Data and Program Analytics

Market Volatility Analysis with Traditional and Non-Traditional Indicators

03/20/2019

Prepared by: Downey, Caleb; Lechner, Max; Merkler, Jamie; Nguyen, Loi; Vosoughian, Azadeh

University of California, Irvine

Introduction

This project evaluated how the bond inversion curve (STPP used as a proxy), market volatility (VIX used as a proxy), and stock related Twitter data (from key investing personalities) can help predict market volatility (S&P 500 (aka SPY) used as proxies). The VIX indicator uses short-term information to measure day to day fear on the market. Conversely, the STPP measures the long-term relationship between bond yields and is the most accurate indicator for predicting recessions. The bond inversion curve is frequently cited by economists and business planners who drive both public and private policies off of these implications. Creating a more comprehensive volatility index could improve market understanding for anyone from an investor to strategic planners.

We used information obtained from the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) also known as the "fear index" to evaluate market volatility at a given time. The VIX was created to represent the market's expectation of 30-day forward-looking volatility. It is generated from the price inputs of S&P 500 index options. Despite its name signaling a bearish for stocks, VIX is a measure of market perceived volatility, whether the volatility is positive and/or negative. High VIX reading means that investors perceive a risk for a sharp movement in the market, both positive or negative. On the contrary, low VIX is an indication of slow movements in the market, whether that movement is positive or negative. Below is the formula provided by CBOE's Index Rules & Methodology:

The VIX Index Formula

The generalized formula used in the VIX Index calculation[1] is:

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2$$

WHERE...

σ	$VIX/100 \Rightarrow VIX = \sigma \times 100$
T	Time to expiration
F	Forward index level derived from index option prices
K_0	First strike below the forward index level, F
K_i	Strike price of i^{th} out-of-the-money option; a call if $K_i > K_0$ and a put if $K_i < K_0$; both put and call if $K_i = K_0$.

ΔK_i	Interval between strike prices – half the difference between the strike on either side of K_i: $$\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$$ (*Note*: ΔK for the lowest strike is the difference between the lowest strike and the next higher strike. Likewise, ΔK for the highest strike is the difference between the highest strike and the next lower strike.)
R	Risk-free interest rates to expiration
$Q(K_i)$	The average of the bid quote and ask quote for each option with strike K_i.

The next variable we included in our analysis is 10-year & 2-year note yields. Traditionally when 10-year & 2-year bond yield curves invert, it has been a signal of an impending recession. Inversion means that the longer-term debt (10-year) carries a lower yield than shorter-term debt (2-year).

The third variable is tweets from several key investing personalities. This is our curated list of investment related Twitter handles based on our secondary research:
'cnbc', 'benzinga', 'stocktwits', 'breakoutstocks', 'bespokeinvest', 'WSJMarkets', 'stephanie_link', 'nytimesbusiness', 'IBDinvestors', 'TheStalwart', 'MorganStanley', 'MohamedEl-Erian', 'SeekingAlpha', 'SimsOnFinance', 'EPBResearch', 'ReformedBroker', 'StLouisFed', 'SallieKrawcheck', 'benthompson', 'TruthGundlach', 'LizAnnSonders', 'morganhouselFXCM', 'MarkYusko', 'CiovaccoCapital', 'Greenbackd', 'TMFJMo', 'CitronResearch', 'benthompson' TruthGundlach', 'MebFaber', 'AswathDamodaran', 'financialsamura', 'USNewsInvesting', 'howardrgold', 'DCBorthwick', 'DailyFXTeam', 'zerohedge', 'russian_market', 'MarketWatch', 'KeithMcCullough'.

We used python to scrape twitter data and develop an aggregate daily sentiment analysis measure. This measure will take a slice of market related tweets across the Twittersphere and score them using an off the shelf language processing tool known as VADER. VADER is tuned for use with social media and understands how to treat Emojis and commonly used social media acronyms. By employing this tool we hope to capture how key pundits view the market on a particular day and develop a mean score of these tweets to measure how well the market is performing. This analysis is a non-learning algorithm and instead relies on VADER to provide accurate measures. Additionally, it assumes that market pundits on twitter are providing relevant and timely information related to key macro market trends. The final and major assumption that this measure requires, is that bias and agenda can be neutralized by measuring and averaging means across many different pundits. If successful we assume that this measure from tweets will provide a more comprehensive and timely measure of volatility than the VIX or STPP alone.

- **Null Hypothesis:**

  Twitter data will not improve market forecasting when integrated into traditional indexes.

- **Alternative hypothesis:**

  Twitter data will provide better market forecasting better than traditional indexes alone.

To test this we first compared the r-squared values of a linear regression model using only traditional indexes STPP and VIX as our independent variables with SPY as our dependent variable. We then compared that to the same linear regression model with a daily aggregate sentiment analysis measure for twitter handles related to the market. The measures were similar and therefore we failed to reject the null hypothesis, with the adjusted R-squared being lower than STPP & VIX alone.

