IMPACT OF INTERNET FORUM TRENDS AND RETAIL INVESTORS ONLINE DISCUSSION ON STOCK PRICE: ANALYZING *WALLSTREETBETS* USERS' ACTIVITIES IN THE 1ST & 2ND QUARTER OF 2021

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Jiachen Song

August 2021

ABSTRACT

A series of short squeeze activities during the last six months has drawn the public's attention to an online forum called *WallStreetBets*. As the community which starts the whole short squeeze series, the posts and discussions on *WallStreetBets* are highly related with the stocks they choose as targets to short squeeze. The aim for us is to clarify the influence of *WallStreetBets* discussions on stock prices. In this paper, the Granger Causality tests conducted on 15 stocks prove that for a certain stock, either a one-directional relationship or a bi-directional relationship exists between its stock price and its discussion frequency on *WallStreetBets*. To further examine the relationship, we use 2SLS and OLS regressions to acquire numerical and specific results for the elasticity between 15 stocks' prices and discussion intensity on *WallStreetBets* respectively.

BIOGRAPHICAL SKETCH

Jiachen Song was born in Xi'an, China in 1997. He spent his childhood in Xi'an. During his adolescence, he studied in and graduated from Xi'an Tieyi middle school and Xi'an Tieyi international high school. Soon after, he went abroad to study in the United States. He received his bachelor degree from the Department of Economics at Boston University with minors in Business Administration as well as Japanese Language & Culture in 2019. For further study in Economics, he started his journey to his Master's degree at the Dyson School of Applied Economics and Management at Cornell University. During his graduate study in Cornell University, he followed Professor Calum Turvey and focused on livestock insurance and stock trading.

ACKNOWLEDGMENTS

For my 2 years of study at Cornell and the completion of this thesis, I would like to thank deeply to my advisor, Professor Calum Turvey, for the knowledge he has taught me and his effort in helping me. After the COVID-19, our original project was cancelled due to quarantine policies. Our whole year of preparation was ruined. Professor Turvey and I did not give up and turned our sights into studying the stock market, which leads to this thesis. Moreover, Professor Christopher Allen Wolf, who serves as my committee minor, gives precious understanding and support.

I would also like to thank my dear friends in Ithaca, Bingyan Dai, Guanghan Xie, Hongyao Fei and so on. The time spent with them enriched my two years at Cornell.

Apparently, the two people who own my deepest respect and gratitude are my parents, Wentao Song and Huaping Chen. With their love and support, I become who I am now.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH ---------- iii

ACKNOWLEDGMENTS ---------- iv

TABLE OF CONTENTS ----------v

CHAPTER 1: INTRODUCTION---------- 1-9

CHAPTER 2: LITERATURE REVIEW----------9-13

CHAPTER 3: GRANGER CAUSALITY TEST

Section 1: Methods ----------13-16

Section 2: Results ----------16-28

Section 3: Discussion----------29-30

CHAPTER 4: REGRESSION ANALYSIS

Section 1: Methods ----------30-36

Section 2: Results ----------36-47

Section 3: Discussion ----------47-51

CHAPTER 5: CONCLUSION----------51-54

REFERENCES----------55-57

APPENDIX----------59-63

CHAPTER 1: INTRODUCTION

The development of network technology and the surge of Internet users during the last decades is one of the most significant events shaping modern society. The Internet, along with the access to unlimited knowledge resources and the rapid global communication brought by it, affects nearly all aspects of government policies and daily life; Finance, undoubtedly, is one of those aspects. The Internet has reshaped the financial sector with new financial industries and financial activities evolving in-step with the technologies. For the stock market the Internet and the traditional stock trading market have been integrated to form online stock trading systems and platforms, increasing low-cost access to new retail investors and accelerating both the quantity and the speed of trading. In addition, abundant online information provides investors with strategic insights to make trade decisions.

An important economic issue is the understanding of how stock prices are influenced by changes brought about by the Internet. In order to solve this question, researchers have examined the impact of the popularization of the Internet and the development of network technology on stock prices from various angles. Bogan (2008) studied the relationship between higher stock market participation rates from online retail investors and the stock price; Hirschey et al (2005) focused on the effects of online announcements and recommendations from financial giants on stock prices; Xu et al(2021) used a quasi-natural experiment on how the availability of online information searching affected stock prices by using the withdrawal of Google in China in 2010. Despite research such as these, the reality is that no single study can explain the entirety of the impact of the Internet on stock prices. In this vein the economic motivation driving the current study was the recent *WallStreetBets* Short Squeeze

phenomenon and whether forum trends and discussions actually have/had an impact on stock price valuations.

To this end, the overall objective of this paper is to determine whether there is a direct relationship between the intensity of chatter about a particular stock on *WallStreetBets* and the follow-on change in stock prices. There is a possible simultaneity underlying this problem in that it is not so clear whether changes in prices lead to more chatter, or whether more chatter leads to changes in prices. or both? Or neither?. To investigate, I use Granger causality testing. The Granger results are central to defining which (of 15 chosen) stocks evaluated were identified as short squeeze targets, which were successful if so, and which were not. In this context the Granger tests are a means to the end, and not an end in itself. Follow-on regressions use two-stage least squares (2SLS), and OLS as appropriate. To my knowledge, given the recency of events related to WallStreetBets, this is the first investigation to make a determination that forum chatter and stock price movements are causally related. Having defined stocks according to degrees of causality, I then explore the relationships with greater granularity. Using a number of control and instrumental variables I seek to determine the magnitude of effect from forum chatter in isolation of other factors including general price movements in the S&P 500, and short positions.

WallStreetBets Short Squeeze is a series of short squeezes on several stocks initiated by users of *WallStreetBets. WallStreetBets* is a sub-Reddit online forum devoted to stock discussion and high risk trading. The forum is known for its profane and juvenile nature and members' blind, almost cult-like following of trading recommendations. The forum is the World's largest retail investors' gathering place. with 10.5 million members as of June 10, 2021. ("Wallstreetbets", 2021) Based on the massive volume of *WallStreetBets* users and posts, Dolbec (2020) identified the potential hidden within *WallStreetBets* on affecting the stock market and even

reshaping financial markets. Throughout 2020 and into 2021, WallStreetBets began to focus on heavily shorted stocks and collectively determined to short squeeze investors. Short Squeeze is a term describing a stock price rising situation mainly due to market supply and demand, rather than changes in the company's fundamentals. When short sellers (Investors who borrow stocks and sell to others, planning to buy them back later for less money to gain profit) are forced to close their positions and continue to buy the underlying stocks, the demand for stocks on the market far exceeds the market circulation, and when the price rises sharply due to insufficient supply comparing to the demand, a short squeeze may occur.

The 2021 *WallStreetBets* short squeeze series began with the Electronics Retailer – GameStop Corporation. Due to the competition from online gaming platforms and services, GameStop continued to lose customers. In addition, after the covid-19 lockdown, the number of people shopping in person decreased sharply, causing the situation of GameStop to worsen. These forces caused the stock price of GameStop to fall which in turn brought about increasingly negative opinions on GameStop fundamentals thereby attracting institutional investors to short-sell GameStop stocks and options. On the opposite side, so-called juvenile *WallStreetBets* users neglected the pessimistic fundamental analysis of GameStop and continued purchasing Gamestop stock. GameStop was known to many forum members because of their childhood gaming memories which involved purchasing games from GameStop. Institutional investors and *WallStreetBets* individual investors were in a delicate balance. The triggering event was on Jan 19th, 2021 when a report written by the famous short-selling institution *Citron Research*, addressed and labeled Gamestop stock buyers as “suckers”. This, apparently, struck a collective nerve. Angry *WallStreetBets* members accelerated activity on Gamestop stocks, buying more and holding. The forum was full of discussions among Gamestop champions. Following this epic trend, *WallStreetBets* succeeded in squeezing Gamestop shorts, causing its

stock price to soar in the following days. As of Jan 28th, 2021, the highest GameStop share price was US$483 (excluding after-hours trading hours), which was nearly 190 times its lowest price of US$2.57 (NYSE, 2021).

For the *WallStreetBets* community, it was just the beginning. On Jan 25th and Jan 27th, they led two more short squeeze trends on BlackBerry Limited and AMC entertainment stock. This second wave was even larger, drawing attention from retail investors outside of *WallStreetBets.* Individual investors around the US were madly buying and holding Blackberry and AMC stock regardless of the consequences. Due to this online carnival, according to Bloomberg (2021), the trading volume of the US stock market on Jan 27th (calculated by the number of shares) exceeded the peak of the financial crisis in October 2008 and it was the third highest daily trading volume in US in the past 13 years. Shortly after the GameStop, BlackBerry and AMC short squeezes, the *WallStreetBets* community triumphantly created a few more short squeeze trends, on Tilray, Sundial Growers, Nokia and Microvision respectively. Table 1 records short squeeze events caused by *WallStreetBets* and Tables 2 through 5 illustrate the daily stock open and close price change for several stocks during the short squeeze period. To the present day, the *WallStreetBets* community has expanded exponentially and is preparing to rise to the next trend. (To add on, *WallStreetBets* successfully short-squeezed Clover Health Investments on June 8th, 2021)

Table 1: Short Squeeze Record

Stock Code	Company Name	Short Squeeze Count	Short Squeeze Date
AMC	AMC Entertainment	2	Jan 27th ; May 25th
BB	BlackBerry Limited	1	Jan 25th
GME	GameStop Corp.	3	Jan 21st; Feb 25th; May 26th
TLRY	Tilray	1	Feb 10th

SNDL	Sundial Growers	1	Feb 10th
NOK	Nokia	1	Jan 29th
MVIS	Microvision	1	Apr 26th
CLOV	Clover Health Investments	1	June 8th

Table 2: AMC Stock Price during Short Squeeze

Date	**Stock Code**	**Company Name**
	AMC	AMC Entertainment
	Stock Open Price	**Stock Close Price**
1/15/2021	$2.20	$2.33
1/19/2021	$2.80	$3.06
1/20/2021	$3.29	$2.97
1/21/2021	$3	$2.98
1/22/2021	$2.91	$3.51
1/25/2021	$4.71	$4.42
1/26/2021	**$5.09**	**$4.96**
1/27/2021	**$20.34**	**$19.90**
1/28/2021	$11.98	$8.63
1/29/2021	$14.31	$13.26
5/27/2021	**$18.61**	**$26.52**
5/28/2021	**$31.81**	**$26.12**
6/1/2021	$31.89	$32.04
6/2/2021	$37.52	$62.55
6/3/2021	$58.10	$51.34

