How do twitter bots affect your revenue?
--A measurement between Twitter bots' activity rate and Stock market volatility

Shiwei Dong Yinan Wang
Email: {shido949, yinwa337}@student.liu.se
Supervisor: Alireza Mohammadinodooshan, { alireza.mohammadinodooshan@liu.se}
Project Report for Information Security Course
Linköpings universitetet, Sweden

Abstract

Twitter bots use cashtag piggybacking to increase the buzz of small companies by inserting their ticker symbols after the ticker symbols of large companies with high buzz, thus influencing the stock prices of small companies. This additional method, which is not in line with the actual economic market development, hurts the stock market and requires further research. In this study, we focus on the impact caused by Twitter bots on the stock market. We propose a new measure to describe the likelihood that a company is affected by bots. Then, we capture companies highly affected by bots, collect their stock data, and calculate their volatility and the probability that the users who mention the company are bots. The impact of Twitter bots on stock market volatility is verified by observing the temporal pattern of both.

1. Introduction

Since social platforms started to expand their sphere of influence, "bot" has become a presence in this environment that cannot be ignored. Social media bots are some bot programs that control accounts on social media platforms. They could be used to spam misinformation to manipulate public opinion and stock markets. The influence of social media bot activities is more and more prominent.[1] There have been many cases where the volume of bots twittering increases suddenly while the sharp stock changes happen. Therefore it is essential to detect and analyze the relations between stock market volatilities and bots' activities.

Thus we concentrate on analyzing Twitter bot activity rate and specific stocks to reach these two goals:

1. Explain how Twitter bots will influence the stock markets.
2. Investigate whether bots are more activated during the period when some stocks have more volatility.

Also, this research can lead to a further investigation of the correlation between the stock and bot activity.

2. Background

2.1 Twitter bot

A Twitter bot is a type of bot software that controls a Twitter account via the Twitter API. The bot software may autonomously perform actions such as twittering, re-twittering, liking, following, unfollowing, or direct messaging other accounts. These simple and basic operations allow bots to spread misinformation, such as rumors and meaningless spam. Some large-scale bot activities could even affect political elections, further affecting social stability. [2]

Financial bots are also one of them. "...on 1 May 2017, there was an upsurge in the number of 'bot' tweets with positive sentiment for Pearson from 18 to 4339. On the same day, the Pearson stock price increased by 1.01%, compared to a 0.63% rise of the FTSE 100." [3] Such kinds of bots will post a large amount of spam on Twitter, to mislead the user, and affect users' actions on stocks, like buying and selling.

2.2 Stock market behavior

Serena Tardelli[5] et al. study the U.S. economic market after discovering the cashtag piggyback phenomenon, a way for bots to influence the Twitter online economic market. They capture all tweets that mention one of the 6689 ticker symbols on the NASDAQ list on Twitter, a list of companies traded on the four major U.S. markets: NASDAQ, NYSE, NYSEARCA, and NYSEMKT. They also find another smaller market, OTCMKTS, which contains 22,956 stocks and has almost no overlap with the NASDAQ list. Stocks within the OTCMKTS market are chosen for the study because the scientists find that the ticker symbols within this market had an extremely high retweet rate of 76% after being mentioned on Twitter.

Usually, larger companies with higher assets have a higher volume of discussion on social media. Tardelli et al. test the social and economic impact using the Spearman method, and the experimental results show a correlation coefficient close to 0.5, which verified their suspicions. However, if only considering the peak discussion segment, the correlation coefficient between the social and economic impact of the OTCMKTS market is about -0.3. It can also be summarized that the fewer the stocks assets in this market, the higher the amount of discussion on Twitter. Therefore this market is more likely to be affected by bots than other large stock markets. Therefore, we select stocks within the OTCMKTS market in our study.

After the initial selection of the market, we proceed to understand the market further. The OTCMKTS market[1] is subdivided into two segments: OTCQX and OTCQB. The OTCQX market is mainly for larger and more established U.S. and international companies. If a company wants to enter the OTCQX market, it needs to meet specific economic criteria. In contrast, the OTCQB market faces smaller companies at their early stage, and the rules are relatively lax. All that is required to enter this market is to pass an annual inspection and not go bankrupt. Based on Tardelli's research, we believe that the OTCQB market within the OTCMKTS market is more susceptible to bots because these companies tend to have smaller market capitalizations. Also, it can be analyzed from a social level that companies that are just starting are more inclined to promote their companies through an easy and quick method like buying bots. In this way, more people can learn about their companies through Twitter and thus buy their shares.