Methods

To test our hypothesis we developed a database from real time data pulled off of IEX and Twitter to provide measures of percent change from the daily closing price and an average measure for sentiment per each day of tweets.

IEX was accessed using pandas datareader tool to pull data from IEX's API with a manually inputted start and end date (January 1, 2015 to present). This information was then sorted into a pandas dataframe through the standard calling technique and only closing prices were retained. These closing prices were then used to calculate daily movements by employing the pct_change function to calculate closing cost change from day to day. Twitter API was accessed using the tweepy library and manually limited by 40 pages rather than dates. An arbitrary number of page numbers was used because limiting by dates was not effective for setting limits using tweepy's standard functions. Twitter is fairly new and therefore, there is a limited amount of data available which limits the scope of this analysis. The tweets were processed through the VADER sentiment analysis tool to develop a compound score per each tweet. This database was massive and contained over 700,000 tweets. The processing time to retrieve and score the tweets was nearly one hour on a standard Lenovo laptop with a quad core i5 processor. Tweets were then organized by day and using the mean function were aggregated to one market indicator measure.

Another step that we took but eventually abandoned was to filter the tweets to be more S&P 500 specific. This was done by running a manual filter based off of a CSV file that contained nearly 4,000 keywords specific to this dataset. The filtering process added another hour to the processing time and provided the same R-squared value as the non-filtered data. We determined that this step was both ineffective and processing cost prohibitive as it slowed down the processing times considerably. It is left in the jupyter notebook as a vestigial line in the code that is commented out.

The database was then visualized using pandas head and tail to ensure that days matched and values were logical given normal market/ twitter expectations. This data visualization revealed that the daily percent change for the STPP was near zero for most days and led us to believe that it would not be an effective measure for daily market fluctuations. We also visualized the data from the closing prices and twitter score to gain an initial appreciation of our dataset and to understand the dynamics of our data set. For visualization of these database elements we used the matplotlib library.

Our next step was to develop several regression models to measure R-squared values and evaluate our null hypothesis. We developed these models using the statsmodel library and the ols feature to run a simple linear regression. To visualize these linear regressions we used the statsmodels.graphics.regressionplots tool and labelled the y and x axis for ease of understanding.

Our final step in the code was to evaluate the predictive value of our tool using training and testing datasets. These datasets were constructed manually with the intention to let the training dataset be at least twice as large as the testing dataset. A linear regression was run on the larger dataset and then using the predict function from the statsmodel library we developed a measure of the effectiveness of our regression model at prediction. This model revealed a fairly accurate predictive tool with the average deviation in the (+/-) 2% range. Our results for this prediction were then visualized using a scatterplot from the matplotlib library.

Results

The results of multiple regression analyses failed to reject the null hypothesis indicating the addition of Twitter Scores to the model did not increase the variance explained in S&P 500.

We conducted a series of regression analyses to evaluate the relationship between the three independent variables (bond inversion, VIX, and Twitter data) with S&P 500 as the dependent variables.

The results showed that there is a strong relationship between option swaps (VIX) and S&P 500 (R squared= 0.629, p<.000). See Figure 1.

Additionally, the results of the regression analyses show that there is not a significant relationship between bond inversion and S&P 500 (R Squared = 0.01). See Figure 2.

Furthermore, there was no significant relationship between the Twitter sentiments scores and the S&P 500 (R Squared =0.001). See Figure 3.

Lastly, we decided to include Vanguard Total World Stock (VT) as an additional independent variable to examine whether the model can be improved to explain more variance in S&P 500. We ran multiple regression analyses with Bond Inversion, VIX, and Twitter scores, and VT as independent variables and S&P 500 as dependent variables. The results show that the multiple regression model can explain more variation in S&P 500 compared to each individual independent variables alone. The results of the multiple regression analyses were statistically significant (R squared = 0.858, p<.000).

Implications/Conclusion

The major takeaway we have from our tool is that if information is free and easily accessible (tweets, etc) incorporating it will not give users better information than the market readily provides. Our simplistic model using freely available sentiment analysis tools is not robust enough to offer usable information to exploit market fluctuations. The best indicator for short-term variations is the widely used and freely available VIX. This information is a good standalone indicator and can explain approximately 63% of the variation of the S&P 500. The world market offers more information because the S&P500 and the world market are highly correlated and the S&P 500 could be considered a subset of the world market indicator.

Web Scraping Ethical Dilemma

SeekingAlpha is a popular blog for traders and provides market news and analysis for traders. This information is hosted on the SeekingAlpha site and the group initially planned to use a web scraper to pull daily market information for sentiment analysis. However, after reading the user agreement license which explicitly states that web scraping is a violation of policy, we made a decision to not scrape this site due to ethical concerns. As a proxy we included the site's twitter handle as well as those of many of the contributing bloggers. We assessed that our sentiment analysis would not have been robust with the addition of this data, even if it were used as VADER is tuned for use with social media. For the data to be useful, we would instead need to employ a program that uses training data to interpret new sources.

Improvements

A major limitation of our tool was the use of an off the shelf sentiment analysis tool and basic treatment of data to develop the linear regression models that we used to evaluate the effectiveness of these indicators.