Table 3: BlackBerry Stock Price during Short Squeeze

Date	**Stock Code**	**Company Name**
	BB	BlackBerry Limited
	Stock Open Price	**Stock Close Price**
1/15/2021	$10.68	$9.84

1/19/2021	$10.85	$12.35
1/20/2021	$13.23	$12.79
1/21/2021	$12.62	$12.85
1/22/2021	**$12.37**	**$14.04**
1/25/2021	**$19.55**	**$18.03**
1/26/2021	$19.09	$18.92
1/27/2021	$20.25	$25.10
1/28/2021	$19.47	$14.65
1/29/2021	$17.98	$14.10
2/1/2021	$14.64	$14.63

Table 4: Sundial Growers Stock Price during Short Squeeze

Date	**Stock Code**	**Company Name**
	SNDL	Sundial Growers
	Stock Open Price	**Stock Close Price**
2/8/2021	$1.15	$1.28
2/9/2021	**$1.39**	**$1.65**
2/10/2021	**$2.21**	**$2.95**
2/11/2021	$3.96	$2.38
2/12/2021	$1.93	$2.08
2/16/2021	$2.24	$2.19
2/17/2021	$1.94	$1.72

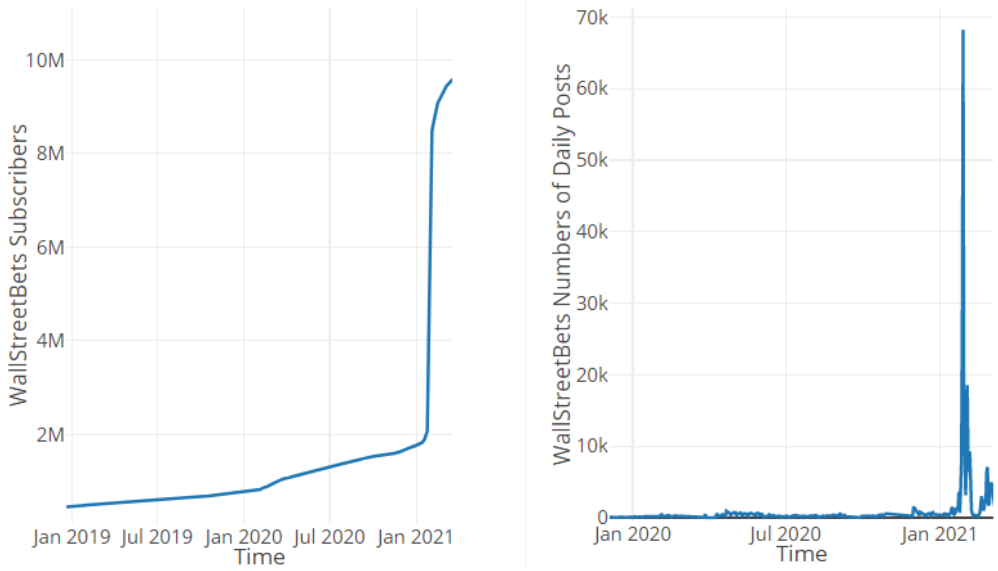
Table 5: GameStop Stock Price during Short Squeeze

Date	**Stock Code**	**Company Name**
	GME	GameStop Corp.
	Stock Open Price	**Stock Close Price**
1/15/2021	$38.49	$35.50
1/19/2021	$41.55	$39.36
1/20/2021	$37.37	$39.12
1/21/2021	$39.23	$43.03
1/22/2021	$42.59	$65.01
1/25/2021	$96.73	$76.79
1/26/2021	**$88.56**	**$147.98**
1/27/2021	**$354.83**	**$347.51**
1/28/2021	$265.00	$193.60
1/29/2021	$379.71	$325.00

2/1/2021	$316.56	$225.00
2/23/2021	$44.97	$44.97
2/24/2021	**$44.70**	**$91.71**
2/25/2021	**$169.56**	**$108.73**
2/26/2021	$117.46	$101.74
3/1/2021	$104.54	$120.40
3/2/2021	$116.93	$118.18
2/23/2021	$44.97	$44.97
5/24/2021	$175.85	$180.01
5/25/2021	$181.00	$209.43
5/26/2021	$229.00	$242.56
5/27/2021	$229.80	$254.13
5/28/2021	$262.97	$222.00
6/1/2021	$233.48	$249.02

WallStreetBets after January 2021 is no longer like before; it has become the symbol of both the power of Internet trends and the unity of retail investors. Not only short squeeze activities themselves, along with growing numbers of active users and short squeeze trends, the numbers of discussions and posts on other common stocks, which are not short squeeze targets, have also shown a rapid upward tendency. The rapid growth of both *WallStreetBets* users and daily posts represent the increasing importance of Internet trends. *WallStreetBets* number of subscribers and daily posts' increasing trend is shown in Graph 1 and Graph 2.

Graph 1: *WallStreetBets* Subscribers data



Graph 2: *WallStreetBets* Daily Post Data

With the increasing popularity of *WallStreetBets*, how Internet trends brought about by *WallStreetBets* discussions affect the stock market and stock prices has become a new subject for economists to study. This paper addresses this question by using the number of times a particular stock was mentioned in a day, and whether there was a relationship between forum intensity and changes in stock prices. To assess this relationship this paper evaluates 15 WallStreetBets favorites. These 15 stocks are listed in Table 6.

Table 6: Information on Stocks used

Stock Code	**Company Name**	**Industry**	**Market Cap**	**Short Squeeze Count**
AMC	AMC Entertainment	Entertainment	$31.32B	2
BB	BlackBerry Limited	Software	$8.57B	1
GME	GameStop Corp.	Electronics retail	$19.97B	3
TLRY	Tilray	Pharmaceutical	$8.48B	1
SNDL	Sundial Growers	Pharmaceutical	$2.1B	1
NOK	Nokia	Telecommunications	$30.56B	1
MVIS	Microvision	Technology	$3.01B	1
CLOV	Clover Health Investments	Health	$1.3B	1
BBBY	Bed Bath & Beyond	Retail-store	$4.71B	0
NIO	NIO Inc.	Electric Vehicle	$64.58B	0
WKHS	Workhorse Group	Manufacturing	$1.42B	0
PLTR	Palantir Technologies	Software	$45.89B	0
TSLA	Tesla, Inc.	Electric Vehicle	$582.93B	0
AAPL	Apple	Technology	$2.09T	0
AMD	Advanced Micro Devices	Semiconductor	$99.6B	0

The 15 stocks were chosen from Top 30 stocks with the highest mentions in posts and discussions on *WallStreetBets* in the past 6 months. Besides high

popularity on *WallStreetBets,* another main criteria of choosing stock was variety. In order to examine the influence of *WallStreetBets* trend on stocks with different market capital, in different industries and whether being short-squeezed or not are important factors to consider. I divided stocks into two groups, with and without *WallStreetBets* short squeeze attempts. tocks with *WallStreetBets* short squeeze attempt experiences, including AMC, Blackberry, Gamestop, Tilray, Sundial Growers, Microvision and Clover Health Investments were successfully short-squeezed during the trend and underwent a price soar; Nokia was the only stock that *WallStreetBets* tried to short squeeze but aborted. For stocks with no short squeeze experience, I selected representative stocks in their industries and in their market capital size. For instance, Tesla represents the electric vehicle industry, Workhorse Group represents the traditional manufacturing industry, Apple represents giant companies with huge market capital, and NIO, as a China Concept stock, represents foreign assets. The following contents of this paper review existing literature on related topics, explain research methods as well as regression models, demonstrate research results and discuss the findings.

CHAPTER 2: LITERATURE REVIEW

Literature related to the impact of online forums trend on stock market can be divided into two parts to discuss: the pre *WallStreetBets* era and the post *WallStreetBets* era.

I. Pre *WallStreetBets* era Literatures

Financial economists have noticed that online financial forum messages may impact stock markets; Antweiler and Frank wrote the first paper focusing on this

subject in 2004. However, during the period between 2004 and 2020, before the first GameStop short squeeze event and *WallStreetBets* officially became famous, researchers tended to ignore the *WallStreetBets* community. No academic examinations directly linked to *WallStreetBets* and its users in this era have (to my knowledge) been published.. From 2004-2020, a number of papers evaluating the influence of either online forums in general or online public opinions on stock markets from different perspectives were published. Three of them, written by Antweiler and Frank (2010), Lin(2013) and Wu et al(2012), are most relevant to this paper. These are discussed below..

Antweiler and Frank's (2004) work was the earliest literature found online under this topic. In 2004, the predecessors of online forums today were called message boards. In their paper, they proposed a hypothesis that all the talks on message boards are just noise and have no influence over stock market and stock trading. After their empirical research, they rejected the hypothesis and concluded that Internet stock message boards can influence the market in some aspects. Here is the summary of their three core findings: (1) Messages cannot accurately predict subsequent stock returns; a positive shock to message board posting predicts negative returns; (2) Disagreements among messages posted on one day leads to fewer trades in the stock market on the following day; (3) Messages can help to predict volatility of stock prices.

Wu et al (2012) proposed a stock market analysis method based on public sentiment reflected from Chinese online forums. They used an Internet crawler to obtain all the posts in online forums and did statistical analysis on the number of daily new posts. They also designed an analysis tool to capture positive and negative sentiment towards stock market trends from the text content in each post. Then, they simulated the trend of stock markets using both the intensity of posts and the

sentiment within each post. By comparing with Chinese stock market main index in the same period, they claimed that the stock market trend created by their analysis method is relatively accurate.

Lin (2013) did statistical research on the relationship between public opinion from online forums and stock prices. He has four main findings. Firstly, forum discussions are more intense when stock prices are falling. By calculating the correlation between stock prices and forums' attention, Lin finds that the discussion intensity in online forums tends to be negatively correlated with stock prices, that is, compared with rising periods, online forum members pay more attention when stock prices are falling. Secondly, when the stock market is volatile, individual investors' behaviors in the market are more consistent; to some extent, there is a "herding effect" in investor behavior. Thirdly, investors' sentiment reflected from their posts in online forums has a predictive effect on stock prices and can help to explain fluctuations on stock yields. Lastly, the stock price predictions in online forums' messages are more accurate when the stock price is either rising or dropping by a significant margin.

In comparison with this paper, which emphasizes the impact on stock price only, three papers listed above went into the impact of online forum discussions on more aspects of stock trading, such as trading volume and stock yields. In comparison, this paper illustrates the impact of Internet chatter trends on specific stocks.