2.3 Stock volatility

Volatility is a measure of the magnitude of changes in the price of a security, and it is often measured as either the standard deviation or variance between returns from the same stock or market. Usually, the higher volatility means huge price fluctuations. Therefore the risk and reward are greater. There are two types of volatility in the definition of economics, Implied Volatility(IV) and Historical Volatility(H.V.). The first one is also known as projected volatility, and it could be used to determine the volatility of the futures markets. The historical volatility, which is analyzed in our research, estimates the fluctuations over a determined period of time of stocks by measuring price changes. Compared to the IV, H.V. has less predictive.

We use a formula to measure the historical volatility of stocks[6]:

$$VOL_i = \sqrt{\frac{\sum_{t=1}^{N}\left[\left(R_{it} - \frac{\sum_{t=1}^{N} R_{it}}{N}\right)^2\right]}{N-1}}$$

Where P_{it} is the closing price of stock i on day t, and N is the window period.

R_{it} could be calculated with the formula as follows:

$$R_{it} = \frac{P_{it}}{P_{i,\,t-1}} - 1$$

According to the calculation method we mentioned, we aim to find some stocks with high volatility. We can carry on to the next step to set up a filter with the stocks we found.

2.4 Bot activity rate

Normal users and bot users behave differently on Twitter[8]. Human users mainly create different textual content and interact with others, while bot users are more mechanized in liking and retweeting or sending the same message repeatedly. Therefore, based on each user's behavior on Twitter, such as sending tweets, retweeting, liking, following, etc., we can give the probability that the user is a bot.

We capture the Twitter users who mention the cashtag of the target company, analyze them, give the probability that each user could be a bot, and finally give the average value. The formula is as follows.

$$BotRate_i = \frac{\sum_j p_{ij}}{N_j}$$

2.5 Yahoo Finance API

Yahoo! Finance, established by Yahoo, is a website where users can search for stock market information. Users can search for real-time trading information, closing prices, trading volume, analysis information, etc. Currently, Yahoo! Finance can also search cryptocurrencies and currency exchange. The site also provides a variety of APIs for developers to use efficiently, which can be used to query and download information in large quantities. Unfortunately, Yahoo! Finance has discontinued the API for downloading historical stock information. Therefore, based on the API of Yahoo! Finance, various libraries were born. In this study, we chose to use this library of yfinance to conduct research, downloading closing data in bulk for the research period (2021.01-2022.04) and calculating volatility.

2.6 Botometer Pro API

Botometer® is a Python API that offers a methodology to check how much a Twitter account is likely to use automation. We use Botometer-V4 in our research to obtain the most accurate results possible, and there are three main steps:

1. We obtain user info and tweets from the accounts we choose through Twitter's official API.

2. We send data to Botometer's server in JSON format.

3. Botometer analyses the data and return scores in JSON format.

By using this API, the program will give a score for such an account, and higher scores indicate more bot-life activities.

3. Method

3.1 Analyze Stock behavior

In our research, a method is used by bots called cashtag piggybacking, which is perpetrated by coordinated groups of bots and aims at promoting low-value stocks by exploiting the popularity of high-value ones.[4]

Similarly, the size of a company is proportional to its social importance (average number of Twitter). Nevertheless, if Twitter bots use cash piggybacking to contribute to a small company's stock price, the number of Twitter goes higher. To give a quantitative measure, we set up an index called the ***stock twitter behavior index (STBI)***. The formula is below:

$$STBI_i = \frac{N_i(Tweets)}{Size_i(Company)}$$

The higher STBI is, the more suspicious a company is influenced by Twitter bots.

3.2 Process user data

In the OTCQB market, we collect mini companies with high STBI values.

We defined ***the stock twitter behavior index (STBI)*** (Section 2.3), calculated by tweet counts divided by company size. According to actual data, we recognized market size as its capitalization.

Firstly, we observed the volatility of the QTCQB market and found that from 2021-11-12 to 2022-01-04, there was a sharp fluctuation during that period, which suggested more bots activity. Then, we calculate the quotient of tweet counts and market capitalization and the distribution is figure 1.

