TRADING & QUANTITATVE RESEARCH REPORT

INVESTWEETS

TRADING ON SENTIMENT

ANALYSTS: DAVID FRYKMER & ABDELJALIL MOUSSA

2018-12-09

THESIS, TRADING STRATEGY AND METHOD

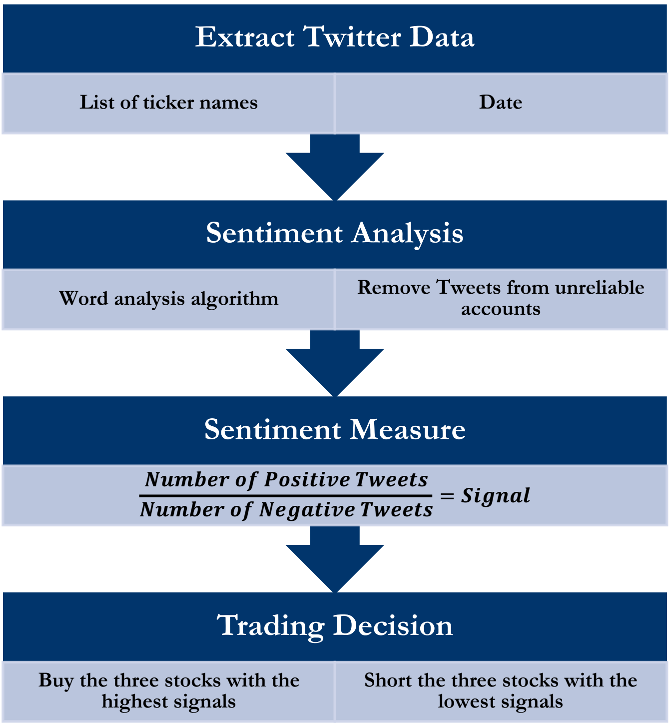

Exhibit A: Schematic overview of the trading system

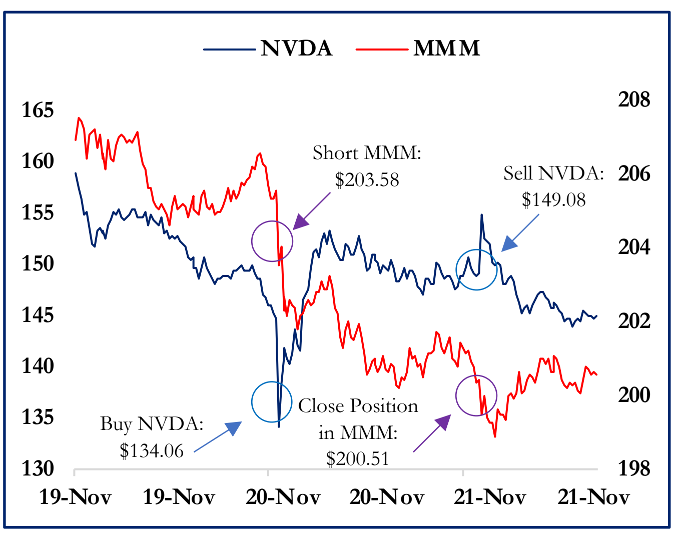

Exhibit B: Twitter posts mentioning the 26 stocks during the previous day were analyzed to predict changes in prices the following day (2018-11-20). The word analysis algorithm scored each Tweet: -1 for negative, 0 for neutral and 1 for positive. Then, the sentiment was estimated for each stock according to the formula above (see Exhibit A). Long positions were taken in the three stocks with the highest signals (JNJ, NFLX and NVDA) and short positions were taken in the three stocks with the lowest signals (MMM, FB and UTX). The positions were taken when the market opened and closed at the end of the trading day.

Investment Thesis

There is no doubt that sentiment influences decisions and thus market prices. To provide a measure of the market sentiment, Twitter data have been analyzed. Twitter is a unique platform for sharing thoughts instantly to followers around the world and hence allows us to capture changes in sentiment. Sentiment is becoming increasingly important for high-caliber traders. In this regard, Twitter is a great platform where professional investors as well as more unsophisticated traders discuss seriously about their investments and share their opinions on the stock market.

Why is the sentiment a useful way to predict stock movements? John Maynard Keynes was one of the first to incorporate sentiment, or in his own words “animal spirits”, into the financial markets and is consider to be a pioneer in the field of behavioral finance. Keynes reasoned that a rational agent does not necessarily price securities based on their fundamental value, but rather on what they believe everyone else thinks their value is. In other words, a rational agent would try to predict what the average person expects the average opinion to be. Our system will capture just this - the average opinion.

