.

## 4.2 Granger Causality

Similar to relevant literature, we obtained results that indicate relevant causality between stock and sentiment [ACG22] in both directions–whether sentiment Granger causes ETFs volatility and volatility Granger causes sentiment– depending on the sector.

Our results also show that the sectors with volatility prediction that is best improved by sentiment data inclusion are technology and consumer discretionary, suggesting that the users we chose and the users who follow them are likely more involved in technology stocks than others. Consumer discretionary products are highly sensitive to the whims and moods of consumers, since they are by nature not necessities. On the other hand, we find that the volatility of sectors such as real estate, consumer staples, financials, and energy are not at all linked closely with historical sentiment at any time lag, which suggests that changes in these sectors are not tied closely to sentiment.

| | **Close** | **Δ%** | **PV** | **GKV** | **RSV** |
|---|---|---|---|---|---|
| Health. | $0.286^{***}$ | 0.058 | $-0.121^{**}$ | $-0.155^{***}$ | $-0.156^{***}$ |
| Tech. | $0.154^{***}$ | $0.113^{**}$ | $-0.078$ | $-0.103^{*}$ | $-0.109^{*}$ |
| R.E. | $0.104^{*}$ | 0.055 | $-0.107^{*}$ | $-0.109^{*}$ | $-0.096^{*}$ |
| Fins. | $0.265^{***}$ | $0.125^{**}$ | $-0.109^{*}$ | $-0.143^{**}$ | $-0.163^{***}$ |
| C.D. | $0.146^{***}$ | $0.104^{*}$ | $-0.092^{*}$ | $-0.125^{**}$ | $-0.128^{**}$ |
| Mats. | $0.245^{***}$ | $0.117^{**}$ | $-0.120^{**}$ | $-0.130^{**}$ | $-0.117^{**}$ |
| Energy | 0.043 | 0.084 | $-0.134^{**}$ | $-0.148^{***}$ | $-0.143^{**}$ |
| Inds. | $0.272^{***}$ | $0.104^{*}$ | $-0.126^{**}$ | $-0.143^{**}$ | $-0.133^{**}$ |
| Utils. | 0.004 | 0.033 | $-0.116^{**}$ | $-0.094^{*}$ | $-0.060$ |
| Comms. | $0.131^{**}$ | $0.103^{*}$ | $-0.122^{**}$ | $-0.153^{***}$ | $-0.160^{***}$ |
| C.S. | $0.204^{***}$ | $0.096^{*}$ | $-0.125^{**}$ | $-0.090^{*}$ | $-0.040$ |

Table 2: Pearson correlation test results.

**Δ%** : Percentage change in price from previous day
**PV**: Parkinson's volatility
**GKV**: Garman-Klass volatility
**RSV**: Rogers-Satchell volatility
$^{*}p < 0.05$, $^{**}p < 0.01$,$^{***}p < 0.001$

We visualize one relationship between sentiment and Parkinson's volatility of technology ETFs in Figure 3. We applied a Gaussian filter to our sentiment and volatility time series to filter out distracting noise for presentation, with the true data still displayed beneath. The graph shows the relationship between the two variables over a shorter time span of around 200 days, and there are rough trends (reflected in our Granger causality data) that can be seen. For example, as the sentiment (red line) rises and falls, there is usually an opposite motion in the volatility (purple line) shortly after. Similar to our results from Pearson's correlation value, this suggests that negative sentiment corresponds to a more volatile market.

### 4.3 Modeling

We ran the models both with and without sentiment data input. Table 4 shows our results. The LinR, DT and XGB models have a similar MAPE with and without sentiment data input. On the other hand, the LSTM model improves significantly upon inclusion of the sentiment data. The $p$-value tests for whether there is a statistically significant difference between the MAPEs of the models with and without sentiment.

Figure 4 compares the predictions made on the Garman-Klass volatility of the technology sector made by the LinR and LSTM, with sentiment data included.