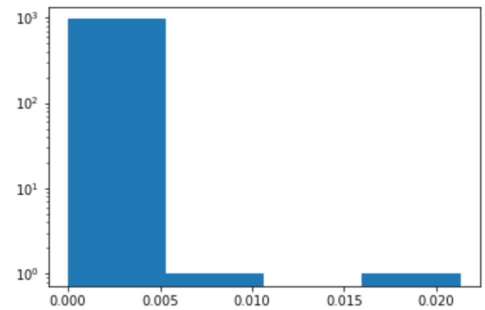

Figure 1: STBI-distribution histogram

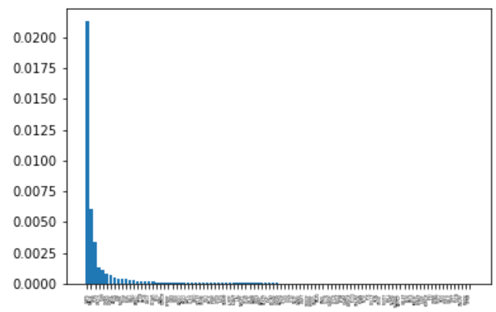

Figure 2: Target companies and its STBI

After calculating the STBI of each company in the OTCQB market, we sort the list by STBI and choose the first 100 out of 1280 companies as our target companies. We expand the period with target companies found during this period and obtain a more expansive users list to make our research more convincing. The users' list starts from 2021-01-01 and lasts until now. Moreover, we use Botometer to test the bot scores of each user.

Finally, there are 20174 users, with 2917 of them being unique. In the user lists, one of them, user 1404188416556797952, failed. One possible explanation is that Twitter has detected and removed this user.

3.3 Process Stock data

After catching a high volatility period for each target company, we collect corresponding users who mention the target company in the corresponding period.

Considering the number of companies is much more than expected, we bucketize 100 companies into 3 clusters according to their volatility.

The purpose of clustering is to coarsely classify the companies based on volatility to observe the general pattern of volatility and robot activity rate. By analyzing the stock price volatility (figure 3), the stock prices of the target companies do not differ much from each other and have a general downward trend. Therefore we calculate the highest and lowest values for each stock and classify them according to the difference between the highest and lowest values. The rationale for choosing the difference as the basis for clustering is that: The difference represents the maximum range of volatility, and the larger the range, the greater the volatility, given that the stock prices are similar overall.

We refer to the clusters of companies resulting from clustering as the high volatility group, the medium volatility group, and the low volatility group.

Besides, by observation, we find that the stock price fluctuation of the company with the highest STBI value with the code CAPS is abnormal compared to the others, and the stock price stays unchanged for a period of time and then suddenly falls or rises, almost in a vertical pattern. Therefore, we considered this stock as an outliner and removed it from the list of companies. We then sort the remaining 99 companies by difference calculation and group them into three groups.

Then, after the clustering step, we use the volatility formula mentioned in 2.3 to calculate the average volatility of each group, and the final result is shown in Figure x. we found no stocks with much higher market capitalization, so the first possibility is excluded.

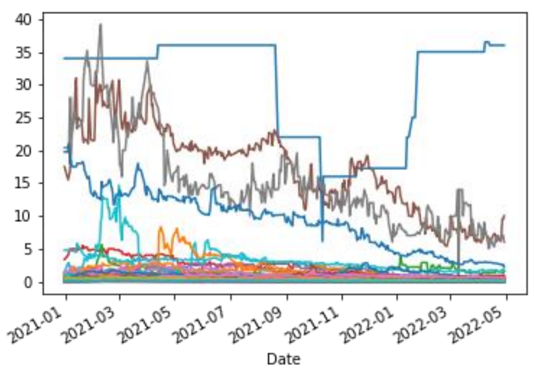

Figure 3: Stock price plot (from 2021-01-01 to 2022-04-30)

3.4 Analysis of CAPS

Capstone Therapeutics Corp., ticker symbol on OTCQB is "CAPS." It is worth noting that this stock is moving somewhat unusually. By observation in Figure 3, its stock price barely changed during the past year. Thus we presume that this company has applied for suspension of trading for a number of reasons. Moreover, "caps" is a common word with its own meaning, and it has also been associated with certain topical events. For example, there is an E-sport player whose game ID is "caPs." His ID is often mentioned in some game-related tweets, which will pollute our dataset and lead to reduced accuracy of the dataset.

Therefore, we considered this stock as an outliner and removed it from the list of companies. We then sort the remaining 99 companies by difference calculation and group them into three groups.

After grouping, we use the volatility formula mentioned in 2.3 to calculate the average volatility of each group, and the final result is shown in Figure 5-10.

