|  |
| --- |
| SYRACUSE UNIVERSITY |
| Financial Analysts on Twitter & Market Sentiment |
| IST 736 – Text Mining |
|  |
| **By Jeff Levesque, Kent Appiah & Seán Hayes** |
| **6/12/2019** |

|  |
| --- |
|  |

Table of Contents

[Problem Definition 2](#_Toc9595867)

[Data Descriptions, Cleaning and Preparation 3](#_Toc9595868)

[Data Set Descriptions 3](#_Toc9595869)

[Data Collection and Cleaning 4](#_Toc9595870)

[Modeling 6](#_Toc9595871)

[VADER Sentiment Analysis 6](#_Toc9595872)

[Correlation Analysis 7](#_Toc9595873)

[Result Interpretation 7](#_Toc9595874)

[Correlation Analysis 7](#_Toc9595875)

[Conclusion 9](#_Toc9595876)

## Problem Definition

Do the opinions of Financial Analysts on Twitter impact market sentiment and volatility? There are many financial analysts on twitter offering investment advice to millions of people on a daily basis. These analysts also have other mediums outside of Twitter in which the can communicate with investors like TV shows, blogs and regular newspaper columns but Twitter can act as a gauge for the overall sentiment towards the market that these analysts are portraying on a daily basis through all means. Taking the most famous financial analyst Jim Cramer for example, he has over 1.1 million followers on Twitter[[1]](#footnote-1) and he is the host of the popular finance television show Mad Money where he gives investment advice to an audience that can reach 328,000 viewers[[2]](#footnote-2) on daily basis, is the sentiment he is portraying to investors, impacting their investment decisions and in turn, impacting market volatility?

Market volatility is defined as “Volatility is a statistical measure of the dispersion of returns for a given security or market index. In most cases, the higher the volatility, the riskier the security.”[[3]](#footnote-3) Market volatility can be tracked by using the CBOE Volatility Index (VIX) aka the Fear Index. The VIX is a market index that represents the market's expectation of 30-day forward-looking volatility. It is derived from the price inputs of the S&P 500 index options; it provides a measure of market risk and investors' sentiments[[4]](#footnote-4). The VIX is considered a reflection of investor sentiment; the higher the VIX is up, the higher the levels of investor anxiety and market volatility.

If there is a correlation between the sentiment being portrayed by these analysts towards the market and the level of market volatility for that day, then this information can be a valuable tool for investors to incorporate into their investment strategy. Having advanced knowledge of when a market is about to increase in volatility would be invaluable to investors allowing them to adjust their portfolios preempting the market risk. Tweets have been known to impact the market before, on May 5th 2019, President Trump tweeted negatively about a Trade War with China and the VIX rose by as much as 46.1% intra-day the next market day, while the Dow Jones Industrial Average plunged by as much as 471 points[[5]](#footnote-5).

## Efficient Market Hypothesis (EMH)

Investors believe that stock prices reflect the all the information about a company and that there’s no arbitrage opportunity for either side of market participants – buyer and seller. The theory specifies that at any given point in time the market is operating at a full potential and any shifts in this underlying believe would change the utility of investors. If this theory is true then in markets, a particular investor does not possess dominance over other investors.

Arbitrage opportunity, according to Investopedia, is simply buying a security in one market and selling that same security in another market for profit. If and when this happens, one investor is better-off at the expense of another investor. Here’s why. Suppose there are two investors in the market for a trade. The seller investor is willing to sell 100 shares of Apple stock, which is currently priced at $175.00. On the other hand, the buyer investor is willing to buy these shares at the current price. The seller, speculatively, believes that Apple stock will fall in price and the buyer, with a greater degree of uncertainty, believes Apple stock will rise in the future. Under the market efficiency hypothesis, when this trade happens, these two investors are trading on speculations and thus regardless of the direction of Apple stock, none of the investors are better or worse-off than the other.

Suppose further that the buyer investor knows with 100% probability that Apple stock will increase by $4. In this case the buyer will borrow money from the bank at, infinitesimally, low interest rate to buy the stock and make a profit without bearing any investment risk. Few things that would’ve happened in this second scenario:

* Arbitrage opportunity would’ve occurred
* The buyer would’ve taken advantage of the arbitrage to increase their utility
* The seller loses on this trade because of the imbalanced information in the market.

