IIMB-WP N0. 644/2021

WORKING PAPER NO: 644

WallStreetBets Against Wall Street: The Role of Reddit in the GameStop Short Squeeze

Abhinav Anand
Assistant Professor
Finance & Accounting
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
abhinav.anand@iimb.ac.in

Jalaj Pathak
Assistant Professor
Finance & Accounting
Great Lakes Institute of Management
Chennai
jalaj.p@greatlakes.edu.in

Year of Publication – June 2021

WallStreetBets Against Wall Street: The Role of Reddit in the GameStop Short Squeeze

Abhinav Anand*

Jalaj Pathak†

June 22, 2021

Abstract

By January 2021, the stock of GameStop (GME) was heavily shorted by institutional investors but an unprecedented retail campaign for pushing-up its stock price culminated in a short squeeze by the end of January 2021. Adapting recent innovations in text analysis in finance and on microblogging platforms, we present evidence that both the tone and the volume of discussions on the subreddit `r/wallstreetbets` had significant predictive associations with the GME return, volatility and put-call ratio.

Keywords: Short Squeeze, GameStop, WallStreetBets, Text Analysis, Social News Platform

JEL Classification: D91, G14, G40

---

*Indian Institute of Management Bangalore

†Great Lakes Institute of Management, Chennai

"Yea there's deep value, then there's deep fucking value."

Keith Gill (u/DeepFuckingValue), r/wallstreetbets.[1]

1 Introduction

At the beginning of 2021, financial markets in the US witnessed an up-to-then unprecedented event: social-media-coordinated trading actions of numerous retail investors long on GameStop (GME)—a video game company—forced its price to appreciate rapidly leading to major losses for several institutional investors who had massively shorted the stock. What was unprecedented was not the (extremely rapid) price appreciation, nor losses for the short sellers, but in fact the decentralized, coordinated buying of GME stock by retail investors on such a vast scale as had never been accomplished prior to the advent of social media and app-enabled access to trading opportunities. Although the digitally mass-coordinated stock-buying did inflict damage on some savvy hedge funds who had shorted GME stock, the retailers' risky trading strategies and the subsequent GME stock price decline led to a Pyrrhic victory for retailers since many went on to endure extremely heavy eventual losses.

Our paper is among the very few studies that is able to answer—in the affirmative—the following important questions in the light of the developments cited above: were GME returns and/or their volatility indeed influenced by the volume and the tone of discussions on online portals, over and above the factors that impact returns and/or volatility? In the absence of digital social media, such a concerted, synchronized trading action over a short span of time by countless retail investors dispersed all over the world could not have happened; but now that such unprecedented app-enabled retail cooperation is possible, our paper is among the first to be able to quantify the extent of the influence of such concentrated retail attention on stock prices and volatilities.

[1] See archived link for the quote here.

Yet another important but often overlooked motivation for our study is the implicit overweighting of such 'meme stocks'—which are those stocks that command such excess retail attention—in passive, exchange traded funds (ETFs). For example, the GME and AMC stocks—two most noteworthy meme stocks from early 2021—are currently the two largest holdings in BlackRock's iShares Russell 2000 Value ETF.[2] Passive index trackers will by construction become over-exposed to such stocks, especially during short squeeze episodes, and the fact that portfolio rebalancing is done usually quarterly or at a later frequency (to minimize transactions costs) could lead to potential losses for the unsuspecting ETF holder. By providing evidence that retail trading campaigns coordinated by discussions on online portals may lead to over-exposure to meme stocks in ETFs, we show that this source of concentration risk can add to the portfolio risk for ETF managers.

Reddit is a social media firm which relies on news aggregation, discussion and user-generated content. Users post text, links, images, videos etc. which are then upvoted or downvoted by other members. Posts on similar topics are organized in the form of 'subreddits' or 'communities'. The community centered around the subreddit `r/wallstreetbets` (WSB) is active in stock and option trading. Over the course of two weeks in late January and early February 2021, the fortunes of the GME stock oscillated rapidly, without there having been any major change in the fundamentals of the company. By late January, several institutional investors had shorted GME and had publicly explained their viewpoint over several media platforms. However, the (mostly) retail trading community centered on the Reddit portal `r/wallstreeetbets` (WSB) had a different perception of the stock and discussions therein showed most members to have advocated a long position in it. Over the last week in January 2021, due to unprecedented retail interest in the GME stock, it increased in value from USD 30 to USD 483 at its peak. This decimated investors who had short-sold GME stock who were then forced to close their short positions at massive losses on account of the

[2] See the news article link here.

squeeze.