II. Post *WallStreetBets* era Literatures

After the GameStop short squeeze event, studies focused attention on *WallStreetBets* and its community can be defined as Post *WallStreetBets* era literatures. Although it is just half a year after the GameStop short squeeze event

started, due to its sensation, a rich literature has emerged that assesses its impact.

Jarrow and Li (2021) built mathematics models to value how forum traders perform under the influence of social media platforms. What relates to *WallStreetBets* in their study is that they divide traders into two groups: large traders and media group traders (*WallStreetBets* community). From the model they built, they demonstrate that the unionized media group traders can discipline large trader's interest to do short selling, which coincides with what happened in reality. As for the formation of the WallStreetBets community and the ability for it to achieve what other chatting groups can't achieve, Christian et al (2021) used social psychology theory to analyze this chaotic community. In their study, they described *WallStreetBets* as a high-risk style gambling casino instead of a professional finance-related discussion forum. The conclusion they present was that the nature of *WallStreetBets* is to bring legal online group gambling and excitement to its users --- mostly young men ---who are bored with normal, flat lives. Its popularity comes from fulfilling people's needs; and its giant power to shake the stock market comes from both users' unity and its popularity.

Instead of calling *WallStreetBets* community's activities as natural behaviors, in Chohan's paper (2021), he put forward an idea that the chain of short squeeze series was abetted by elite sympathizers. He also pointed out several structural socio economic issues as main reasons for *WallStreetBets* short squeeze to succeed. In his point of view, "so long as structural socio economic issues persist, such tactics will continue to generate mass appeal." (Chohan, 2021)

These studies discuss the nature of *WallStreetBets* short squeeze series with different understanding and from various angles. In this paper, more numerical and empirical results would be given to present the effect of short squeezes on specific stocks; beyond that, how *WallStreetBets* discussions affect stocks that they did not

short squeeze would be something new to add to the current literature.

CHAPTER 3&4: MAIN FINDINGS

The main findings of this paper shall be presented in two parts: the first part focuses on answering the question about how discussion trends on *WallStreetBets* affect stock price, or they affect each other. In order to answer this question, we conduct the Granger Causality test. The second part of the paper focuses on the specific and numerical effect of discussion trends on *WallStreetBets* on stock prices. In this part, regression analysis is the main tool.

CHAPTER 3: GRANGER CAUSALITY TEST

Section 1: Methods

Phase 1: Data Collection

In order to conduct the research, data collection is the very first and the most necessary step. Data used in this paper include, but are not limited to, daily stock open price, daily stock close price, daily stock mention times on *WallStreetBets*, stocks' short interest and stocks' days to cover from 01/15/2021 to 06/15/2021. Stock prices are collected from the NYSE on the Yahoo Finance website (https://finance.yahoo.com/quote). Short interest volume and days to cover are collected from a website called *marketbeat* (https://www.marketbeat.com/market-data)

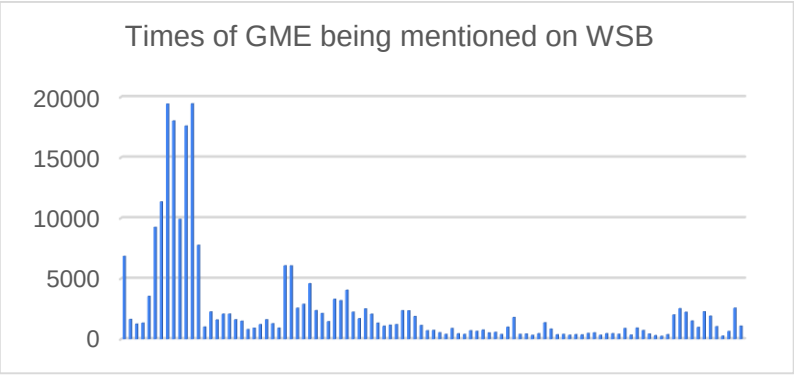
. Half of the *WallStreetBets* data are collected from a website called *quiverquant* (https://www.quiverquant.com/wallstreetbets/); the other half is scraped from *Reddit/WallStreetBets* by using R language:First download *WallStreetBets* daily discussion *URL*, then using rvest package to get plain text, then using a frequency function recommended by the data analyst Martin Geissmann to collect words'

counts or using text mining package in R instead. Data collection websites are credited in the reference and related code is attached in the appendix.

Phase 2: Performing Granger Causality Test on Stock Prices and *WallStreetBets* Discussion Data

Before I estimate the relationship between daily stock prices and *WallStreetBets* daily discussion trends, the question awaiting to answer is the relationship is one-directional or bidirectional; in other words, when *WallStreetBets* discussions and trends affect stock prices, stock price changes simultaneously affect *WallStreetBets* discussions or not. To respond to the question above, I chose to perform Granger Causality Test on 15 stocks' prices and their daily mentioned times in *WallStreetBets* discussions. Graph 3 is presented here to give a sense of how many times a stock's name would appear on *WallStreetBets* by using GME as an example.

Graph 3: GME mentioned times data



The Granger causality test is a statistical method of hypothesis testing, which tests whether a set of time series x is the cause of another set of time series y. It is based on the vector autoregressive model in regression analysis. Regression

analysis usually can only get the correlation between different variables at the same time; the autoregressive model can only get the correlation between the same variable before and after; but by using the Granger Causality test, through a series of tests in the autoregressive model, we can reveal the correlation between the time gap between different variables.

In order to do the Granger causality test, first, a vector auto-regression model is built between daily stock close price and daily stock's name, mentioned in *WallStreetBets* discussions with 4 lags. (A "lag" is a fixed amount of passing time, 4 lags is commonly used for monthly data and daily data). After the model is built, I list the question to test as whether stock prices 'cause' *WallStreetBets* discussions or the *WallStreetBets* discussions cause stock prices or both. Consider the following pair of regression to see the Granger Causality:

Stock Price t = C1 WallStreetBets discussions t-i + C2* WallStreetBets discussions t-j + u1t* *(1)*

*WallStreetBets discussions t = C3 * Stock Price t-i + C4 * Stock Price t-j + u2t* *(2)*

Here "Stock Price t" and "*WallStreetBets* discussion t" represent daily stock close price and daily stock's name mentioned times in *WallStreetBets* discussions on any given day t. All the "t-I" and "t-j" terms represent lag periods, and "u1t" as well as "u2t" are residuals. Thirdly, two sets of null hypothesis and alternative hypothesis are stated as follows:

Null hypothesis: lagged stock prices does not cause effect on daily stock's name mentioned times in WallStreetBets discussions

Alternative hypothesis: lagged stock prices causes effect on daily stock's name mentioned times in WallStreetBets discussions (3)

Null hypothesis: lagged daily stock's name mentioned times in WallStreetBets discussions does not cause effect on stock prices

Alternative hypothesis: lagged daily stock's name mentioned times in WallStreetBets discussions causes effect on stock prices (4)

Lastly, the Null hypothesis is either accepted or rejected based on the probability value from the Granger Causality tests.

Section 2: Results

Granger Causality Test Results

Vector auto-regression and Granger causality tests are done for all 15 stocks. Table 7 shows the vector auto-regression results for AMC. The vector autoregression only works as a pre-step for conducting the Granger causality test in our study. For this reason, Table 7 is just for presenting this mid-step and not for discussion of results. The vector autoregressions for other stocks are not shown in the paper.

Table 7: AMC Vector Auto-regression Results

VARIABLES	(1) StockClosePrice	(2) Mentioned Time in WSB
L.StockClosePrice	1.188***	218.3***
	(0.121)	(72.19)
L2.StockClosePrice	-0.194	85.26
	(0.253)	(150.7)

L3.StockClosePrice	0.340	-190.5
	(0.308)	(183.2)
L4.StockClosePrice	-0.391**	-28.20
	(0.199)	(118.7)
L.Mentioned Time in WSB	-0.000485**	0.328**
	(0.000221)	(0.131)
L2.Mentioned Time in WSB	0.000628***	0.123
	(0.000209)	(0.124)
L3.Mentioned Time in WSB	-0.000940	0.177
	(0.000643)	(0.383)
L4.Mentioned Time in WSB	0.000501	-0.103
	(0.000455)	(0.271)
Constant	0.587	-789.8**
	(0.555)	(330.7)
Observations	16	16

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Based on the Granger Causality test results, 15 stocks can be divided into three groups to deliver the results.

(1) GME AMC BB TLRY SNDL CLOV NIO AMD

The Granger causality test results for stocks in this group are shown in Table 8-Table 15.

Table 8: GME Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	47.872	4	0.000
StockClosePrice	ALL	47.872	4	0.000
MentionedTimesi~B	StockClosePrice	17.772	4	0.001
MentionedTimesi~B	ALL	17.772	4	0.001

For GME, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock

price and discussion times.

Table 9: AMC Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	10.189	4	0.037
StockClosePrice	ALL	10.189	4	0.037
MentionedTimesi~B	StockClosePrice	34.158	4	0.000
MentionedTimesi~B	ALL	34.158	4	0.000

For AMC, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 10: BB Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	18.87	4	0.001
StockClosePrice	ALL	18.87	4	0.001
MentionedTimesi~B	StockClosePrice	28.292	4	0.000
MentionedTimesi~B	ALL	28.292	4	0.000

For BB, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 11: TLRY Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	18.111	4	0.001
StockClosePrice	ALL	18.111	4	0.001
MentionedTimesi~B	StockClosePrice	9.5011	4	0.050
MentionedTimesi~B	ALL	9.5011	4	0.050

For TLRY, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 12: SNDL Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	12.219	4	0.016
StockClosePrice	ALL	12.219	4	0.016
MentionedTimesi~B	StockClosePrice	10.639	4	0.031
MentionedTimesi~B	ALL	10.639	4	0.031

For SNDL, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 13: CLOV Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	11.053	2	0.004
StockClosePrice	ALL	11.053	2	0.004
MentionedTimesi~B	StockClosePrice	18.524	2	0.000
MentionedTimesi~B	ALL	18.524	2	0.000

For CLOV, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 14: NIO Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	19.324	4	0.001
StockClosePrice	ALL	19.324	4	0.001
MentionedTimesi~B	StockClosePrice	15.819	4	0.003
MentionedTimesi~B	ALL	15.819	4	0.003

For NIO, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

Table 15: AMD Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	28.457	4	0.000
StockClosePrice	ALL	28.457	4	0.000
MentionedTimesi~B	StockClosePrice	12.326	4	0.015
MentionedTimesi~B	ALL	12.326	4	0.015

For AMD, The P-value for examining mentioned times on *WallStreetBets* on Stock close price is less than 0.05; The P-value for examining stock close price on mentioned times on *WallStreetBets* is less than 0.05 as well; suggesting that we reject the null hypothesis and confirm the bi-directional relationship between stock price and discussion times.