By estimating the sentiment, our trading strategy is able to exploit two major human behavioral biases. Firstly, basing the trading decision on what the majority believe allows the system to capture herding behavior, i.e. that people tend to do as the majority. Secondly, the system exploits that humans have limited attention span. No one can process all available information. Hence, people tend to only consider stocks that come to their attention. It is likely that the stocks that receive the most attention are the ones that the majority have a strong opinion about. Thus, the trading system exploits the limited attention span as well.

Trading Strategy

Initially, the trading strategy was to analyze Tweets from only 100 selected accounts. However, to derive a reliable sentiment measure, the originally selected Twitter accounts were not enough. The main problem was that for some days, none of the traders mentioned the stocks of interest. Consequently, a new approach was developed as part of the optimization process. Instead of restricting the analysis to only 100 accounts, the system analyzes the aggregate sentiment for 26 stocks listed on the S&P 500. The algorithm analyzes the Tweets mentioning the stocks of interest X hours before the market opens. Then, the trading system buys the three stocks that people are most positive about and shorts the three stocks that people are most negative about. The orders are placed before the market opens and the positions are closed at the end of the trading day (see exhibit A for a schematic overview). Additionally, an example of our strategy is illustrated in Exhibit B.

Method

The Tweets are extracted using the Twitter API in Python. To find the posts mentioning the stocks of interest, the API searches for Tweets containing the respective stock’s cash tag ($TickerName). Once the relevant Tweets are filtered, a word analysis algorithm is applied to the Tweets in order to estimate a sentiment score for each Tweet. Subsequently, an aggregate sentiment measure is attached to each stock. The three stocks with the most positive and negative sentiment respectively are then traded accordingly.

While this approach increases the noise in the data compared to only considering a list of selected traders, analyzing the aggregate sentiment is necessary to ensure enough data. The issue with noise is partly solved by removing the Tweets posted from accounts with less than 300 followers. To put in perspective, the median number of followers for the Tweets gathered is 515.

$$\frac{\textit{Number of Positive Tweets}}{\textit{Number of Negative Tweets}} = \textit{Signal}$$

Exhibit C: The adjusted formula used to calculate the signal that determines whether the stock should be bought or shorted.

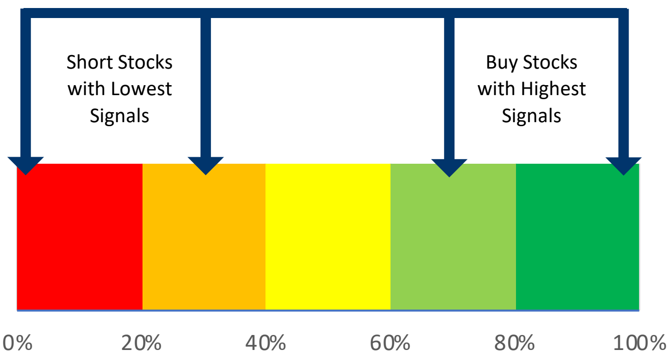

Exhibit D: Sentiment scale. A low percentage corresponds to a bearish sentiment.

Exhibit E: Sentiment categories illustrated.

Parameters

The main feature of the system is the sentiment measure derived by the algorithm and is based on how many positive and negative Tweets that are mentioning the stock. The sentiment measure captures the intensity of the sentiment as well as the overall mood. There are two factors that directly affect the sentiment algorithm; the restriction on how many followers the accounts need to have in order to be considered reliable and the number of hours considered before the market opens, i.e. the memory of the system. Both factors have been optimized and are discussed in more detail in the paragraph "Optimization".

As part of the optimization, the formula for the signal has been revised. The formula that was developed initially also took the number of total Tweets into account. The problem with the first formula was that the resulting signal was extremely insensitive to changes in the parameters (number of followers and memory). In other words, the system almost always predicted the same stocks regardless of the parameter values. Therefore, after trying some different formulas, the ratio of positive and negative Tweets was deemed to be appropriate for capturing changes in the sentiment (see Exhibit C). If there are no positive or negative Tweets mentioning the stock, there is not enough data for estimating a reliable signal and hence the resulting signal is not considered when determining which stocks to buy and short respectively.

Code Conditions

As illustrated in Exhibit D, the sentiment measure is normalized and converted to a percentage which makes it easier to visually compare the sentiment measures to each other. Once each stock has received a sentiment score, an algorithm finds the three stocks with the highest and lowest scores respectively and forms a portfolio where long positions are taken in the three most positive and short positions are taken in the three most negative.