To analyze whether the volume and tone of the discussions on the WSB subreddit influenced GME stock returns and/or volatitilities, we extract the tone of WSB threads using a dictionary-based approach. We employ the recently introduced innovations in financial text analysis in Anand et al. [2021a]—which recommends usage of the sentence as a unit of text analysis; and quantifies the effect of adverbs and adjectives which modify the texts' meaning—to accurately quantify the tone expressed in the subreddit's threads. These innovations help to correctly quantify the tone of sentences containing 'valence shifters'—words like 'but', 'although', 'despite', 'faintly' etc.—which add context to the meaning of sentences but have been ignored so far in dictionary-based bag-of-words approaches [Anand et al., 2021b]. Further, as a proxy of the volume of discussion, we test whether the number of threads on the WSB subreddit (thread count) has any putative influence on GME's stock. Since text analysis on microblogging platforms and portals such as Reddit are especially challenging, we use multiple, independent lexicons such as Mohammad and Turney [2010], Jockers [2017], MPQA and Sentiword etc. to correctly quantify polar words and phrases. The WSB subreddit features heavy usage of emojis for which we rely on techniques presented in Kralj Novak et al. [2015].

Our main finding is the documentation of significant predictive associations of WSB subreddit tone as well as the number of threads (proxy for discussion volume) on both the GME stock return, as well as on its volatility. Further, we also show significant predictive impact of the subreddit's tone and thread count on GME's put-call ratio—all of which is broadly consistent with recent papers on the short squeeze [Allen et al., 2021, Lyócsa et al., 2021, Long et al., 2021]. Our results survive a battery of auxiliary tests, out-of-sample analysis, as well as the inclusion of additional controls at both the daily and intraday frequency. On account of much lower barriers to entry for retail investors due to easy app-based investing, it is quite possible that such retail mass-coordinated retail campaigns are able to move stock prices and volatilities significantly off-course during such future episodes. Our paper is among the first which is able to rigorously unearth evidence confirming the popular suspicion that a large number of retail traders acting in concert can move financial variables off their equilibrium paths, at least in the short run.

Text analysis in finance has become an active field over the past several years. Among the earliest such works was Antweiler and Frank [2004] where the authors tested the impact of Yahoo! messages on stock returns. In a similar vein, Das and Chen [2007] studied Amazon sentiment from message boards using machine learning algorithms. Campbell et al. [2012] and Garcia [2013] examined the impact of media sentiment during periods of market volatility. Further, Chahine et al. [2015] analyzed the impact of media news on earnings management before equity offerings and reported significant results. Similarly, Danbolt et al. [2015] examined the relationship between investor sentiment and bidder annoucement abnormal results. Ahmad et al. [2016] investigated the impact of media tone and firm level stock returns. Somewhat similarly, Bajo and Raimondo [2017] scrutinized the impact of media sentiment on IPO underpricing.

The microblogging platform Twitter has been shown to be associated with movements in the stock market [Behrendt and Schmidt, 2018]. Similarly Feng and Johansson [2019] examine how microblogging by top executives on Weibo—the microblogging alternative popular in China, where Twitter remains banned—is significantly associated with dissemination of more firm-specific information to the capital market. Relatedly, Cioroianu et al. [2021] examine the corporate effects of blockchain related technological development and report that social media response is a significant variable. Even more recently, easy access to free, app-based investing, especially during the Covid-19 pandemic, has changed the landscape of the trading population by lowering the barriers to entry for retail investors. Eaton et al. [2021], Aharon et al. [2021] and Friedman and Zeng [2021] are some recent studies which analyze the impact of the influx of Robinhood platform users on trading in the financial markets.