Key takeaways from the above scenario is that in an event of arbitrage, stock prices are not in an equilibrium and therefore this inefficiency causes an imbalanced trading. Although, there’s a broad believe that markets are efficient and that any arbitrage opportunity, when they exist, are quickly repriced and eroded, there’s also a common believe that markets can be timed. With the right information, an investor can trade on public information before that information is repriced into markets. Keep in mind that public information are information available to all investors such that one investor is not advantageous over the other.

## Data Descriptions, Cleaning and Preparation

### Data Set Descriptions

#### CBOE Volatility Index (VIX) aka the Fear Index

* URL: <http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data>
* The daily high, low and closing prices are available historically back to 2004

#### Financial Analysts’ Tweets

* URL: <http://www.twitter.com>
* Financial Analysts chosen:
  + @jimcramer – Jim Cramer
    - Host of Mad Money on CNBC
    - 1.1 million followers with over 95,000 tweets
  + @ReformedBroker – Josh Brown
    - Analyst on CNBC’s Halftime Report and CEO of Ritholtz Wealth Management
    - 1 million followers with over 88,000 tweets
  + @TheStalwart – Joe Weisenthal
    - Host of What Did You Miss? on Bloomberg TV
    - 170,000 followers with over 320,000 tweets
  + @LizAnnSonders – Liz Ann Sonders
    - Chief Investment Strategist for Charles Schwab
    - 70,000 followers with over 8,000 tweets
  + @SJosephBurns – Steve Burns
    - Founder of NewTraderU.com
    - 160,000 followers with over 100,000 tweets

### Data Collection and Cleaning

#### VIX data

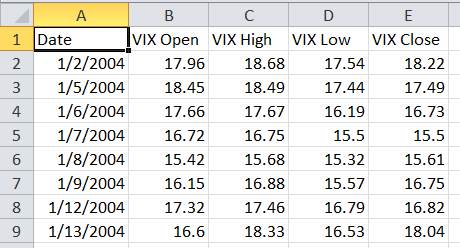


Figure - Before: VIX data in Excel

The VIX file contained daily high, low and closing prices available historically back to 2004 for the VIX index, only the closing price was needed. A Daily Returns column was needed to be created for the analysis by finding the difference between the current day’s closing value and the previous day’s closing value. The first row was dropped from the data as the return could not be calculated due to the previous close value was not available. The Date column was converted into a Python date data type so it could be joined with other datasets.

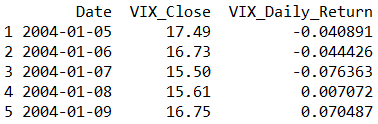


Figure - After: VIX data in Python

#### Twitter Data

Talk about API data collection…

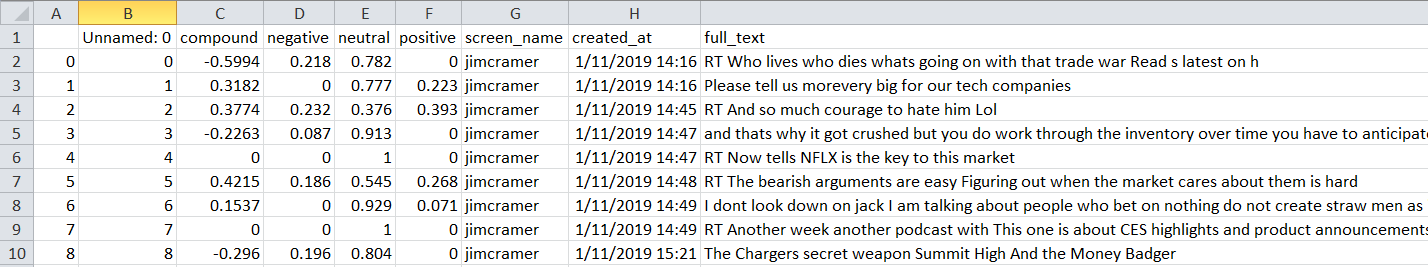


Figure - Before: Twitter data in Excel

The tweets from each analyst were outputted into separate excel files with their Vader sentiment scores. When loading the data into Python, only the compound, screen\_name, and created\_at columns were required. The compound column was converted to a float data type and the created\_at column was converted into a Python date data type so it could be joined with other datasets. Then for each day the average compound score was calculated for each analyst.