From Tables 8-15, all stocks in this group have probability value less than 0.05 for both stock price causes mentioned times in *WallStreetBets* and mentioned times in *WallStreetBets* causes stock price, suggesting rejecting both null hypothesis 1 and null hypothesis 2. To conclude, stock prices and daily mentioned times in *WallStreetBets* are proven to have influence over each other for stocks GME, AMC, BB, TLRY, SNDL, CLOV, NIO, and AMD.

GME, AMC, BB, TLRY, SNDL and CLOV are stocks that have been short squeezed by *WallStreetBets* trends; 6 out of 8 short squeeze stocks are proven to have a bidirectional relationship between stock price and *WallStreetBets* trend, suggesting the assumptions that changes on stock prices can in return encourage the discussion trends in *WallStreetBets* is on the right track. In discussion of non-

short squeeze stocks in this group, which are NIO and AMD, the bidirectional relationship between their share prices and discussion times on *WallStreetBets* is worth exploring more. NIO is a new Chinese electric car concept stock with medium-size market capital, having the shortest listing time among stocks being studied. For a new listed stock in a new industry, a great performance on its price can improve its popularity, attracting more discussion on *WallStreetBets*; with more discussions on *WallStreetBets*, NIO is known for more potential buyers and higher price. This is my speculation on the bidirectional relationship on NIO stock price and its *WallStreetBets* discussions. AMD is a chip-manufacturing oriented technology company. AMD chips are commonly used to do crypto mining. When *WallStreetBets* discusses AMD, it is generally mentioned together with popular crypto like Bitcoin, Ethereum and Litecoin. (Dogecoin is not included due to it not accounting for a large share in crypto mining) Therefore, the bidirectional relationship obtained from the Granger causality test for AMD may be biased due to the fact that some discussions mentioned AMD are actually talking about crypto mining. Besides stocks, Crypto is another hot topic on *WallStreetBets*.

(2) NOK BBBY TSLA WKHS

The Granger causality test results for stocks in this group are shown in Table 16-Table 19.

Table 16: NOK Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	3.0432	4	0.551
StockClosePrice	ALL	3.0432	4	0.551
MentionedTimesi~B	StockClosePrice	13.078	4	0.011
MentionedTimesi~B	ALL	13.078	4	0.011

For NOK, The P-value for examining mentioned times on *WallStreetBets* on stock close price is more than 0.05, suggesting that we can not reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is less than 0.05, suggesting that we can reject the null hypothesis that stock price does not cause discussion times and accept the alternative hypothesis that stock price does cause discussion times.

Table 17: BBBY Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	5.6234	4	0.229
StockClosePrice	ALL	5.6234	4	0.229
MentionedTimesi~B	StockClosePrice	10.721	4	0.030
MentionedTimesi~B	ALL	10.721	4	0.030

For BBBY, The P-value for examining mentioned times on *WallStreetBets* on stock close price is more than 0.05, suggesting that we can not reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is less than 0.05, suggesting that we can reject the null hypothesis that stock price does not cause discussion times and accept the alternative hypothesis that stock price does cause discussion times.

Table 18: TSLA Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	8.9409	4	0.063
StockClosePrice	ALL	8.9409	4	0.063
MentionedTimesi~B	StockClosePrice	22.818	4	0.000
MentionedTimesi~B	ALL	22.818	4	0.000

For TSLA, The P-value for examining mentioned times on *WallStreetBets* on stock close price is more than 0.05, suggesting that we can not reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is less than 0.05, suggesting that we can reject the null hypothesis that stock price does not cause discussion times and accept the alternative hypothesis that stock price does cause discussion times.

Table 19: WKHS Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	3.638	4	0.457
StockClosePrice	ALL	3.638	4	0.457
MentionedTimesi~B	StockClosePrice	14.667	4	0.005
MentionedTimesi~B	ALL	14.667	4	0.005

.

For WKHS, The P-value for examining mentioned times on *WallStreetBets* on stock close price is more than 0.05, suggesting that we can not reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price. The

P-value for examining stock close price on mentioned times on WallStreetBets is less than 0.05, suggesting that we can reject the null hypothesis that stock price does not cause discussion times and accept the alternative hypothesis that stock price does cause discussion times.

From Tables 16-19, stocks in this group, NOK, BBBY, TSLA and WKHS, have probability value less than 0.05 for lagged stock price causes mentioned times in *WallStreetBets* and probability more than 0.05 for lagged mentioned times in *WallStreetBets* causes stock price, suggesting that we reject the first null hypothesis and accept the second null hypothesis. To conclude, lagged stock price influences the stock's *WallStreetBets* discussion times; however, lagged *WallStreetBets* discussion times do not have influence over stock price.

Among these four stocks, BBBY, TSLA and WKHS are not *WallStreetBets* short squeeze targets. They vary from three different industries and have medium market capital size, large market capital size and small market capital size respectively. NOK had one short squeeze trend on *WallStreetBets*, but its price showed nearly no change during that short squeeze trend. The Granger causality test results for these four stocks are strong evidence to state that *WallStreetBets* discussions have a tiny influence on stocks that they do not try to short squeeze and even if they try to short squeeze a certain stock, the consequence may not meet their goals. Regressions are still conducted for these four stocks, but the results may not be very meaningful. On the flip side, the changes in these four stocks' price can shift their discussion times on *WallStreetBets*, coincides with the idea that stock prices can in return affect online forum trends and implies that instead of stocks with smooth price, *WallStreetBets* prefer to focus on stocks with fluctuated price.

(3) PLTR MVIS AAPL

The Granger causality test results for stocks in this group are shown in Tables 20-22.

Table 20: PLTR Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	115.79	4	0.000
StockClosePrice	ALL	115.79	4	0.000
MentionedTimesi~B	StockClosePrice	8.9117	4	0.063
MentionedTimesi~B	ALL	8.9117	4	0.063

For PLTR, The P-value for examining mentioned times on *WallStreetBets* on stock close price is less than 0.05, suggesting that we can reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price and accept the alternative hypothesis that discussion times does cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is more than 0.05, suggesting that we can not reject the null hypothesis that stock price does not cause discussion times.

Table 21: MVIS Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	24.095	4	0.000
StockClosePrice	ALL	24.095	4	0.000
MentionedTimesi~B	StockClosePrice	4.701	4	0.319
MentionedTimesi~B	ALL	4.701	4	0.319

For MVIS, The P-value for examining mentioned times on *WallStreetBets* on

stock close price is less than 0.05, suggesting that we can reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price and accept the alternative hypothesis that discussion times does cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is more than 0.05, suggesting that we can not reject the null hypothesis that stock price does not cause discussion times.

Table 22: AAPL Granger Causality Test Result

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
StockClosePrice	MentionedTimesi~B	16.361	4	0.003
StockClosePrice	ALL	16.361	4	0.003
MentionedTimesi~B	StockClosePrice	4.9058	4	0.297
MentionedTimesi~B	ALL	4.9058	4	0.297

For AAPL , The P-value for examining mentioned times on *WallStreetBets* on stock close price is less than 0.05, suggesting that we can reject the null hypothesis that mentioned times on *WallStreetBets* does not cause stock price and accept the alternative hypothesis that discussion times does cause stock price. The P-value for examining stock close price on mentioned times on WallStreetBets is more than 0.05, suggesting that we can not reject the null hypothesis that stock price does not cause discussion times.

From Tables 20-22, PLTR, MVIS and AAPL have probability value less than 0.05 for lagged mentioned times in *WallStreetBets* causes stock price and probability more than 0.05 for lagged stock price causes mentioned times in *WallStreetBets*, suggesting that we accept the first null hypothesis and reject the second null

hypothesis. To conclude, lagged stock price does not influence the stock's *WallStreetBets* discussion times; however, lagged *WallStreetBets* discussion times cause changes in stock price.

MVIS is the only short-squeezed stock whose share price has no influence over *WallStreetBets* discussions. To find the cause for the specialty of MVIS, I compare MVIS's short squeeze price raise percentage and short squeeze persisting days with GME, AMC, BB, TLRY, SNDL and CLOV. For short squeeze price raise percentage, MVIS is 45%, the average of the other six stocks is 340% (GME is over 900%) For short squeeze persisting days, MVIS is 4 days, the shortest of the other six is CLOV with 8 days. Based on the statistical data measures, I address the MVIS short squeeze as a semi-successful short squeeze. It shakes the stock market, but not as much as the other short squeezes. This may be the main reason for MVIS not showing the bidirectional relationship between its stock price and mentioned times in *WallStreetBets*. AAPL is the stock with the smoothest daily mentioned times change among the Top30 hottest stocks on *WallStreetBets.*(quiverquant) Its stock price does not affect its discussion trends on *WallStreetBets* may due to its status as the leading stock in technology industry, electronic products industry and software industry. For stocks like AAPL, companies' fundamentals are more worthy of discussion than short term price change. PLTR is a middle size software company that specializes in big data analytics. It does not have much specialty and runs normally during the last six months, I can't find any persuasive reasons to explain why its test result is in this group.

Section 3: Discussion

The relationships between stock price and stock's discussion trend on *WallStreetBets* for stocks studied are summarized in the following table:

Stock Code	Short Squeezed or Not	Price-Discussion Relationship
AMC	Yes	Affect each other
BB	Yes	Affect each other
GME	Yes	Affect each other
TLRY	Yes	Affect each other
SNDL	Yes	Affect each other
NOK	No	Price affect Discussion only
MVIS	Yes	Discussion affects Price only
CLOV	Yes	Affect each other
BBBY	No	Price affect Discussion only
NIO	No	Affect each other
WKHS	No	Price affect Discussion only
PLTR	No	Discussion affects Price only
TSLA	No	Price affect Discussion only
AAPL	No	Discussion affects Price only
AMD	No	Affect each other

For the majority of the short squeeze stocks, their stock price and discussion times on *WallStreetBets* display a bidirectional relationship: stock price causes change on discussion times and discussion times in return causes change on stock price as well. For stocks without short squeeze experience, the price-discussion relationship varies from stock to stock. NIO and AMD still show the bidirectional price-discussion relationship; for TSLA, WKHS and BBBY, their stock prices cause changes on *WallStreetBets* discussion times, but *WallStreetBets* discussions have no effect on their prices; for AAPL and PLTR, their *WallStreetBets* discussion times causes changes on stock prices, but changes on their stock prices have no effect on their

discussion times on *WallStreetBets.*

To conclude the finding in this part of the paper, there exists a bidirectional relationship between stock prices and discussion times on *WallStreetBets* for stocks that were short -squeezed by the *WallStreetBets* community before. As for stocks that have not been short squeezed before, that is no identical conclusion. For stocks in traditional manufacturing and vehicle industries, like BBBY, WKHS and TSLA, stock prices are more likely to have influence over *WallStreetBets* discussions. For stocks in software and high-technology industries, like AAPL and AMD, *WallStreetBets* discussion trends are more likely to have influence over stock price. For all 15 stocks being studied, all of them passed at least one-side the two-side Granger causality test on price causes discussion and discussion causes price, suggesting the two variables, discussion times and stock prices, are related.