However, quite naturally, there are only a very few recent papers which explicitly examine the Jan-Feb 2021 short squeeze episode in detail. In fact, other than GameStop, there were several other stocks which saw extreme retail attention during that period which led to short squeeze episodes, although their scale was not as dramatic as in the case of GameStop.[3] Allen et al. [2021] is a notable study which examines the aforementioned short squeeze episodes in all such stocks but from the perspective of its effects on market quality (or lack thereof). Lyócsa et al. [2021] studies the short squeeze episodes in GME, AMC, Blackberry and Nokia and report that WSB subreddit activity drives daily price changes in these stocks. Finally, Long et al. [2021] is a recent working paper which comes closest in spirit to our objective: it performs text analysis on the WSB subreddit during the short squeeze episode at the intraday 1-min interval level and finds that both the tone and the number of comments influence GME intraday returns. Our main results are in broad agreement with those of the papers discussed above, especially Long et al. [2021]. However, our anlysis is more comprehensive and our results more robust, since i) we conduct analysis on both daily and intraday (1-min interval) frequency; ii) we employ a theoretically sound multiple-lexicon based ngram analysis based on new innovations in financial text analysis [Anand et al., 2021a]; iii) we analyze the impact of WSB subreddit on GME volatility—which has not been examined previously, iv) we include a larger collection of control variables, and v) we corroborate our results by out-of-sample cross-validation.

The rest of the paper is organized as follows. Section 2 briefly recounts the GME short-squeeze episode in early 2021, section 3 describes our data sources; section 4 explains our methodological execution; section 5 presents our results and analyses. Section 6 describes our robustness exercises and finally section 7 offers concluding remarks.

[3]Some examples are AMC, Nokia, Koss, Blackberry, American Airlines etc.

2 A Brief Overview of the GameStop Saga

2.1 Background

GameStop (GME) is a video game company headquartered near Dallas, Texas. From 2016 onwards, due to a variety of factors it had been performing poorly, leading to several institutional investors having shorted its stock. By January 22 2021, about 140% of its public float had been shorted—an uncommonly high figure.[4]

The online community centered around the subreddit 'r/wallstreetbets' (WSB) features a large group of users who are interested in trading and investing in the financial markets. A substantial fraction of WSB users are young retail investors who tend to prefer risky day-trading and employ aggressive, high-leverage strategies such as using borrowed student loans to bet on certain stocks.[5] The subreddit's tone is often irreverential and expletive-laden and features heavy usage of slang and in-jokes.[6] At the peak of the GME short-squeeze episode, its number of users surged by over 6 million.

Among the users of WSB, there had been prior interest in the GME stock over the perception of it being underpriced. For example, the well-known WSB user with the alias 'DeepFuckingValue'—whose quote we feature at the beginning of our paper—claimed to have had bought GME stock options worth USD 53,000 in 2019 whose value subsequently rose to about USD 48 million by January 27, 2021. The identity of this user was later revealed to be that of Keith Gill, a financial advisor from Massachussets in a news report by Reuters.[7] Keith Gill's bullish stance regarding the GME stock and his subsequent gains have been cited to be among the important factors which led to the GME short squeeze in late January 2021.

[4] See link to the news story here.

[5] See archived link for the news story here.

[6] For example, the use of 'stonks' for stocks, 'to the moon' for a price that is predicted to rise a lot, 'DD' for due diligence, 'paper hands' for someone who sells at the first sign of loss etc.

[7] See link to the archived news story here.

2.2 Ascent

At the beginning of 2020, GME's stock price was around USD 5.88 per share which subsequently increased marginally to USD 6.11 per share by September 2020. Interest regarding the GME stock among short sellers remained high during December 2020 and early January 2021. In order to improve its financial performance and profitability, GME announced the appointment of three new directors on January 11, 2021. While this did not improve the short sellers' sentiment, users at WSB reacted positively to these developments and postings regarding GME heightened during the build-up to the short squeeze in late January 2021.

On January 19, 2021, Citron Research—an online investment newsletter which had shorted GME—published its analysis on GME and effectively called GME buyers 'suckers'. In response to this, the posts on WSB regarding GME increased exponentially and retail interest in the stock intensified manifolds. By January 27, GME's stock price had increased over 1500% over the levels from two weeks prior. On the same day, the highest intraday stock price for GME was recorded as USD 483—about 190 times its lowest prior price of USD 2.57 recorded in April 2020. At its peak on January 28, during pre-market hours the stock price hit its highest value of USD 500. From December 2020 to January 2021, an investor in GME would have accumulated returns of over 900%, without there having been much corresponding change in the fundamentals of the company.