3.5 Comparision

With the data we obtained, we draw two groups of figures:

The volatility of stock clusters.

The bot activity of corresponding stock clusters.

Then, we will analyze these two groups of figures.

The stock price after removing the outliner is shown below:

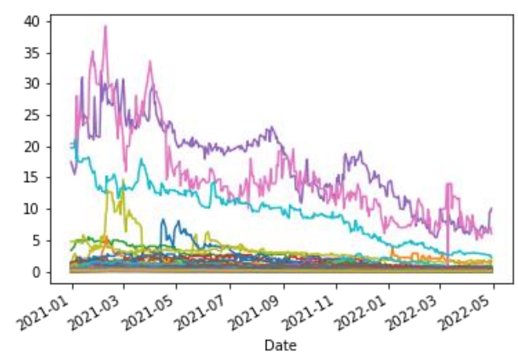

Figure 4: Stock price of target companies

Due to the wide disparity in the distribution of stock prices of different companies, we divided our target companies into three groups: the high, med and low. The volatility is calculated after grouping and the results are as follows:

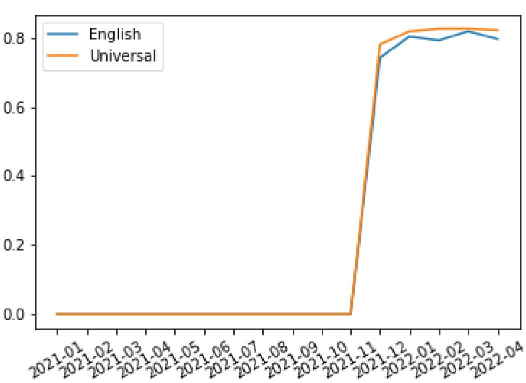

Figure 5 Bot activity rate in High group

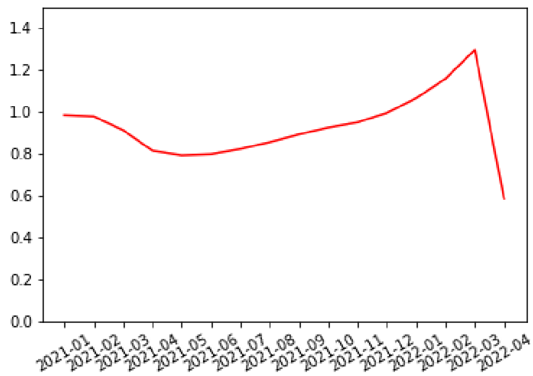

Figure 6 Stock volatility in High group

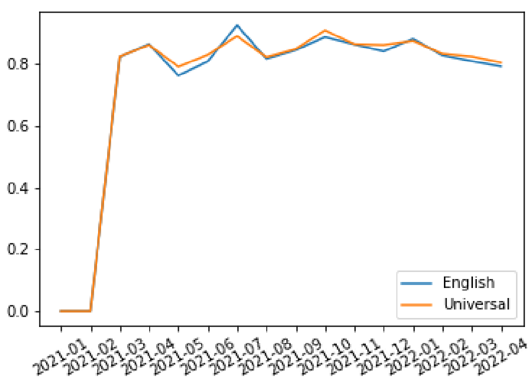

Figure 7 Bot activity in Medium group

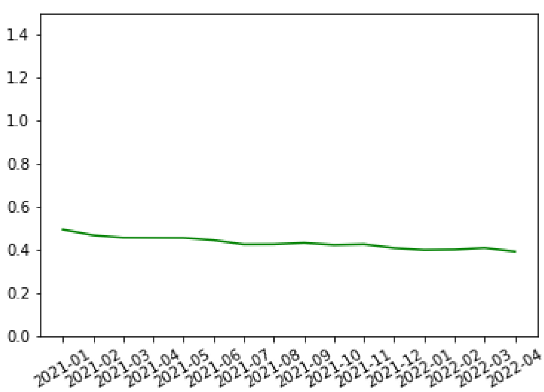

Figure 8 Stock volatility in Medium group

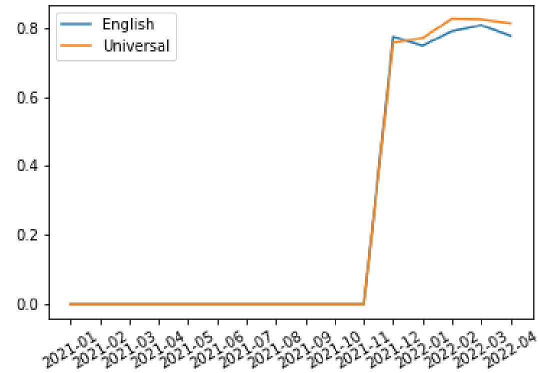

Figure 9 Bot activity in Low group

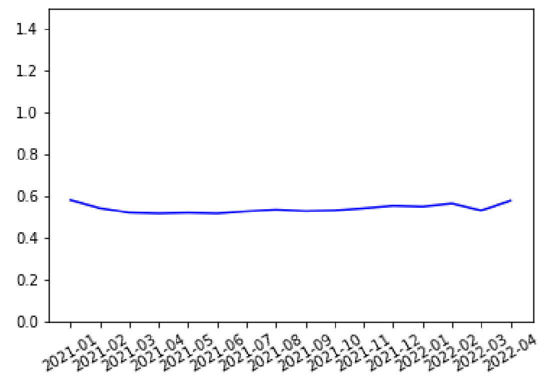

Figure 10 Stock volatility in Low group

4. Result

There is a time similarity between stock volatility and bot activity, bot activity is higher when stock volatility is high, especially in the High-volatility group: bot activity started in December 2021, while stock volatility is also very high in the high volatility group, both English users and Universal users. In the Medium-volatility group and Low-volatility group, stock volatility are stable. It can be inferred that bots activity effects them slightly. From an overall perspective, we cannot give a clear positive correlation between stock volatility and bot activity rate.

Another point is that since twitter itself detects bots and blocks them, it is possible that the bots before January 2021 have already been blocked, so the user data is not available now.

5. Related work

Because of the increasing activity of bots in recent years and their impact on the stock market, more and more scholars are studying the activity of bots on Twitter. Stefano et al. demonstrated that bots increase the discussion of small companies by appending the stock code of large companies to the stock code of small companies, which impacts the stock price fluctuations of small companies. Later, Tardelli et al. conclude that if the social impact is much higher than the economic impact, the market is likely to be heavily manipulated by bots by looking at different stock markets' social and economic impacts.