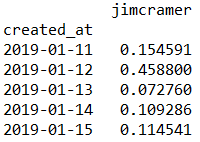


Figure - Twitter data in Python

#### Combining VIX and Twitter Data

All the twitter data for each analyst was combined with the VIX data joining on the created\_at and Date columns so analysis could be performed on the data.

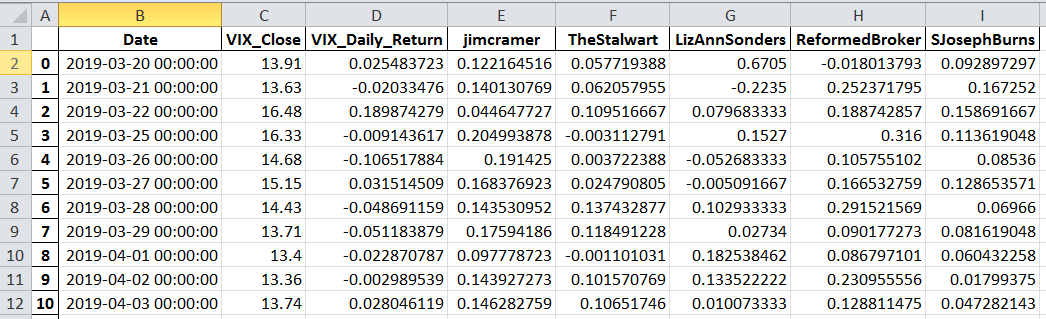


Figure - Combined Data sets (in Excel for easier reading)

## Modeling

### VADER Sentiment Analysis

The VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion[[6]](#footnote-6). The VADER model was applied to each tweet outputting a compound score.

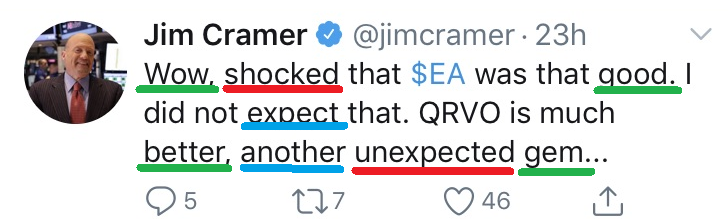


Figure - VADER Sentiment Analysis on a Tweet

The compound score is metric that calculates the sum of all the lexicon ratings which have been normalized between -1 meaning the most extreme negative and +1 meaning the most extreme positive. Each word in a tweet is scored as negative, neutral or positive where the compound score is calculated off those values. A daily average compound score was calculated for each analyst from all their tweets in that day.

### Correlation Analysis

Correlation analysis is a method of statistical evaluation used to study the strength of a relationship between two (or more), numerically measured, continuous variables[[7]](#footnote-7). Correlation analysis was performed comparing the VIX Daily Return value against all the daily average compound scores of each analyst.

## Result Interpretation

### Correlation Analysis

The correlation analysis between the VIX Daily Return value and all the daily average compound scores of each analyst uncovered very little correlation between the market sentiment portrayed by the analysts and VIX.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VIX Daily Return** | **jimcramer** | **TheStalwart** | **LizAnnSonders** | **ReformedBroker** | **SJosephBurns** |
| **VIX Daily Return** | 1 | -0.19352293 | 0.063617964 | 0.21215351 | 0.165338168 | 0.081536073 |
| **jimcramer** | -0.19352293 | 1 | -0.176836905 | -0.208063147 | -0.126110348 | -0.038933127 |
| **TheStalwart** | 0.06361796 | -0.176836905 | 1 | -0.070329108 | 0.087018995 | -0.045209402 |
| **LizAnnSonders** | 0.21215351 | -0.208063147 | -0.070329108 | 1 | -0.231889267 | -0.127426701 |
| **ReformedBroker** | 0.16533817 | -0.126110348 | 0.087018995 | -0.231889267 | 1 | -0.242437727 |
| **SJosephBurns** | 0.08153607 | -0.038933127 | -0.045209402 | -0.127426701 | -0.242437727 | 1 |