2.3 Descent

On January 28, 2021, the app-based brokerage Robinhood halted the purchase of GME stock citing their inability to post sufficient collateral at clearing houses for the execution of their clients' orders. Trading was also halted in some other stocks which were also caught in short squeezes due to unprecedented retail interest e.g., AMC Theatres, Blackberry, American Airlines, Koss, Nokia etc. On January 29, 2021 Robinhood revealed that it had raised an additional USD 1 billion to protect itself from the emergent financial pressures and to meet the increased collateral requirements.[8]

A sequence of major corrections soon followed, especially over February 1 and 2, 2021 as the GME stock declined precipitously and shed almost 90% of its stock price to end up at USD 90 per share. Notwithstanding these adverse developments, some WSB users continued to hold on to GME shares due to a variety of viewpoints which subsequently led to heavy losses including a majority of savings lost for some users.[9]

2.4 Aftermath

Although retail traders who had gone long suffered serious losses due to the reversal in the stock price by early February 2021, several institutional investors who had shorted GME stock on account of its poor perceived performance faced significant losses as well. The price appreciation of GME stock due to activity on the WSB subreddit led to a major short squeeze which inflicted heavy damage to those who had shorted the stock and were hoping to profit from low prices. For example, Melvin Capital, a hedge fund that had heavily shorted GME was reported to have lost 53% of its value by the end of January 2021 although it claimed to have closed its short position by January 26, 2021.[10] Similarly, it was reported that short sellers lost USD 6 billion on January 26, 2021 due to the short squeeze.[11]

The GME short squeeze episode—especially the halt in trading on the Robinhood app—has proven controversial politically. The US Securities and Exchange Commision, the Congress, as well as the Senate Banking Committee have launched investigations into the matter. A major concern voiced has been the protection of retail investors from allegedly manipulative trading activity. On the other hand, the concerted action of retail traders that led to the short squeeze has been criticized and deemed manipulative by Allen et al.

[8] See link to archived news story here.

[9] See links to the news stories here and here.

[10] See archived link to the news story here.

[11] See link to the archived news story here.

[2021]. They present evidence that it impeded market quality and advocate active monitoring of social media platforms to curb such behavior.

3 Data

There are two major sources of data used in this paper. The data from the subreddit WSB are downloaded using the 'pushshift API'; and the data on GameStop's daily and intraday returns have been collected from Bloomberg. Our sample duration comprises the date range from December 1, 2020 to April 31, 2021. Our initial Reddit sample consists of 36,691 threads containing the following keywords: 'GAMESTOP', 'gamestop', 'GameStop', 'Gamestop', 'GME' and 'gme'. We filter out all threads which are empty and those with fewer than 10 characters. Some of these include threads which were taken down by Reddit due to violation of its rules while others were removed by the users themselves. Our final sample consists of 10,997 threads with an aggregate of 832,360 comments.

The data for the Google Search Index (GSI) has been collected via the R library 'gtrendsR' [Massicotte and Eddelbuettel, 2021]. To calculate GSI we average the GSI score of 14 terms across the specified time period. These include the following keywords: 'GameStop', 'gamestop', 'GAMESTOP', 'GME', 'gme', 'short squeeze', 'short sell', 'call option', 'wall street bets', 'Melvin Capital', 'to the moon', 'Keith Gill', 'Dave Portnoy' and 'Justin Sun'. The data for Twitter Sentiment is taken from Bloomberg. Bloomberg classifies respective tweets and the sentiment of those tweets for each company. The sentiment is calculated using using a machine learning algorithm and for each company the tweet count and sentiment can be downloaded using the 'GN' Bloomberg function.

4 Methodology

4.1 Tone Quantification

In order to extract the tone of posts on the subreddit WSB, we rely on a dictionary-based ngram approach. Further, we augment this approach by employing two recent innovations introduced in financial text analysis: ngram analysis based at the sentence level, and assignment of appropriate weights for 'valence shifters' which modify the meaning of sentences but have been granted zero weight in all current dictionary-based approaches. We elaborate on these aspects in detail below.