Natural language processing is a complicated task and using it to inform a trading algorithm may be helpful if analyzed in greater detail. The current analysis uses the mean of aggregate twitter information in the form of a slice of the overall Twittersphere using a curated list of twitter handles. The theory that this aggregated score developed through the VADER tool will provide a realistic measure of market trends is too simplistic to provide a successful analysis of the market. Our group ran tests to see how filtering would affect the efficacy of this model, and we saw no improvement in how well it informed our models. The filtering mechanism used a list of S&P 500 keywords to filter out non-relevant tweets.

To improve our sentiment analysis model we can easily use market trends with a machine learning tool to develop a better tool to interpret the sentiment of the market. This will allow for a better interpretation of incoming data from twitter while also refining our sentiment algorithm to be geared specifically to market trends.

Additionally, twitter is not a robust platform for market intelligence. While we opted to spare our ethics from the gray area debate of web scraping, a trader with different ethical standards would be able to develop a scraping tool to exploit readily available information on the internet for potential financial gain.

Finally, our data is treated equivalently across all indicators which does not allow the indicator to be used to its full potential. For instance, the STPP indicator is not an effective measure of short-term market trends and using its daily variation is likely ineffective. If instead the model used the STPP to project long-term volatility and offered the model a more long-term outlook, then it would provide a better measure than a simple percent change feedback. This indicator should also be exposed to machine learning to potentially develop an algorithm which uses STPP as a feeder to relay a more impactful measure of market cardinality or volatility than the simple measure our model uses.

Conclusion

While web scraping off Twitter and sentiment analysis offers great potential for unique trading algorithms, the data is not well curated. Using an off the shelf solution to conduct sentiment analysis, paired with loosely filtered and understood data, resulted in a weak indicator to incorporate into our analysis. This statement assumes that with better data treatment we can develop a better indicator. However, beating the market is not an easy task and even with the best minds and most advanced web scraping tools, many hedge funds still struggle to post an alpha off their tracked index. Our testing of the S&P 500 trends vs. our algorithm would have steady returns of (+/-)2% off of the normal trend. This means that our trading algorithm would most likely achieve around the SPY index, and if broker costs are entered into the calculation would give us a negative alpha. Overall, this is an ineffective prediction tool, and a passive investment in a low cost S&P 500 index would be a much better bet.

Citations:

CBOE. “CBOE Volatility Index (VIX) Index Rules and Methodology”. CBOE Online. April, 2018.

Hershman, Brett. https://www.benzinga.com/news/17/12/10968604/10-financial-twitter-names-to-follow-in-2018

Investopedia. https://www.investopedia.com/terms/v/vix.asp

Investopedia. https://www.investopedia.com/bonds-signaling-inverted-yield-curve-and-potential-recession-4579830

https://www.marketwatch.com/story/finance-twitter-the-50-most-important-people-for-investorsto-follow-2018-12-13

https://www.thestreet.com/slideshow/14114006/1/15-of-the-best-finance-twitter-accounts-to-follow.html

Pandey, Parul. “Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)” https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

Appendices

Figure 1: The relationship between Option Swaps and S&P 500

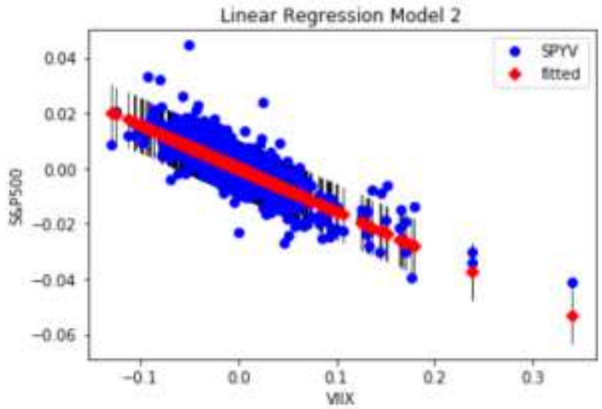

Figure 2: The relationship between Bond Inversion and S&P 500

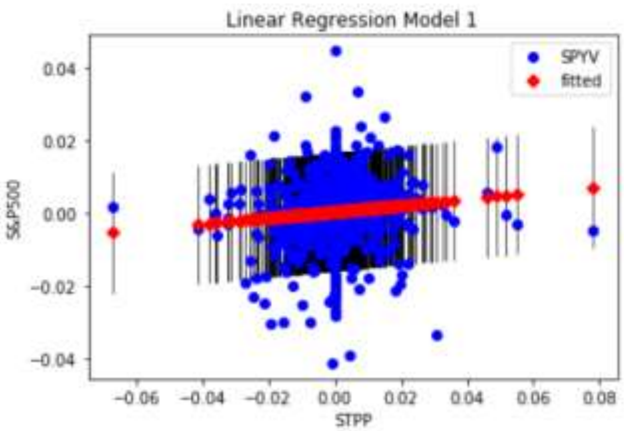

Figure 3: The relationship between Twitter Scores and S&P 500