From a broad perspective, scholars have studied the impact of Twitter on stock markets. Johan Bollen's article "Twitter mood predicts the stock market"[7] analyzes this phenomenon.

6. Conclusions

The aim of this paper is to demonstrate that bot influences companies on Twitter through the cashtag, thus affecting the stock market. Those bots post a large number of tweets to pollute the "information pool", and use this method to manipulate the stock markets. The results showed that, with the companies stock price fluctuating, more bots are activated, this is more evident in the group of companies with higher stock prices.

The main limitation of this study is that, due to time, companies were selected after obtaining short-term tweet counts. Then the list of users was expanded by directly expanding the period to enhance persuasiveness. This expansion practice is not rigorous, and further argumentation should be required to collect more extended time company tweet counts and cross-compare the selection of companies.

References

[1] Ferrara, Emilio; Varol, Onur; Davis, Clayton; Menczer, Filippo; Flammini, Alessandro. *The Rise of Social Bots*. Communications of the ACM 59 (7), 96-104, 2016. https://doi.org/10.1145/2818717

[2] Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A. and Menczer, F. *Detecting and tracking political abuse in social media.* In Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (2011). 297–30

[3] Fan R, Talavera O, Tran V. *Social media bots and stock markets.* Eur Financ Manag. 2020;26:753–777. https://doi.org/10.1111/eufm.12245

[4] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. 2019. *Cashtag Piggy-backing: Uncovering Spam and Bot Activity in Stock Microblogs on Twitter.* ACM Trans. Web 13, 2, Article 11 (April 2019), 27 pages. https://doi.org/10.1145/3313184

[5] Tardelli, S., et al. (2020). *Characterizing Social Bots Spreading Financial Disinformation.* Cham, Springer International Publishing.

[6] Zhou, Dongyi; Zhou, Rui. *ESG Performance and Stock Price Volatility in Public Health Crisis: Evidence from COVID-19 Pandemic*. JAN 2022; 19; 1. https://doi.org/10.3390/ijerph19010202

[7] Bollen, J., et al. (2011). "*Twitter mood predicts the stock market.*" **2**(1): 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

[8] Chu, Zi & Gianvecchio, Steven & Wang, Haining & Jajodia, Sushil. (2010). Who is Tweeting on Twitter: Human, Bot, or Cyborg?. Proceedings - Annual Computer Security Applications Conference, ACSAC. 21-30. 10.1145/1920261.1920265