We calculate the tone for each WSB subreddit thread by classifying it as a collection of sentences. For each thread we download the text, user id, time, date, upvotes and comments from the WSB subreddit. We parse the content and remove all html tags and remove symbols such as '@' and urls. Following this, all emojis in the text are identified and assigned weights according to their text meaning. For example, the emoji ':)' is characterized as 'smiling face'. Since the threads are downloaded from the subreddit `r/wallstreetbets`, money related emojis are encountered frequently.

All text-related content is then converted to lower case. We identify all possible punctuation marks in the text and following this, the text between two full stops; a full stop and a question mark; and between two question marks is classified as a sentence. A complete thread is thus broken down into a collection of sentences. For each sentence, words are classified into two categories: valence shifters—adjectives and adverbs such as 'few', 'but', 'despite' etc. which modify the meaning of sentences—and polar words (which signify a positive/negative connotation). Usage of valence shifters in financial text analysis has been introduced recently in Anand et al. [2021a] and Anand et al. [2021b]. Relatedly, tone/sentiment quantification analysis at the sentence level has been advocated by Andreevskaia and Bergler [2008].

Since we rely on dictionaries to assign meanings to words, our approach is only as good as the dictionaries we employ. Choosing a wrong or irrelevant dictionary will lead to wrong tone quantification. The subreddit `r/wallstreetbets` poses a special challenge in this regard. Most of the discussion on its threads is clearly financial in nature and hence one must employ a dictionary that correctly assigns meanings to financial terms used in the discussions. However, there is a lot of content in the form of emojis as well as several varieties of slang for which standard financial dictionaries such as those by Loughran and McDonald (LM) [Loughran and McDonald, 2011] are of no use.

In order to surmount these problems, we cover a wide range of dictionaries by referring to multiple lexicons in order to correctly identify polar words and phrases. In addition to the standard LM dictionary, these include Mohammad and Turney [2010] and Jockers [2017]. These lexicons contain polar words popular in informal usage such as 'awful' (negative polarity, weight -1), 'awesome' (positive polarity, weight +1) etc. We augment these lexicons with those of MPQA (https://mpqa.cs.pitt.edu/); and Sentiword (https://www.aclweb.org/anthology/L06-1225/) which have been shown to perform with a precision of over 85% for microblogging content [Ghiassi and Lee, 2018]. For quantifying emoji tones, we employ Kralj Novak et al. [2015] which assigns polarity to emojis like 'thumbs up/down', 'smiling face', 'money bag' etc. among others. Finally we manually assign weights to several idiosyncratic slang-origin terms popular on WSB such 'paper hands', 'YOLO', 'to the moon' etc. Moreover, we also specify weights for words such as 'call', 'put', 'short squeeze' etc. which have not been assigned weights in financial dictionaries such as LM.

The methodology of valence shifters and ngram analysis at the sentence level has been adapted from Anand et al. [2021a] which in turn source them from Kennedy and Inkpen [2006], Polanyi and Zaenen [2006] and Schulder et al. [2018]. These papers show how the presence of valence shifters in the text quantification process improves accuracy over the unigram analysis. These valence shifters can be further classified into four categories: amplifiers ("absolutely", "acutely", "very"), de-amplifiers ("barely", "faintly", "few"), negators ("not", "cannot") and adversative conjunction ("despite", "but"). The amplifiers, de-amplifiers, and adversative conjunction are given a weight a 0.8—positive for an amplifier, negative for a de-amplifier and negative for the words before adversative conjunction and positive for the words after adversative conjunction.[12] This is done because adversative conjunction such as "but" will amplify the argument after it and weight down the argument before it.[13] The negators are given a value of -1. We find that 10 percent of posts in our sample have one or more valence shifters.

As for the tone quantification process, for each sentence, first, the polar words are identified and given the weight of +1/-1, following which valence shifters are identified around each polar word from the beginning till the end of the sentence. Thus, each polar word along with its set of valence shifters is classified as a word cluster for each sentence.

The process is elaborted below for a sentence taken from a sample WSB thread in this study:

> *"my trade would have been up about $130k from oct 9 to oct 10, but failure to take proper action only allowed me to realize about $90k in realized profits in one day's time."*

Using the LM dictionary-based "bag-of-words" approach the tone of the above sentence is calculated as:

$$\frac{(+1)[\text{=up}] + (-1)[\text{=failure}] + (+1)[\text{=profits}]}{17} = 0.058$$

Now, using the methodology borrowed from Anand et al. [2021a], the tone is calculated as below:

Firstly, polar words/phrases are identified from the sentence followed by valence shifters around these polar words/phrases. Thus each sentence is divided into clusters with respect to polar words/phrases such as:

[12]The weight, 0.8, is as per the existing literature. We verify the results by varying the weight of valence shifters from 0.5 to 0.9 and our results continue to hold.