CHAPTER 4: REGRESSION ANALYSIS

Performing Two Stage Least Squares Regressions for studying relationship between Stock Prices and Daily Discussion Times in *WallStreetBets*

Section 1: Methods

Phase 1: Performing Two Stage Least Squares Regressions and OLS regressions for studying relationship between Stock Prices and Daily Discussion Times in *WallStreetBets* for all 15 selected stocks by using daily data

Based on the bidirectional relationship between stock price and daily stock's name mentioned times in *WallStreetBets,* 15 stocks are divided into two groups to conduct regression. Majority of the stocks, 12 of them passed the test and 2SLS (two stage least squares) regressions were run to study the relationship between Stock Prices and Daily Discussion Times in *WallStreetBets* for these 12 stocks respectively. Due to the fact that both stock prices and discussion frequencies vary over a large range for stock to stock, instead of the normal form of data, we use natural log form of data to view relationships in percentage form for more accurate understanding. On the natural-log scale are directly interpretable as approximate proportional differences: with a coefficient of 0.05, a difference of 1 in x corresponds to an approximate 5% difference in y, and so forth.The log form on both sides provides us the elasticity between stock price and discussion frequency, which will be called price-discussion elasticity later.

The 2SLS regression model information is listed as below, taking AMC as an instance:

The base OLS regression:

$$Ln\ AMC\ Stock\ Price = \beta_0 + \beta_1\ Ln\ AMC\ Mentioned\ Times\ in\ WSB + \beta_2\ Short\ Interest + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

Add stock price again into the model as an instrumental variable on the endogenous variable Mentioned times in WSB:

$$Ln\ AMC\ Mentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ Stock\ Price + \epsilon_t *$$

Finalize the model:

$$Ln\ AMC\ Stock\ Price = \beta_0 + \beta_1\ (\ Ln\ AMC\ Mentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ Stock\ Price + \epsilon_t *)\ + \beta_2\ Short\ Interest\ + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

List of other stocks performing 2SLS regression model (by stock code): BB, GME, TLRY, SNDL, NOK, CLOV, BBBY, NIO, WKHS, TSLA, AMD.

For stocks PTLR, AAPL and MVIS, their stock prices are proven to have nearly no effect on discussion times, and OLS regression would be used instead of 2SLS regression.

Phase 2: Performing Two Stage Least Squares Regressions for studying relationship between Stock Prices and Daily Discussion Times in *WallStreetBets* by using 15 days period average data for short squeeze stocks

For stocks with *WallStreetBets*-related short squeeze experience, one additional 2SLS regression model is designed to better evaluate the effect of short squeeze and the effect of the short squeeze. Firstly, using non overlapped 15 days as a time scale instead of daily data; to achieve that, take 15-days average value of stock prices and stock discussion times. Secondly, using days to cover to represent the potential of being short squeeze, and making days to cover as another instrumental variable of stock's discussion times, since a high days to cover measurement can signal a potential short squeeze and drive more discussions on *WallStreetBets*.

The 15-day scales 2SLS regression model information is listed as below, taking AMC as an instance:

The base OLS regression:

$$Ln\ AMC\ Average\ Stock\ Price = \beta_0 + \beta_1\ Ln\ AMC\ Average\ Mentioned\ Times\ in\ WSB + \beta_2\ AMC\ Short\ Interest\ + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

Add average stock price and days to cover as two instrumental variables on endogenous variable average Mentioned times in WSB:

$$Ln\ \ AMC\ AverageMentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ Average\ Stock\ Price + \beta_6\ Ln\ AMC\ Days\ to\ Cover + \epsilon_t *$$

Finalize the model:

$$Ln\ AMC\ Average\ Stock\ Price = \beta_0 + \beta_1\ (\ Ln\ AMC\ Average\ Mentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ \ Average\ Stock\ Price + \beta_6\ Ln\ AMC\ Days\ to\ Cover + \epsilon_t\ *) + \beta_2\ AMC\ Short\ Interest\ + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

List of other stocks performing 15-days scale 2SLS regression model (by stock code): BB, GME, TLRY, SNDL, NOK, MVIS.

Phase 3: Performing Two Stage Least Squares Regressions for studying relationship between Stock Prices and Discussions with specific goals in *WallStreetBets*

After the first and second rounds of regressions, the third round of 2SLS regressions were conducted on GameStop, AMC and Blackberry, the three stocks with highest popularity and the most mentioned times on *WallStreetBets* over the last 6 months. For the third round of regression, the data on stock's name daily mentioned times are replaced with the daily times of both stock's name and the phrase "to the moon" being mentioned together. Taking a quick detour here to explain the meaning of the phrase "to the moon": In *WallStreetBets* community's culture, when they express the emotion that wish a certain stock price to go up, they post certain stock's name plus "to the moon" as the slogan. For instance, On Feb 24th, "GME to the Moon" was posted 4056 times, indicating the forum's trend that wanted GameStop price to raise.

The third round of regression is used for comparing with the results from the first round to assess the differences of the impact of discussion trends with purpose on stock price and the impact of discussions without knowing purpose on stock price. Results from the first round regression on GameStop, AMC and Blackberry represent the impact of discussions without knowing purpose on stock prices. Results from the third round regression on GameStop, AMC and Blackberry represent the impact of proposed online trends on stock price.

The 2SLS regression model information is listed as below, taking AMC as an instance:

The base OLS regression:

$$Ln\ AMC\ Stock\ Price = \beta_0 + \beta_1\ Ln\ Both\ AMC\ \ and\ To\ the\ Moon\ Mentioned\ Times\ in\ WSB + \beta_2\ Short\ Interest\ + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

Add stock price again into the model as an instrumental variable on endogenous variable Both AMC and the phrase “To the Moon” Mentioned times in WSB:

$$Ln\ Both\ AMC\ and\ To\ the\ Moon\ Mentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ Stock\ Price + \epsilon_t *$$

Finalize the model:

$$Ln\ AMC\ Stock\ Price = \beta_0 + \beta_1\ (\ Ln\ Both\ AMC\ and\ To\ the\ Moon\ Mentioned\ Times\ in\ WSB = \beta_4 + \beta_5\ Ln\ AMC\ Stock\ Price + \epsilon_t *)\ + \beta_2\ Short\ Interest\ + \beta_3\ S\&P500\ D2_t + \epsilon_t$$

List of other stocks performing third round regression (by stock code): BB, GME.

Phase 4: Performing Additional OLS Regressions

After the main tests and regressions are done, two more simple OLS regressions are conducted for research interests and additional study.

(1) Study the relationship between *WallStreetBets* daily discussion times and daily stock price range, for examining the impact of *WallStreetBets* discussion on daily stock price fluctuation.

The OLS regression, taking AMC as an instance:

$$Ln\ AMC\ Daily\ Stock\ Price\ Range = \beta_0 + \beta_1\ Ln\ AMC\ Daily\ Mentioned\ Times\ in\ WSB + \epsilon_t$$

(2) Study the relationship between short interest and *WallStreetBets* daily discussion times, to examine the impact of stock's short volume change on its popularity on *WallStreetBets.* To answer the question: Would *WallStreetBets* choose their short squeeze target by looking at the volume of short interest.

The OLS regression, taking AMC as an instance:

$$Ln\ AMC\ Daily\ Mentioned\ Times\ in\ WSB = \beta_0 + \beta_1\ Ln\ Short\ Interest + \epsilon_t$$

The codes for conducting regressions analysis listed above are included in the appendix.

Section 2: Results

Due to the fact that regressions are conducted in a natural log-log form which gives us percentage change, the results we collected from regressions can be defined as stock price-discussion elasticity. The elasticity here refers to the degree to which changes on certain stock's name are mentioned on *WallStreetBets* posts and discussions in response to that stock's price in percentage form. By evaluating and comparing this elasticity, how much stock prices are influenced from *WallStreetBets* discussions can be examined.

Phase 1: Performing Two Stage Least Squares Regressions and OLS regressions for studying relationship between Stock Prices and Daily Discussion Times in *WallStreetBets* for all 15 selected stocks by using daily

data

For the Two stage least squares regression results, firstly the results for AMC, GME, BB, TLRY, SNDL and CLOV are presented in Table 23.

Table 23: 2SLS Regression Results for AMC, GME, BB, TLRY, SNDL, CLOV

VARIABLES	AMC lnprice	BB lnprice	GME lnprice	TLRY lnprice	SNDL lnprice	CLOV lnprice
lntrend	0.491***	0.169***	2.620***	0.272***	0.299***	0.280***
	(0.0730)	(0.0135)	(0.979)	(0.0417)	(0.0589)	(0.0436)
SP500	0.00378***	-3.37e-05	0.0138***	-0.000996***	-0.000233	-0.00243***
	(0.000777)	(0.000117)	(0.00466)	(0.000193)	(0.000278)	(0.000460)
ShortInterest	-0.000154	0.00105**	-0.000540	0.000551**	-8.15e-05	-0.000516
	(0.000185)	(0.000399)	(0.000340)	(0.000251)	(0.00106)	(0.000471)
Constant	-15.72***	1.156**	-68.33***	5.497***	-0.276	11.27***
	(3.355)	(0.479)	(25.01)	(0.814)	(1.145)	(1.683)
Observations	101	101	101	101	101	95
R-squared	0.409	0.551				

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

AMC, GME, BB, TLRY, SNDL and CLOV are stocks that are short-squeezed by *WallSreetBets* and are proven by the Granger causality test that their share prices and discussion times on *WallStreetBets* affect each other. From the table 23, keeping other variables constant, for 1% increase in mentioned times on *WallStreetBets*, AMC stock price goes up by 0.49%, BB stock price goes up by 0.17%, GME stock price goes up by 2.62%, TLRY stock price goes up by 0.27%, SNDL stock price goes up by0.30%, and CLOV stock price goes up by 0.28%; all these results are statistically significant ($p<0.01$). They all show a relatively strong relationship between stock price and discussion trends; GME has the strongest price-discussion relationship, followed by AMC, and then the other four are on a par with each other. This order completely

coincides with their short squeeze times count; GME has been short squeezed 3 times, AMC has been short squeezed twice and the others have once short squeeze experience. Overall, *WallStreetBets* discussion trends have considerable effect on these six stocks. Usually when a stock is under the *WallStreetBets* short squeeze trend, its name's mentioned times increase by more than 500% in one day. Even for BlackBerry, which has the lowest coefficient among these six stocks, 500% increases in mentioned times would increase its price by 84.5%, a substantial amount.