| | **DC** | **Δ%** | **PV** | **GKV** | **RSV** |
|---|---|---|---|---|---|
| Health. | | $\rightarrow 1, 2$ | | | $\rightarrow [2, 6], [8, 12]$ |
| | | | | | $\leftarrow 3$ |
| Tech. | $\rightarrow 2, 9$ | $\rightarrow [1, 14]$ | $\rightarrow [2, 14]$ | $\rightarrow [3, 14]$ | $\rightarrow [4, 14]$ |
| | | $\leftarrow 3, 4, [7, 9]$ | $\leftarrow 5$ | $\leftarrow [8, 11]$ | $\leftarrow [8, 11]$ |
| R.E. | | | | | |
| | | $\leftarrow [3, 7]$ | | | |
| Fins. | | $\rightarrow 2$ | $\rightarrow [1, 3]$ | | |
| | | | | $\leftarrow 1$ | $\leftarrow 1, 2$ |
| C.D. | $\rightarrow 2, 9$ | $\rightarrow [1, 5], [9, 12]$ | $\rightarrow [1, 14]$ | $\rightarrow 1, 10, 11$ | |
| | | | | $\leftarrow [1, 4], 8, 10$ | $\leftarrow [1, 5], [7, 13]$ |
| Mats. | | | $\rightarrow [1, 14]$ | $\rightarrow [1, 6], [10, 14]$ | $\rightarrow 2$ |
| | $\leftarrow [4, 8]$ | $\leftarrow [3, 10]$ | | $\leftarrow [1, 3]$ | $\leftarrow 1, 2$ |
| Energy | | | | | |
| | | $\leftarrow 1$ | $\leftarrow [1, 3]$ | $\leftarrow [1, 3]$ | $\leftarrow [1, 3]$ |
| Inds. | | $\rightarrow 3$ | $\rightarrow [1, 3], [10, 14]$ | $\rightarrow 1, 14$ | |
| | | | $\leftarrow [1, 3]$ | $\leftarrow [1, 3]$ | |
| Utils. | | | $\rightarrow [3, 5]$ | $\rightarrow 3$ | |
| | $\leftarrow 3, 4, 6$ | $\leftarrow 3$ | | $\leftarrow 1, 3$ | |
| Comms. | | $\rightarrow 2, 3$ | $\rightarrow 5$ | $\rightarrow 3$ | |
| | | $\leftarrow 3, 11$ | | | $\leftarrow 1, 2, [8, 10]$ |
| C.S. | | | | | |
| | $\leftarrow [1, 8]$ | $\leftarrow 1, [3, 14]$ | | | |

Table 3: Granger causality analysis results.

**DC**: Differenced closing price
**Δ%**: Percentage change in price from previous day
**PV**: Parkinson's volatility
**GKV**: Garman-Klass volatility
**RSV**: Rogers-Satchell volatility
$\rightarrow$ indicates sentiment Granger-causing the column header and vice versa, and numbers indicate the time lags at which the causality is significant ($p < 0.05$). For example, $\rightarrow [2, 14]$ for technology, under PV, indicates that sentiment Granger causes Parkinson's volatility for technology stocks for a time lag from 2 to 14 days, inclusive.

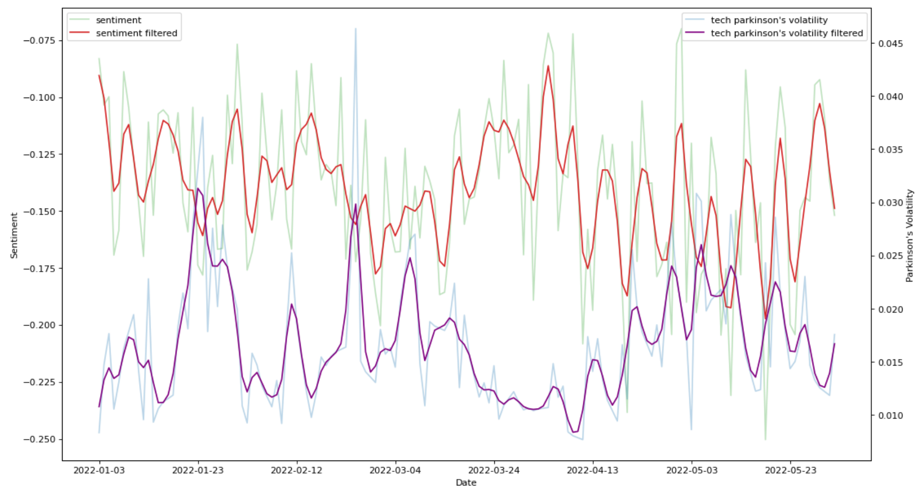



Figure 3: A section of sentiment and Parkinson's volatility in the technology sector.