[13]E.g. "The service is good but there is a lot of scope for improvement."

`1.` *my trade would have been* ***up*** *about $130k from oct 9 to oct 10,* ***but failure*** *to take proper action*

`2.`***only*** *allowed me to realize about **$90k in realized* ***profits*** *in one day's time.*

Thus, the above sentence is divided into two clusters with ***but*** being a valence shifter (adversative conjunction) in the first cluster and ***only*** being a valence shifter (de-amplifier) in the second cluster.

The tone calculated is as follows:

$$(+1)[=\text{up}] + (-0.85)[=\text{but}] = +0.15$$

$$(-1)[=\text{failure}] = -1$$

$$(+1)[=\text{profits}] + (-0.8)[=\text{only}] = +0.2$$

$$\frac{(+0.15)[=\text{first cluster}] + (-1)[=\text{second cluster}] + (+0.2)[=\text{third cluster}]}{19} = -0.034$$

The tone is negative as compared to the "bag-of-words" approach since the valence shifters "but" and "only" bring down the impact of the positive polar words "up" and "profits" respectively. The number of non stop-words in the denominator is higher in case of new methodology due to the introduction of the valence shifters.

4.2 Empirical Design

We examine the impact of WSB subreddit thread tone on GME return as per the following regression equation:

$$R_t = a_0 + b_n Tone_{t-n} + d * Controls + \gamma_t \tag{1}$$

The dependent variable is the GME stock's daily return and the putative impact of the WSB subreddit's tone is tested upto five lags (n assumes values from 0 to 5). The controls include number of comments and upvotes for each thread, the Twitter sentiment, the Google Search Index (GSI); and the lag of GME stock return. The regression methodology is ordinary least squares with heteroskedasticity and autocorrelation consistent (HAC) errors.

On similar lines, we test the impact of WSB subreddit's thread count on GME returns as per the following regression specification:

$$R_t = a_0 + b_n ThreadCount_{t-n} + d * Controls + \gamma_t \tag{2}$$

Controls are the same as in specification (1).[14]

Both specifications are also tested for the intraday 1-min interval GME return as well as the intraday 1-min interval WSB tone and thread count respectively. The two main changes in the intraday regression specification are that i) 10 1-min interval lags are tested instead of 5, and ii) there are only three control variables—lagged return, number of upvotes and the number of comments. The reason we are not able to incorporate the control variables such as the Google Search Index and the Twitter Sentiment is because the highest frequency at which those are available is at the daily frequency.

We specify the same regression equations for the daily and intraday GME volatility testing, with the same set of corresponding lags and controls as elaborated above.[15]