The results for NIO and AMD are presented in Table 24

Table 24: 2SLS Regression Results for NIO and AMD

VARIABLES	NIO lnprice	AMD lnprice
lntrend	3.531	0.191***
	(13.91)	(0.0410)
SP500	0.0108	0.000155
	(0.0355)	(9.83e-05)
ShortInterest	1.848	0.116***
	(4.602)	(0.0299)
Constant	-65.32	2.048***
	(218.2)	(0.669)
Observations	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NIO and AMD are two stocks that passed the granger causality test on stock price and discussion times in *WallStreetBets*, but were not short-squeezed by *WallStreetBets* before. Although they have characteristics mentioned above in common, the regression results for NIO and AMD are quite different and therefore would be discussed separately. For NIO, Keeping all other variables constant, 1%

increase in mentioned times on *WallStreetBets* would increase its share price by 3.53%. Surprisingly, this coefficient is even larger than GME and AMC, but by looking at the P-value, this result is not statistically significant. Due to its statistically insignificant results, the result for NIO can only be concluded as its stock price and its discussion times on *WallStreetBets* are positively correlated. For AMD, Keeping all other variables constant, 1% increase in mentioned times on *WallStreetBets* would increase its share price by 0.19%. This result is similar to BlackBerry's; the dissimilarity is that AMD has not been short-squeezed by *WallStreetBets* before. This gives us a new observation: certain non-short squeeze target stocks that passed the Granger causality test still have relatively large price—discussion trend elasticity.

The results for NOK, BBBY, TSLA and WKHS are shown in Table 25.

Table 25: 2SLS Regression Results for NOK, BBBY, TSLA and WKHS

VARIABLES	NOK lnprice	BBBY lnprice	TSLA lnprice	WKHS lnprice
lntrend	0.007	0.118***	-1.788	1.313
	(0.0054)	(0.0151)	(1.693)	(0.999)
SP500	0.00265	0.000162	-0.00277	-0.00119
	(0.00181)	(0.000118)	(0.00291)	(0.00254)
ShortInterest	-0.00589	-0.000455***	0.00771	0.000408
	(0.00579)	(9.65e-05)	(0.0192)	(0.00295)
Constant	-10.14	2.824***	27.80	4.124
	(8.050)	(0.491)	(21.57)	(11.28)
Observations	101	75	101	99

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From the Granger Causality test, these four stocks' discussion trends on *WallStreetBets* are proven to have no influence over their stock price. The regression results from the tale above agrees with the Granger Causality test; according to the P-

value, the results for NOK, TSLA and WKHS are not statistically significant, suggesting *WallStreetBets* discussions should not be taken into consideration when we analyze stock price changes for NOK, TSLA and WKHS. BBBY is the exception here. Keeping other variables constant, a 1% increase in BBBY mentioned times on *WallStreetBets* would increase BBBY stock price by 0.12%, and this result is highly statistically significant (P<0.01). The odd thing here is that by the Granger causality test, BBBY mentioned times on *WallStreetBets* should not have influence over its price. My explanation is the Granger causality test adds lags into the dataset and tests on a time series basis and the regression evaluates coefficients on a daily-to-daily basis. For BBBY, to be apart from other stocks that have statistically significant results, the prediction of value of stock price based on discussion times only performs better than prediction of value of stock price based on both discussion times and the past value of stock price.

The results for PLTR, MVIS, and AAPL are shown in Table 26

Table 26: OLS Regression Results for PLTR, MVIS and AAPL

	PLTR	MVIS	AAPL
VARIABLES	lnprice	lnprice	lnprice
lntrend	-0.0146	0.0872***	0.0295***
	(0.0156)	(0.0224)	(0.00841)
SP500	-0.000442***	-0.000435	7.71e-05**
	(8.94e-05)	(0.000278)	(3.50e-05)
ShortInterest	0.181***	0.000936***	-0.00304
	(0.0261)	(0.000279)	(0.00287)
Constant	4.787***	3.714***	4.435***
	(0.425)	(1.008)	(0.163)
Observations	101	96	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Instead of 2SLS regression, OLS regressions are run for these three stocks due to the fact that their results from the Granger causality test declaring that discussion trends do not cause changes in stock prices. From the table above, the regression result for PLTR shows a negative relationship between its stock price and discussion times, but it is not statistically significant and can be neglected. For AAPL, Keeping other variables constant, 1% increase in AAPL mentioned times on *WallStreetBets* would increase its stock price by 0.03%, and this result is statistically significant. This is a signal revealing *WallStreetBets* discussion trends have the ability to influence industry leading stocks like AAPL, although the influence is tiny. For MVIS, a 1% increase on discussion times on *WallStreetBets* would bring a 0.09% increase on its stock price, and the result is statistically significant as well. MVIS is one of the short squeeze stocks; its short squeeze process is defined as semi-successful in the above texts. Here Table 27 presents the price-discussion times elasticity from high to low for all short squeeze stocks.

Table 27: Price-Discussion times Elasticity for Short Squeeze Stocks

Stock Code	Price-Discussion times Elasticity
GME	2.62
AMC	0.49
SNDL	0.30
CLOV	0.28
TLRY	0.27
MVIS	0.09
NOK	0.07

From this table, it is clear why MVIS short squeeze is defined as semi-

successful, NOK short squeeze is defined as not successful, and the other short squeezes are successful.

Phase 2 :Performing Two Stage Least Squares Regressions for studying relationship between Stock Prices and Daily Discussion Times in *WallStreetBets* by using 15 days period average data for short squeeze stocks

After providing the results of analyzing daily data, in this section, data for short squeeze stocks are collected in a 15-days scale and are conducted in 2SLS regression. There are three reasons to examine again by using 15-days scale data. 1. Compared with daily data, data on a 15-days scale does a better job on representing trends over a certain period; 2. Discussion times on marker close days are ignored when we use daily data; 3. Usually the short squeeze trend ends within one 15-days period. Due to the fact of that, data under the short squeeze effect would only take up one, mostly two observations in a 15-days scale dataset. The results are presented in Table 28 and Table 29.

Table 28: 2SLS Regression Results for AMC, GME, and BB

VARIABLES	AMC lnprice	GME lnprice	BB lnprice
lntrend	0.565*	2.149*	0.191***
	(0.287)	(1.921)	(0.1605)
SP500	0.00328*	0.00585*	0.000261
	(0.00223)	(0.00281)	(0.000315)
ShortInterest	-4.88e-05		
	(0.000457)		
ShortInterest		0.000245	-0.000712
		(0.000241)	(0.00106)
Constant	-15.74	-26.15	0.708
	(10.61)	(16.29)	(1.231)

Observations	11	11	11
R-squared	0.619	0.212	0.709

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 29: 2SLS Regression Results for TLRY, SNDL, NOK, MVIS.

VARIABLES	TLRY lnprice	SNDL lnprice	NOK lnprice	MVIS lnprice
lntrend	0.163	-0.761	0.0112	0.125
	(0.0971)	(1.505)	(0.0245)	(0.0788)
SP500	-0.000594**	-0.00505	0.000684***	-0.000417***
	(0.000316)	(0.00754)	(0.000126)	(0.000225)
ShortInterest	0.000833	0.0130	0.000670	0.0049
	(0.000551)	(0.0171)	(0.000560)	(0.00342)
Constant	4.136**	21.31	-1.437**	2.46
	(1.420)	(33.94)	(0.569)	(1.557)
Observations	12	12	12	12
R-squared	0.658		0.831	0.755

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

By what tables 28 and 29 provided, an interesting divergence could be found within these short squeezed stocks: For AMC, GME and BB, their results are still statistically significant and the AMC and the BB price-discussion times elasticity are even higher; For TLRY, SNDL, NOK and MVIS, their results are not statistically significant, suggesting mentioned times on *WallStreetBets* should not be considered as a cause for stock price any more. The main cause for this divergence is the differences in the numbers of short squeeze lasting days. AMC, GME and BB short squeezes last several 15-days periods, especially for AMC and GME, they both underwent more than one short squeeze. In comparison, TLRY, SNDL, NOK and MVIS short squeezes end within a 15-day period. The conclusion drawn from here is that even for short squeezed stocks, *WallStreetBets* discussions outside of their short

squeeze period are mostly random noises, showing no influence on their stock price

Phase 3: Performing Two Stage Least Squares Regressions for studying relationship between Stock Prices and Discussions with specific goals in *WallStreetBets* for AMC, GME and BB

In this section, in order to better examine the effect of trends, not only discussions, on stock prices, the regressions are conducted between stock prices and number of posts that contain specific goals, such as “AMC price to the moon” and “GME price to the moon”. The results of regression on AMC, GME, and BB are presented in Table 30.

Table 30: Phase 2.3 Regression Results for AMC, GME, BB

VARIABLES	AMC lnprice	GME lnprice	BB lnprice
lntrend	0.572***	2.842***	0.233***
	(0.3923)	(1.997)	(0.0191)
SP500	0.00599***	0.0104***	-7.49e-05
	(0.00141)	(0.00316)	(0.000156)
ShortInterest	-0.000439	-0.000672*	0.00146***
	(0.000288)	(0.000352)	(0.000549)
Constant	-23.41***	-43.52***	1.669***
	(5.788)	(14.78)	(0.587)
Observations	100	101	100

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From the table 30, the price-trend elasticity for AMC, GME and BB are 0.57, 1.44 and 0.23 respectively, and the results are all statistically significant. Compared with the table in phase 2.1 presenting the results for AMC, GME and BB, the coefficients

for all three stocks increase by a moderate to large amount. The increase in coefficient signals that discussion trends with specific goals have more influence over stock price than discussions without knowing the purpose.