To demonstrate the sentiment analysis algorithm, Tweets from Elon Musk have been analyzed. As mentioned earlier, the algorithm analyzes the content of the Tweets and estimates a sentiment score. The sentiment analysis algorithm assigns each Tweet with a score (+1, 0 or -1) which is then divided into three roughly defined categories. In Exhibit E, a sample output of the sentiment analysis algorithm used in the trading system is presented which illustrates the different sentiment categories. A Tweet is considered positive when the author is using a lot of positive words, but is considered negative when it does not make sense, or when the author uses negative words. For now, the sentiment analysis algorithm works fine and does not need any further adjustments.

Optimization

Our current resources only allow us to do a limited number of calls to the Twitter API every 15 minutes. In addition, it is only possible to extract data ranging from today and one week ago. More resources would make it possible for us to extract older Tweets. To circumvent the current limitations, the Tweets mentioning the stocks have been saved in a separate Excel sheet. This allows us to reload the data and analyze the Tweets using different parameters and hence improve the algorithm.

As of today, 140 000 Tweets are stored externally, which indeed is a considerable amount of data. For every day, the Excel sheet will be updated which allows us to quickly gather more data and hence, the algorithm and parameters can continuously be optimized further. However, the optimization can be improved even more if data ranging over a longer period were readily available.

Since the optimization was done based on a limited data set, there might be an issue with overfitting, i.e. that the model works well for the chosen sample but not for predictions made out-of-sample. As predictions are made each day, potential issues with overfitting will become evident. In addition, because more data is gather every day, it is possible to adjust and optimize the parameters over time.

Memory

Threshold for Number of Followers

	4h	8h	16h	24h	32h	40h	48h
0	0,64%	1,18%	1,11%	1,09%	1,55%	1,54%	0,25%
100	2,77%	1,11%	2,83%	0,64%	0,73%	0,43%	0,65%
200	1,75%	0,97%	1,21%	0,98%	0,65%	0,83%	0,12%
250	1,23%	1,98%	0,26%	0,41%	0,84%	0,79%	0,55%
300	0,82%	1,88%	-0,52%	0,22%	0,03%	0,71%	-0,12%
400	2,35%	0,31%	0,49%	0,44%	0,08%	0,79%	0,84%
500	1,58%	1,23%	0,52%	0,23%	1,09%	1,04%	0,97%
750	1,61%	-0,86%	-0,64%	-0,24%	0,96%	0,65%	-0,06%
1000	-0,50%	-1,67%	-0,34%	0,39%	-0,44%	-0,15%	0,41%

Exhibit F: Outcome of the trading strategy when changing the memory and the threshold for the number of followers needed for the algorithm to take a Tweet into account when calculating the signals.

Exhibit G: Unfavorable position in Facebook on the 19th of November where the loss could have been minimized with a stop loss.

Optimization cont.

Exhibit F shows how the portfolio performed when varying the two parameters; the number of followers required and how many hours to take into account before the market opens. The number of hours taken into account can be viewed as the memory of the algorithm. In general, the longer the memory, the worse is the outcome. This is in line with our intuition and implies that more recent Tweets affect stock prices more. Moreover, the outcome tends to worsen as the threshold for the number of followers increases. Even though the size of the sample used for optimizing the parameters was rather small, the heat map provides useful guidance and shows that the best outcome is generally when the memory is short and the threshold is low. Based on the optimization, a memory of eight hours and a threshold of 300 followers were chosen. The performance of the strategy with those parameters is discussed below.

Minimizing losses

One could argue that potential losses are reduced by closing the positions at the end of the day. However, this aspect can be improved further. For example, if the estimated sentiment for a stock is extremely positive, one would expect to see a rather rapid increase in the price. If the prediction for some reason is wrong and instead there is a rapid decrease in the price, a stop loss can be implemented that tells the system to sell the stock if the price decreases by more than 1% during the first hours of the trading day. An example is illustrated in exhibit G. The algorithm suggested that a long position in Facebook should be taken on the 19th of November. However, the stock price quickly declined and at the end of the day, the position resulted in an individual loss of almost 2% (which was luckily countered by favorable positions in other assets). If the system would have automatically closed the position once the share price had dropped 1%, the loss would have been minimized and hence the overall performance of the portfolio had improved. Erasing historical gains due to one unfavorable prediction would indeed be devastating. Therefore, it would be beneficial to implement a condition that instructs the system to close the position if a certain loss is already made. Optimally, such condition should be based on the historical volatility of the particular stock.