$$Vol_t = a_0 + b_n Tone_{t-n} + d * Controls + \gamma_t \tag{3}$$

$$Vol_t = a_0 + b_n ThreadCount_{t-n} + d * Controls + \gamma_t \tag{4}$$

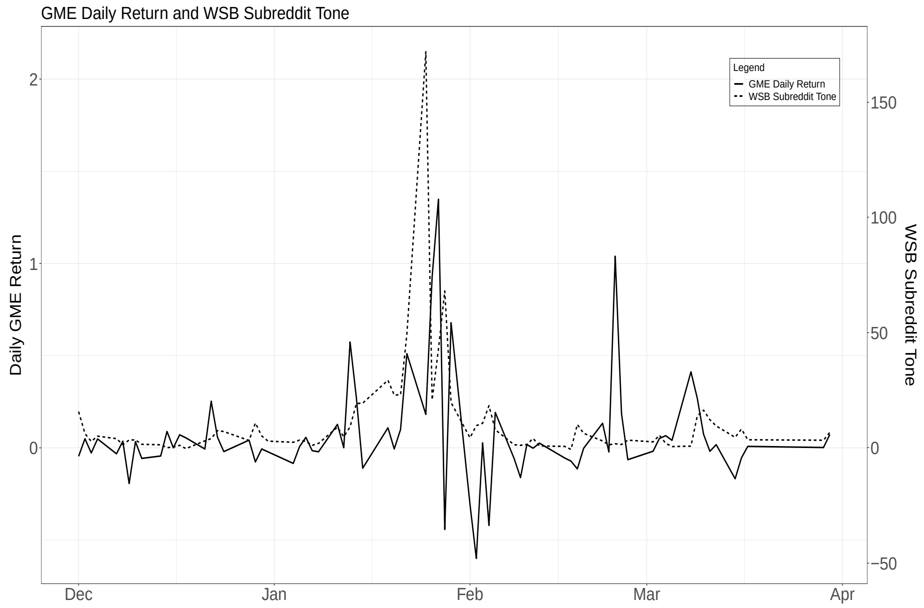

Figure 1: Comovement of GME return (solid line) and the WSB subreddit tone (dotted line) from December 2020 to April 2021.

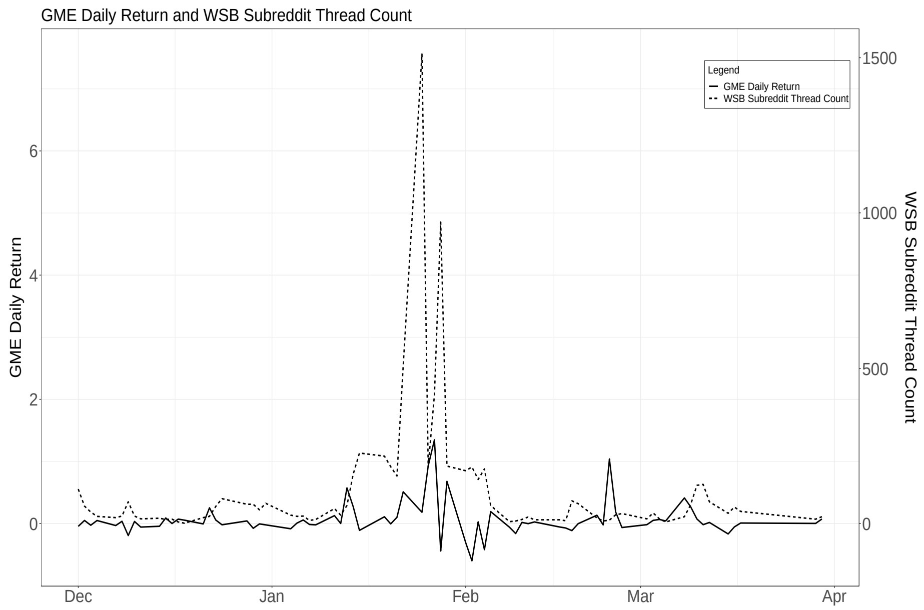

Figure 2: Comovement of GME return (solid line) and the WSB subreddit thread count (dotted line) from December 2020 to April 2021.

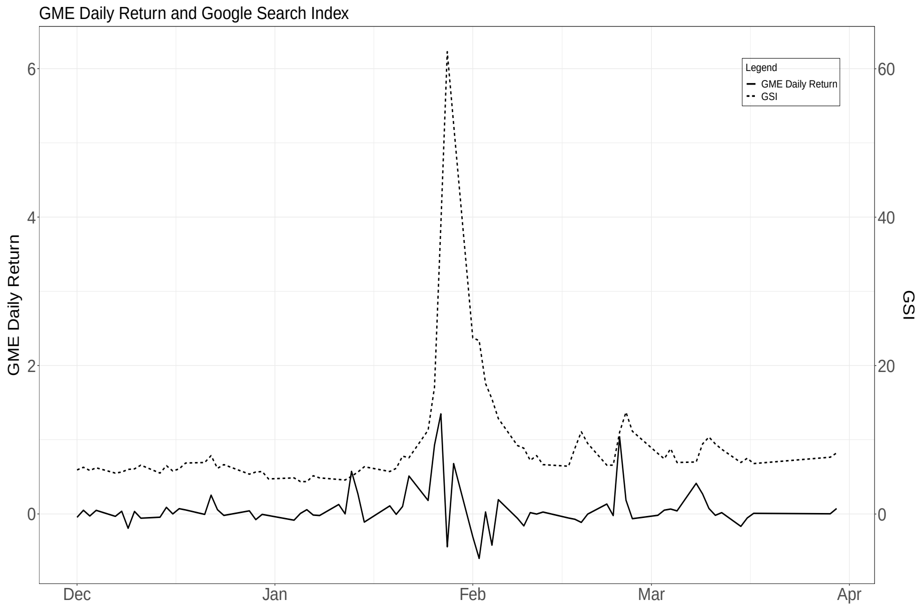

Figure 3: Comovement of GME return (solid line) and the Google Search Index (dotted line) from December 2020 to April 2021.