Phase 4: Performing Additional OLS Regressions

For better understanding the influence of *WallStreetBets* discussions and for personal research interests, two more regression analyses are implemented in this section. First one is studying the relationship between *WallStreetBets* daily discussion times and daily stock price range, for examining the impact of *WallStreetBets* discussion on daily stock price fluctuation. Second one is Studying the relationship between short interest and *WallStreetBets* daily discussion times, for examining the impact of stock's short volume change on its popularity on *WallStreetBets.* The first result is presented in table 31.

Table 31: Results of *WallStreetBets* Discussion on Daily Stock Price Range for AMC, GME, SNDL, MVIS, TSLA and AAPL

VARIABLES	AMC lnpricerange	GME lnpricerange	SNDL lnpricerange	MVIS lnpricerange	TSLA lnpricerange	AAPL lnpricerange
lntrend	0.183***	0.566***	0.409***	0.240***	0.573***	0.395***
	(0.0661)	(0.0657)	(0.0307)	(0.0339)	(0.0939)	(0.0581)
Constant	-0.881**	-0.966**	-3.845***	-0.306**	0.105	-1.060***
	(0.419)	(0.475)	(0.138)	(0.150)	(0.546)	(0.299)
Observations	101	101	101	96	101	100
R-squared	0.072	0.429	0.642	0.347	0.274	0.320

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From the table 31, 1% increase in discussion times on *WallStreetBets* increases daily price fluctuation range by 0.18% for AMC, 0.57% for GME, 0.41% for SNDL, 0.24% for MVIS, 0.57% for TSLA and 0.40% for AAPL; all the results are statistically significant. Among these stocks, AMC, GME and SNDL were short squeezed with remarkable effect, MVIS was short squeezed with slight effect, TSLA and AAPL were not short squeezed. No matter being short squeezed or not and no matter the final effect of the short squeeze, *WallStreetBets* discussions triggers stocks' daily price fluctuations.

The second result is presented in Table 32.

Table 32: Results of stock's short volume change on its popularity on *WallStreetBets* for AMC, GME, SNDL, MVIS, TSLA and AAPL

VARIABLES	AMC lntrend	GME lntrend	SNDL lntrend	MVIS lntrend	TSLA lntrend	AAPL lntrend
lnShortInterest	0.284	0.276	1.437*	1.356	0.809***	-1.240**
	(0.222)	(0.232)	(0.320)	(0.289)	(0.200)	(0.518)
Constant	4.319***	5.082***	-3.297*	-3.716**	3.032***	8.325***
	(1.450)	(1.755)	(1.685)	(1.693)	(0.684)	(1.345)
Observations	101	101	101	96	101	101
R-squared	0.016	0.014	0.169	0.190	0.142	0.055

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results for AMC, GME, SNDL and MVIS are either statistically insignificant or statistically significant at the low degree (0.05<p<0.1). Although the results for TSLA and AAPL are statistically significant, 1% increase in short interest volume increases

TSLA's discussion on *WallStreetBets* by 0.81% and decreases AAPL's discussion on *WallStreetBets* by 1.24%, the R-squares for the regression models are too low. In conclusion, the volume of short interest has no effect on a certain stock's popularity, which is represented by the times of being mentioned on *WallStreetBets.*

Section 3: Discussion

Section 3.1: Discussion for each stock selected

To continue, this part discusses the effect of *WallStreetBets* discussion times on stock price for each stock based on our regression results. The price-discussion elasticity drawn from regressions are ranked from high to low for 15 stocks in the following table:

Stock Code	Short Squeezed or Not	Price-Discussion Elasticity
GME	Yes	2.62
AMC	Yes	0.49
SNDL	Yes	0.3
CLOV	Yes	0.28
TLRY	Yes	0.27
AMD	No	0.19
BB	Yes	0.17
BBBY	No	0.12
MVIS	Yes	0.09
AAPL	No	0.03
NIO	No	Statistically insignificant
NOK	Yes	Statistically insignificant
PLTR	No	Statistically insignificant
TSLA	No	Statistically insignificant
WKHS	No	Statistically insignificant

10 out of 15 stocks' prices are proven to be influenced by *WallStreeetBets* discussions. For these 10 stocks, the price-discussion elasticity for short squeeze

stocks are significantly higher than stocks without the short squeeze experience. Two outliers are AMD and MVIS, comparing with other stocks with no short squeeze experience, AMD has unexpected high elasticity because of its connection with crypto market, another hot topic on *WallStreeetBets*; comparing with other stocks with short squeeze experience, due to MVIS short squeeze has relatively shorter lasting period and smaller scale, MVIS has lower elasticity. Overall, *WallStreetBets* discussions have influence on 67% of the stocks in the research list: 87.5% of the short squeeze stocks and 43% of the non-short squeeze stocks., suggesting that for a *WallStreetBets* short squeeze target stock, *WallStreetBets* discussions most likely affects its stock price; for a random stock chosen, *WallStreetBets* discussion has nearly 50% chance affects its stock price.

Section 3.2: Deeper Research on Short Squeeze Stocks

Stock Code	Original Results	Regression Model 2 Results	Regression Model 3 Results
GME	2.62	2.15	2.84
AMC	0.49	0.57	0.57
BB	0.17	0.2	0.23
TLRY	0.27	Statistically insignificant	Not Conducted
SNDL	0.3	Statistically insignificant	Not Conducted
MVIS	0.09	Statistically insignificant	Not Conducted
NOK	Statically insignificant	Statistically insignificant	Not Conducted

For further examining the price-discussion relationship for short squeezed stocks, regressions using 15 days as time scale instead of 1 day (regression model 2) and regressions between stock prices and number of posts that contain specific goals, such as “AMC price to the moon” and “GME price to the moon” (regression

model 3) were conducted for selected stocks and the results along with the previous results are presented in the table below.

When data are collected, averaged and analyzed per 15 days instead of per day, the regression results for short squeezed stocks are divided into two groups: 1. TLRY, SNDL and MVIS have only one short squeeze trend during the last six months and their short squeeze lasting time was relatively short. When data are collected and averaged on a 15-days basis, the short squeeze-related price change and discussion trends change are shortened within one observation. Therefore, compared with the original daily basis regression, results from this regression model represent more on how discussion trends affect stock price with less bias from stock price changes and discussion data changes due to short squeeze. In this scenario, the results for TLRY, SNDL and MVIS turn out to become statistically insignificant, revealing that the statistically significant results from original regression are mostly due to short squeeze periods' data. After minimizing the effect of short squeeze, *WallStreetBets* discussions failed to continue influencing TLRY, SNDL and MVIS's stock prices. 2. For GME, AMC and BB, it is a totally different story. GME's price-discussion elasticity keeps nearly unchanged, AMC and BB's price-discussion elasticity even increase for a substantial amount. Comparing with TLRY, SNDL, and MVIS, GME, AMC, and BB have much longer short squeeze lasting time and multiple short squeeze counts; in the 15-days basis regression model, their short squeeze related price changes and discussion trends changes cross over several observations. Short squeeze influence is not reduced, but enhanced for these three stocks; thus result in the different results from TLRY, SNDL, and MVIS. This is in accordance with the speculation above: for short squeezed stocks, the relationship between their stock prices and stock mentioned times on *WallStreetBets* are mostly caused by the short squeeze activity, not stocks themselves.

As the fourth column named “Regression model 3 results” in the table above presented, after we run regression using discussions with specific short squeeze goals instead of discussions without knowing the purpose as independent variable, the price-discussion elasticity increase by an attention-getting amount for GME, AMC and BB. This finding suggests that short squeeze-oriented discussions trends cause more changes on stock prices than just discussions, verifying the conclusion above that short squeeze activities are the main causes of the relationship between stock prices and *WallStreetBets* discussions for short squeezed stocks.

Section 3.3: Additional Findings

For all stocks selected, increases in mentioned times on *WallStreetBets* are proven to increase their daily stock price ranges, meaning that discussion trends can cause stock price fluctuations. These results are statistically significant for no matter short squeezed stocks or non-short squeezed stocks. This brings us a new sight: even for stocks that have proven before that *WallStreetBets* discussions do not influence their price changes, *WallStreetBets* discussions actually influence their stock prices within a trading day. If a regression model is built by using hourly stock price data and hourly *WallStreetBets* data, a clear and distinct linear relationship between hourly stock price and hourly *WallStreetBets* discussion data may be found. I will leave this question open to future studies.

Another additional finding is that the volume of short interest does not cause *WallStreetBets* discussion trends to change. This finding helps to prove the idea that the majority of *WallStreetBets* users are truly profane and juvenile. When they choose their short squeeze targets, they rely on personal preference more than numerical and fundamental analysis. In the meanwhile, the reality of how the

WallStreetBets community chose their short squeeze target remains questionable. There are no obvious similarities on stocks they have already short squeezed. From my personal observation and online sayings, there are a few community leaders organizing short squeeze trends. The most famous leader's user ID is "u/DeepFuckingValue". This IS posted seditious contents on short squeezing GME, AMC, SNDL, BB, TLRY and most recently CLOV. He did not involve the short squeeze on MVIS. Does this cause MVIS short squeeze being not as successful as the others? I would say this may be one of the reasons, but the causal relationship still lacks verification.

CHAPTER 5: CONCLUSION

This paper makes a contribution through its examination and analysis of the relationship between stocks' prices and their discussion trends on *WallStreetBets* in the past six months as well as following *WallStreetBets*-related implications and discoveries. Results of this paper describes 15 stocks' causal relationship with *WallStreetBets* discussions, provides numerical information and can be used as reference for future stock price—online forum trends studies and short squeeze studies.

The Granger causality tests in this paper have examined 15 stocks on their bidirectional relationships between stock prices and stocks' mentioned times on *WallStreetBets.* The majority of the stocks that have been short-squeezed by *WallStreetBets* in the past six months have passed the Granger causality tests, suggesting that their stock prices and discussion times on *WallStreetBets* affect each other simultaneously. For stocks that haven't been short squeezed during the past six months, the test results differ by industry. The test results for stocks in the traditional manufacturing and vehicle industry show that stock prices cause changes

in discussion times on *WallStreetBets*, but not the other way around. The test results for stocks in the software and high-tech industry show that discussion times on *WallStreetBets* cause changes in stock price, but not the other way around. This Granger causality test results on Industry basis are drawn from the 15 stocks studied only, if more than 15 stocks can be tested, this result may be more reliable or it may change.