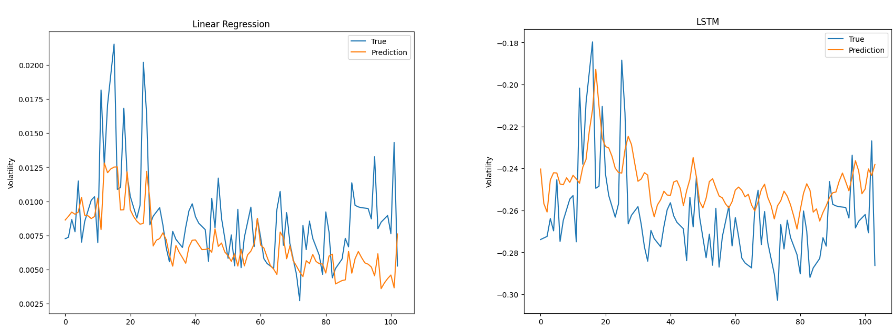
# 5 Conclusion and Future Work

In this paper, we conduct two statistical tests, the Pearson correlation test and the Granger causality test, to determine the relationship between sentiment and sector-wise volatility. We find for most sectors, there is a negative correlation between market sentiment and same-day volatility. From our Granger causality test results, we find that the addition of sentiment data improves the regression performance in certain sectors, such as technology, consumer discretionary, and materials, while not for other sectors, such as real estate, energy, and consumer staples. We hypothesize that the latter sectors are less affected by Twitter sentiment change, as they are related to goods and services that are necessities. Alternatively, these findings may suggest that the Twitter users we selected and those who follow them are more involved in stocks in the former sector, such as technology. It has been well documented that influential figures on Twitter such as Elon Musk can have a large impact on stock performance just from their tweets. We suggest that investors can use Twitter data as an appropriate proxy or indicator of future volatility, especially in sectors like technology and consumer discretionary, and thus adjust their expectations of risk accordingly.

Informed by these statistical conclusions, we then propose four different machine learning models to forecast volatility. We find that the LSTM model is best able to make use of the additional sentiment data, and is the only model which improves upon the inclusion of this data. The average MAPE of the LSTM model with sentiment is significantly lower than all other models, suggesting that the more abstract temporal structure of the LSTM is far better suited for analyzing and extracting useful information from the sentiment data. There are many hyperparameters for the LSTM model such as time steps, layer

| | LSTM | | LinR | | DT | | XGB | |
|---|---|---|---|---|---|---|---|---|
| | s. | n.s. | s. | n.s. | s. | n.s. | s. | n.s. |
| Health. | 0.073 | 0.346 | 0.263 | 0.262 | 0.374 | 0.335 | 0.317 | 0.343 |
| Tech. | 0.072 | 0.446 | 0.294 | 0.294 | 0.967 | 0.795 | 0.650 | 0.513 |
| R.E. | 0.081 | 0.0.225 | 0.095 | 0.096 | 0.244 | 0.244 | 0.175 | 0.182 |
| Fins. | 0.159 | 0.283 | 0.161 | 0.159 | 0.299 | 0.341 | 0.260 | 0.291 |
| C.D. | 0.099 | 0.41 | 0.241 | 0.242 | 0.656 | 0.902 | 0.759 | 0.765 |
| Mats. | 0.077 | 0.331 | 0.159 | 0.148 | 0.371 | 0.360 | 0.268 | 0.202 |
| Energy | 0.072 | 0.236 | 0.243 | 0.246 | 0.311 | 0.301 | 0.369 | 0.354 |
| Inds. | 0.08 | 0.297 | 0.141 | 0.145 | 0.316 | 0.411 | 0.216 | 0.280 |
| Utils. | 0.110 | 0.250 | 0.222 | 0.224 | 0.375 | 0.333 | 0.280 | 0.214 |
| Comms. | 0.086 | 0.362 | 0.201 | 0.200 | 0.347 | 0.321 | 0.373 | 0.387 |
| C.S. | 0.083 | 0.365 | 0.198 | 0.197 | 0.547 | 0.434 | 0.267 | 0.302 |
| Mean | 0.090 | 0.323 | 0.202 | 0.201 | 0.437 | 0.434 | 0.57 | 0.348 |
| $p$ | 0.000 | | 0.687 | | 0.934 | | 0.618 | |

Table 4: MAPE of various machine learning models.




(a) LinR

(b) LSTM

Figure 4: Comparison of models for prediction of Garman-Klass volatility in the tech sector.

size, learning rate, epoch size, and more, that could be further optimized in order to improve model performance. Our average MAPE value for the LSTM-with-sentiment model of 0.090 is close to, but not as low as the other values in previous literature, suggesting that further optimization and training can still be performed, considering that we had only 521 days worth of data, which is very low for machine learning and deep learning applications. We are unfortunately restricted by the newly established Twitter rate limits. We could also include more input datasets for our models, including additional technical market indicators. Overall, we believe that this method can also be applied to the prediction of other market statistics, such as returns.