Figure - Correlation Analysis Output

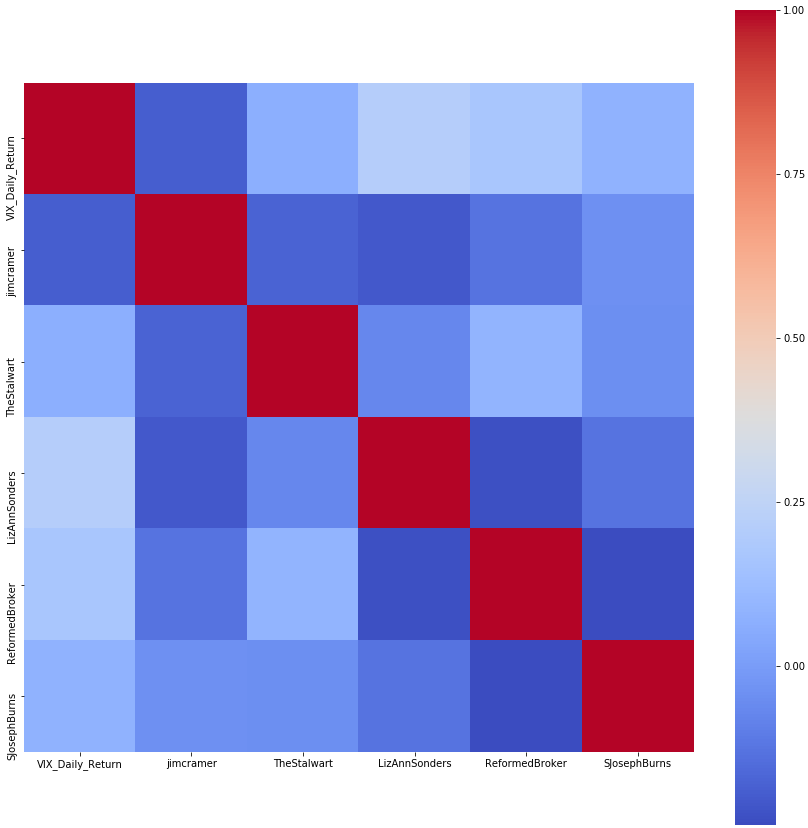


Figure - Correlation Analysis Heat map

There is no correlation lower than -0.24 or greater +0.21 which shows that no financial analyst’s Twitter sentiment towards the market has any significant influence on the daily returns of the VIX.

## Conclusion

Unfortunately, there is no significant correlation between the sentiment being portrayed by these financial analysts towards the market and the level of market volatility for that day. The tweets that the financial analysts post on a daily basis giving investment advice are not significant enough or don’t reach a wide enough audience to impact the daily returns of the VIX index.



Figure - VIX 1 Year Timeline[[8]](#footnote-8)

The research was conducted on Tweets between March and May 2019. This period for the VIX was relatively stable. It would be interesting to redo this analysis during a more volatile period for the VIX like between October 2018 and December 2018 where there was much more volatility and see if there is a more significant correlation between the tweets and the VIX.

1. <https://twitter.com/jimcramer> [↑](#footnote-ref-1)
2. <http://tvbythenumbers.com/2009/03/10/cable-news-ratings-for-monday-march-9/14279> [↑](#footnote-ref-2)
3. <https://www.investopedia.com/terms/v/volatility.asp> [↑](#footnote-ref-3)
4. <https://www.investopedia.com/terms/v/vix.asp> [↑](#footnote-ref-4)
5. <https://www.bloomberg.com/news/articles/2019-05-06/with-two-tweets-trump-shatters-historic-calm-in-global-markets> [↑](#footnote-ref-5)
6. <http://datameetsmedia.com/vader-sentiment-analysis-explained/> [↑](#footnote-ref-6)
7. <https://www.djsresearch.co.uk/glossary/item/correlation-analysis-market-research> [↑](#footnote-ref-7)
8. <https://finance.yahoo.com/chart/%5EVIX/> [↑](#footnote-ref-8)