The first round of regressions is conducted between the log form of stock prices and the log form of mentioned times on *WallStreetBets*. The results are given as price-discussion elasticity. All the stocks with short squeeze experience have a relatively high price-discussion elasticity, except for NOK, due to the fact that the short squeeze on NOK failed. Three out of seven stocks without short squeeze experience have a relatively low price-discussion elasticity; the other four stocks' results are statistically insignificant. The first round regressions tell us that discussion trends on *WallStreetBets* have a larger influence on stocks with short squeeze experience than stocks that haven't been short squeezed before. Additionally, for some stocks, the changes on their stock prices are not related to *WallStreetBets* discussions.

The second round of regressions are conducted on short squeezed stocks. By comparing the different results on stocks with long-lasting short squeeze and stocks with short-lasting short squeeze, we conclude that for short squeezed stocks, the influence of *WallStreetBets* discussions on stock prices mostly come from the short squeeze periods and short squeeze trend. After minimizing the short squeeze effect, their price-discussion elasticity tends to lean closer to non-short squeezed stocks.

The third round of regression further confirms the conclusion above. Instead of using discussions without knowing the purpose as an independent variable,

discussions with specific short squeeze goals are used to run the regression. The price-discussion elasticity becomes larger for the majority of the stocks selected, indicating that discussion trends with specific goals have a larger effect on stock prices than discussions without knowing the aim.

The results from the final round of regressions show that changes in discussion times on *WallStreetBets* are closely related with daily stock price fluctuations for all stocks chosen regardless of being short-squeezed or not. This finding opens a new door for this topic: *WallStreetBets* discussions affect stock prices within a trading day. Hourly discussion trends and changes stand a good chance of influencing the hourly stock price. Future research can set to and study on this possibility.

To sum up, main conclusions for this paper are: 1.The relationship between stock prices and discussion trends is not a linear causal relationship. For more than 70% of the stocks studied, stock prices in return shift discussion trends. 2. For stocks that underwent short squeeze during the past six months, *WallStreetBets* discussion trends have moderate to large effect on stock prices; for stocks without short squeeze experience during the past six months, how *WallStreetBets* discussions influence stocks' prices vary from stock to stock and should be discussed case by case.

The limitations of the paper are: 1. Only 15 stocks are studied; more stocks added into research can bring more convincing and more general conclusions. 2. The models used for stock prices are too simple. There exists omitted variables and selection bias problems. 3. Hourly data on stock prices and stock discussions on *WallStreetBets* are not collected and analyzed. From the third round regression, studying on hourly data may be the correct direction to make conclusions for stocks regardless of being short squeezed or not.

Despite having limitations listed above, this paper could support future studies on *WallStreetBets* and short squeezes. Compared with other influences the Internet brought to the stock market, online forum discussions and trends are relatively lack of attentions. 2021 short squeeze series would change people's traditional opinions on online forums. In Doblec's article, he addressed *WallStreetBets* as an element that can permanently disrupt the stock market. (2021) As this community is continue growing larger, I agree with Doblec and believe *WallStreetBets* has this potency. I am waiting, we are waiting, our world is waiting for more detailed *WallStreetBets* research, to reveal the inherent nature of this Internet monster, which is made of fanaticism and desire.

REFERENCE

AMC Entertainment Holdings, Inc. (AMC) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/AMC?p=AMC

Antweiler Werner, Frank Murray. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, Wiley for the American Finance Association Jun., 2004, Vol. 59, No. 3 (Jun., 2004), pp. 1259-1294

Apple Inc. (AAPL) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/AAPL?p=AAPL&.tsrc=fin-srch

BlackBerry Limited (BB**)** Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/BB?p=BB&.tsrc=fin-srch

Bloombergquint. GameStop Mania Goes Global as Retail Traders Gang Up on Shorts Posted: 2021-01-28

Bogan, Vicki. Stock Market Participation and the Internet. *The Journal of Financial and Quantitative Analysis* , Mar., 2008, Vol. 43, No. 1 (Mar., 2008), pp. 191-211

Chohan W. Usman, Counter Hegemonic Finance: The Gamestop Short Squeeze. UNSW Business School. 2021.

Christian J. Boylston, Beatriz E. Palacios, Plamen T. Tassev. WallStreetBets: Positions or Ban. Georgia Institute of Technology, 2021.

Dolbec, Pierre-Yann. “WallStreetBets” is disrupting financial markets — possibly permanently. Concordia University. 2021

GameStop Corp. (GME) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/GME?p=GME&.tsrc=fin-srch

Jarrow Robert, Li Siguang. Media Trading Groups and Short Selling Manipulation: Are Media Groups Efficiency Enhancing or Reducing? Cornell University. 2021.

Lin Bingshan. Statistical Research on the Impact of Internet Public Opinion on Stock Prices Based on Investor Sentiment. Southwestern University of Finance and Economics, 2013.

MicroVision, Inc. (MVIS) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/MVIS?p=MVIS&.tsrc=fin-srch

NIO Inc. (NIO) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/NIO?p=NIO&.tsrc=fin-srch

N Tomić, V Todorović, B Čakajac The potential effects of cryptocurrencies on monetary policy. *The European Journal of Applied Economics*, volume 17, issue 1, p. 37 – 48. 2020

Tesla, Inc. (TSLA) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/TSLA?p=TSLA&.tsrc=fin-srch

Workhorse Group Inc. (WKHS) Stock Data, 2021. Retrieved from https://www.marketbeat.com/stocks/NYSE; https://finance.yahoo.com/quote/WKHS?p=WKHS&.tsrc=fin-srch

Wu Jing, Chen Yi-xiang, Liu Daoming. Internet-forum-based Stock Market Analysis Method. *Shanghai Key Laboratory of Trustworthy Computing*, East China Normal University, Shanghai 200062, 2012.

Xu Yongxin, Xuan Yuhao, Zheng Gaoping. Internet searching and stock price crash risk: Evidence from a quasi-natural experiment. *Journal of Financial Economics* Volume 141, Issue 1, July 2021, Pages 255-275

APPENDIX

1.Appendix for Data Collection

```
install.packages("tm")
install.packages(“SnowballC”)
install.packages(“rvest”)

library(tidyverse)
library(RedditExtractoR)

library("tm")
library(rvest)

raw.oz.lines =
readLines("https://www.reddit.com/r/wallstreetbets/search?q=flair_name%3A%22Dail
y%20Discussion%22&restrict_sr=1&sort=new.htm")

oz.obj = html(raw.oz)
oz.text.3 = html_text(oz.obj)

oz <- tm_map(oz, tolower) #lowercase
oz <- tm_map(oz, removePunctuation, preserve_intra_word_dashes = FALSE)
oz <- tm_map(oz, removeWords, stopwords("english"))
oz <- tm_map(oz, stemDocument)
tdm = TermDocumentMatrix(oz)
tm_term_score(tdm, "AMC" “To the Moon”)
tm_term_score(tdm, "GME" “To the Moon”)
tm_term_score(tdm, "BB" “To the Moon”)
```

```
 reddit <- get_reddit(subreddit = "wallstreetbets", page_threshold = 5000)
reddit1 <- get_reddit(subreddit = "wallstreetbets", page_threshold = 1, cn_threshold =
5000)
reddit %>% saveRDS("data/reddit.RDS")
reddit <- readRDS("data/reddit.RDS") %>%
  as_tibble() %>%
  mutate_at(vars(contains("date")), as.Date)

https://www.nasdaq.com/market-activity/stocks/screener
stocks <- read_csv("GME AMC BB.csv")

stocks %>%
  filter(Symbol == "GME")

stocks %>%
  filter(Symbol == "AAPL")

reg_expression <- regex(paste0("\\b(?:",
                               paste(stocks$Symbol, collapse = "|"),
                               ")\\b"))

reddit_mentions <- reddit %>%
mutate(stock_mention = str_extract_all(comment, reg_expression)) %>%
unnest(stock_mention)
#
# reddit_mentions %>% saveRDS("data/reddit_mentions.RDS")
reddit_mentions <- readRDS("data/reddit_mentions.RDS")
```

```
reddit_mention_counts <- reddit_mentions %>%
  group_by(post_date, stock_mention) %>%
  count()

reddit_mention_counts %>%
  filter(stock_mention %in% top5) %>%

reddit_mentions %>%
  filter(!(stock_mention %in% fp)) %>%
  group_by(stock_mention) %>%
  count() %>%
  arrange(-n) %>%
  print(n = 20)

reddit_mentions %>%
  filter(!(stock_mention %in% fp)) %>%
```

2. Appendix for Regressions

```
import excel "C:\Users\mis\Downloads\AMC.xlsx", sheet("AMC") cellrange(A3:M108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc
import excel "C:\Users\mis\Downloads\BB.xlsx", sheet("BB") cellrange(A3:M108)
firstrow clear
gen  lnprice=log(StockClosePrice)
```

```
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc
import excel "C:\Users\mis\Downloads\GME.xlsx", sheet("GME") cellrange(A3:M108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc
import excel "C:\Users\mis\Downloads\TLRY.xlsx", sheet("TLRY")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc
import excel "C:\Users\mis\Downloads\SNDL.xlsx", sheet("SNDL")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc
import excel "C:\Users\mis\Downloads\CLOV.xlsx", sheet("CLOV")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am1.doc

import excel "C:\Users\mis\Downloads\NIO.xlsx", sheet("NIO") cellrange(A3:L108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
```

```
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am2.doc
import excel "C:\Users\mis\Downloads\AMD.xlsx", sheet("AMD") cellrange(A3:L108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am2.doc


import excel "C:\Users\mis\Downloads\PLTR.xlsx", sheet("PLTR")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
outreg2 using am4.doc
import excel "C:\Users\mis\Downloads\MVIS.xlsx", sheet("MVIS") cellrange(A3:L108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
outreg2 using am4.doc
import excel "C:\Users\mis\Downloads\AAPL.xlsx", sheet("AAPL")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
outreg2 using am4.doc

import excel "C:\Users\mis\Downloads\NOK.xlsx", sheet("NOK") cellrange(A3:L108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
```

```
outreg2 using am3.doc
import excel "C:\Users\mis\Downloads\BBBY.xlsx", sheet("BBBY")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am3.doc
import excel "C:\Users\mis\Downloads\TSLA.xlsx", sheet("TSLA") cellrange(A3:L108)
firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am3.doc
import excel "C:\Users\mis\Downloads\WKHS.xlsx", sheet("WKHS")
cellrange(A3:L108) firstrow clear
gen  lnprice=log(StockClosePrice)
gen lntrend=log(MentionedTimesinWSB)
reg lnprice lntrend SP500 ShortInterest
ivreg lnprice SP500 ShortInterest (lntrend = lnprice)
outreg2 using am3.doc
```