There are limitations to this study. The volatility predictions are all made based on the performance of the aforementioned Vanguard ETFs. This means that our findings only reflect the aggregate performance of the hundreds of stocks included in each ETF and not the individual stocks, which investors must consider. It may be useful to analyze the relationship between sentiment and individual stocks. With regard to our data, especially in calculations of volatility and sentiment, we also use many assumptions. For the calculation of volatility, we use estimators which are commonly used for an extended period of time instead of a single day. A better way to calculate daily volatility would be using intraday historical data instead of only the opening, closing, high and low prices. However, this data is not easily accessible due to difficulty retrieving them and its large data size. To compare the daily Twitter sentiment with the daily volatility of the trading days, we also assume the weekends' volatility to be the average of the closing price and opening price of the previous and following days respectively. In addition to data from Twitter, other social media platforms, such as Instagram and Reddit, may also contain useful sentiment information.

# References


[ACG22] Faten Alzazah, Xiaochun Cheng, and Xiaohong Gao. Predict market movements based on the sentiment of financial video news sites. In *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*, pages 103–110. IEEE, 2022.

[AM12] Arpit Goel Anshul Mittal. Stock prediction using twitter sentiment analysis. *Standford University, CS229*, 2012.

[BMG20] Seungho Baek, Sunil K. Mohanty, and Mina Glambosky. Covid-19 and stock market volatility: An industry level analysis. *Finance Research Letters*, 37:101748, 2020.

[BNA+21] Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Somayeh Asadi, and U Rajendra Acharrya. A novel fusion-based deep learning model for sentiment analysis of covid-19 tweets. *Knowledge-Based Systems*, 228:107242, 2021.

[DF79] D. Dickey and Wayne Fuller. Distribution of the estimators for autoregressive time series with a unit root. *JASA. Journal of the American Statistical Association*, 74, 06 1979.

[DGPP22] Justina Deveikyte, Helyette Geman, Carlo Piccari, and Alessandro Provetti. A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*, 5, 2022.

[dOCCdS21] Arthur Emanuel de Oliveira Carosia, Guilherme Palermo Coelho, and Ana Estela Antunes da Silva. Investment strategies applied to the brazilian stock market: a methodology based on sentiment analysis with deep learning. *Expert Systems with Applications*, 184:115470, 2021.

[GÖBD16] Mustafa Göçken, Mehmet Özçalıcı, Aslı Boru, and Ayşe Tuğba Dosdoğru. Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Systems with Applications*, 44:320–331, 2016.

[Gra69] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[GZ23] Yang Gao and Chengjie Zhao. Investor sentiment and stock price jumps: A network analysis based on china's carbon–neutral sectors. *The North American Journal of Economics and Finance*, 68:101954, 2023.

[HDC05] Jim Granato Harold D. Clarke. Time series analysis in political science. *Encyclopedia of Social Measurement*, pages 829–837, 2005.

[HRHK19] Zakir Hossain, Atikur Rahman, Md Hossain, and Jamil Karami. Over-differencing and forecasting with non-stationary time series data. *Dhaka University Journal of Science*, 67:21–26, 01 2019.

[HZL+19] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, 57(6):114–119, 2019.

[JWW21] Nan Jing, Zhao Wu, and Hefei Wang. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, 178:115019, 2021.

[LXC+14] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69:14–23, 2014.

[Owe19] Laura Hazard Owen. A typical big news story in 2018 lasted about 7 days (until we moved on to the next crisis). *Nieman Lab*, 25, 2019.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[RAC+15] Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441, 2015.

[Res19] Hudson Thames Quantitative Research. Volatility estimators, 2019.

[RPS+23] Claire E Robertson, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J Van Bavel, and Stefan Feuerriegel. Negativity drives online news consumption. *Nature human behaviour*, 7(5):812–822, 2023.

[SC+02] Thomas Douglas Victor Swinscow, Michael J Campbell, et al. *Statistics at square one*. Bmj London, 2002.

[SZ06] Jinghong Shu and Jin E Zhang. Testing range estimators of historical volatility. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 26(3):297–313, 2006.

[YL18] ShuiLing Yu and Zhe Li. Forecasting stock price index volatility with lstm deep neural network. In *Recent Developments in Data Science and Business Analytics: Proceedings of the International Conference on Data Science and Business Analytics (ICDSBA-2017)*, pages 265–272. Springer, 2018.

[YUH20] Yi Yang, Mark Christopher Siy UY, and Allen Huang. Finbert: A pretrained language model for financial communications, 2020.

[ZGWY21] Weiguo Zhang, Xue Gong, Chao Wang, and Xin Ye. Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting*, 40(8):1479–1500, 2021.