Performance

As mentioned above, our current resources only allow us to collect data from one week ago. The data gathering process began 22/11-2018. Hence, the first trading day for comparing our predictions to the actual outcomes was 16/11-2018. Since then, the trading strategy has performed better than our benchmark portfolio (S&P 500). So far, our strategy has generated a return of 4.25 % (see exhibit H). During the same time period, the S&P 500 had a return of 2.64%. The S&P 500 had an initial sharp decrease mainly due to large outflows from the tech-sector, conflicting statements made by president Trump and a turbulent oil market while the trading strategy generated a steady return.

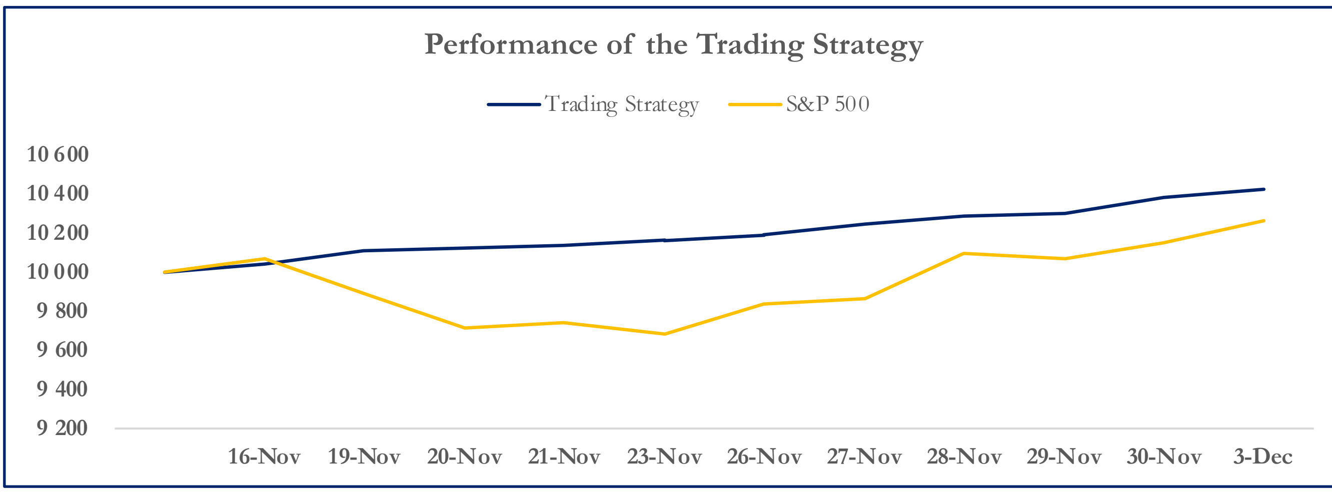

Exhibit H: Portfolio based on our trading strategy compared to the S&P 500. The graph shows the performance of the strategy if 10 000 SEK was invested on the 16th of November.

STATISTICS AND CONCLUSION

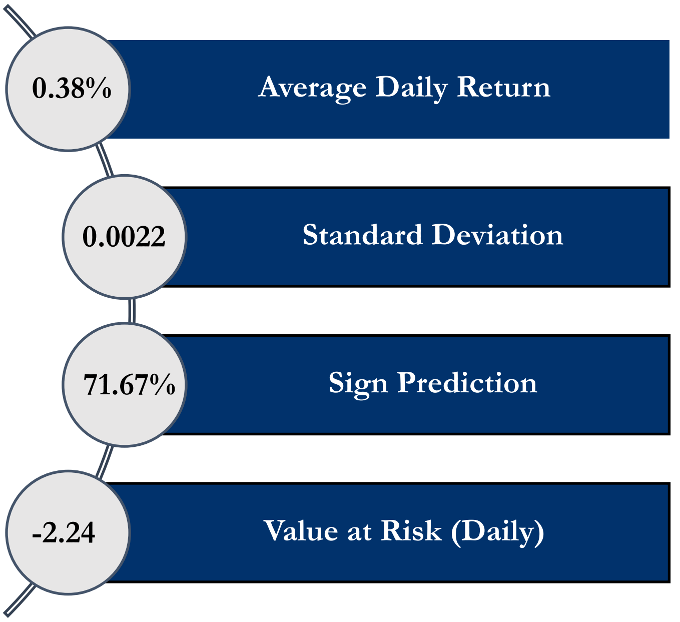

Exhibit I: Statistics for the portfolio. The average daily return is 0.38% with no daily negative returns. The standard deviation is estimated to be 0.0022. In about 72% of the trades, the system predicted the correct sign. If 10 00 SEK is invested in the portfolio, the daily value at risk is 2.24.

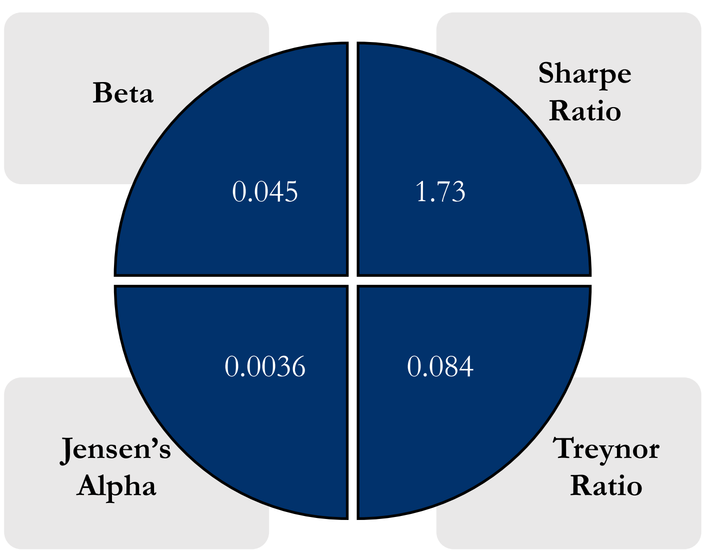

Exhibit J: Measures computed for the portfolio generated by the trading system. Jensen's alpha is significant at a 1% level while the beta is exceptionally low (not statistically different from zero).

Statistics

In Exhibit I, statistics for the trading strategy is presented. On average, the daily return of the strategy is 0.38%. Using the daily yield of US treasury bills, the daily excess return is then 0.373%. Additionally, it is worth highlighting that our strategy has not resulted in any negative daily returns so far. The estimated standard deviation of the portfolio is merely 0.0022.

Another interesting statistic is that the algorithm has predicted the correct sign in almost 72% of the trades. Expressed differently, for 72% of the trades, the algorithm predicted trades that were profitable, i.e. the stocks with a positive sentiment measure increased and the stocks with a negative sentiment measure decreased. Compared to the efficient market hypothesis (EMH), which suggests that assets follow a random walk and therefore one should only be able to predict the correct sign 50% of the time, the predictive power of the algorithm is indeed strong.

Lastly, assuming that the initial amount invested is 10 000 SEK, the daily value at risk (VaR) of the strategy is only 2.24 SEK. On an annual basis, the VaR based on the same initial investment is 35.53 SEK.

Portfolio Measures

Four measures have been computed for the portfolio that was generated by our trading system; the beta of the portfolio, the Sharpe ratio, Jensen's alpha and the Treynor ratio. The measures are presented in Exhibit J. The beta of the portfolio is extraordinarily low (not statistically different from zero), indicating that the trading strategy barely covaries with the market at all. Thus, the portfolio has no or little systematic risk. According to the capital asset pricing model (CAPM), a security with a beta of zero should earn the same return as the risk free rate. However, this is not the case since the average daily return is significantly higher than the risk free rate (US rate converted to daily returns equals 0.007%).

This is confirmed by Jensen's alpha, which implies that the portfolio earns abnormal returns. To test whether alpha is statistically different from zero, an OLS regression was performed. The results from the regression implied that alpha was significant at a 1% level, suggesting that the portfolio indeed generates abnormal returns. In addition, the magnitude of the Treynor ratio confirms that our trading strategy is earning risk-adjusted excess returns. The Sharpe ratio of the portfolio is 1.73 which is generally considered to be good. As more trades are executed, the Sharpe ratio is believed to stabilize around 1.5.

Conclusion

There is no doubt that the sentiment greatly affects future prices. Twitter has proven to be a useful way to estimate the aggregate sentiment. Our trading system is designed to capture changes in sentiment while at the same time exploiting human behavioral biases. Indeed, the trading system looks promising and the portfolio generated by the algorithm has earned a total return of 4.25 % during a short period of time (11 trading days) while at the same time carrying a low risk. Jensen's alpha is significant at a 1% level, implying that the returns are abnormally high. At the same time, the beta of the portfolio is extremely low. Additionally, the volatility of the strategy is low and thus appealing to risk averse investors.

Since the algorithm is able to predict the correct sign in 72% of the trades, the returns can be improved more by optimizing the parameters further as well as implementing more conditions (e.g. conditions for closing an unfavorable position early).

The long-term goal of the strategy is to generate abnormal returns. In the long run, the strategy is believed to generate the same return as the market, but to a substantially lower risk. Financial markets may have become more efficient as information spreads faster and becomes more readily available. However, unless investors suddenly lose their behavioral biases, sentiment will continue to have a substantial impact on asset prices. Hence, the strategy is believed to keep generating risk-adjusted excess returns in the foreseeable future.