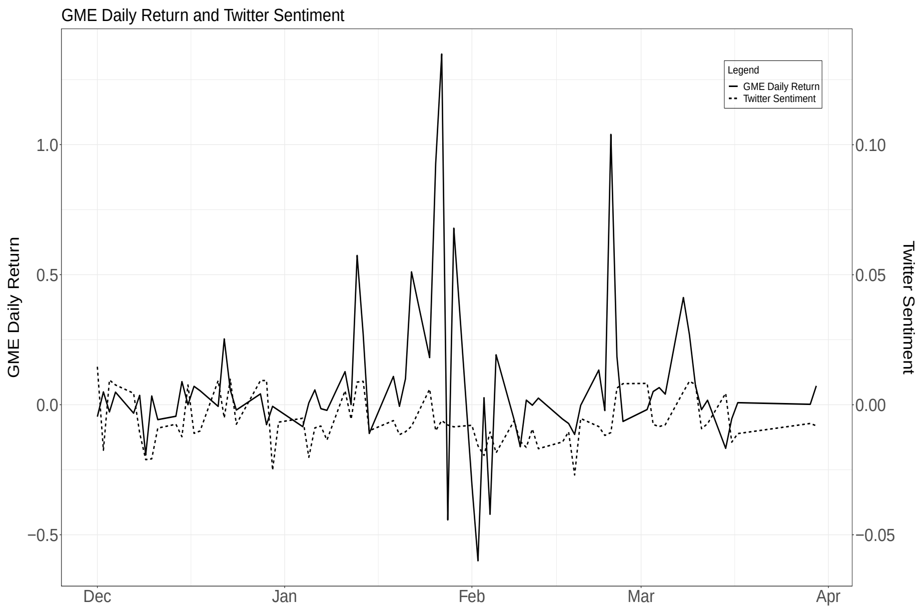

Figure 4: Comovement of GME return (solid line) and the Twitter Sentiment (dotted line) from December 2020 to April 2021.

5 Results and Analysis

Figures 1, 2, 3 and 4 present the plots for the daily comovement of GME stock return with the WSB subreddit thread tone, the subreddit thread count, the Google Search Index (GSI), and the Twitter Sentiment respectively. A visual inspection hints at strong comovement in January and February 2021.

In particular, the figures show remarkable visual evidence that the daily WSB subreddit tone and the daily WSB subreddit thread count seem to lead the movement in the GME stock's daily returns. On the other hand, this relationship seems reversed for the Google Search Index where the search index seems to lag the movement in the daily returns. For the Twitter Sentiment variable, the comovements look somewhat contemporaneous with some evidence for the Twitter Sentiment lagging the GME returns slightly.

Figure 5 presents the plot of daily GME stock volatility on the primary axis and the WSB subreddit tone on the secondary axis. Again there seems strong visual evidence that the dotted line representing the subreddit's daily tone leads movements in the daily GME volatility—especially around late January 2021, corresponding to the short squeeze episode.

In order to ascertain the descriptive aspects of the variables we analyze, we refer to table 1. It contains the summary statistics of all variables used in this study at the daily frequency. For the variable 'GME return', the discrepancy between the daily median return and the daily mean return signifies the heavy-tailed nature of daily GME stock returns. The fact that the median is much lower than the mean points out that there are substantially many positive outliers—implying a fat right tail—which fits the facts of the rapid price appreciation in the GME stock during our sample period. The same information can also be gleaned from the large difference between the standard deviation and the inter-quartile range. In fact, except for the vari-

[14]When the volume of discussion is proxied by the number of threads, the appropriate Twitter control variable is the number of Twitter comments.

[15]One major difference in the control specification is that we do not include lagged volatility as a control.

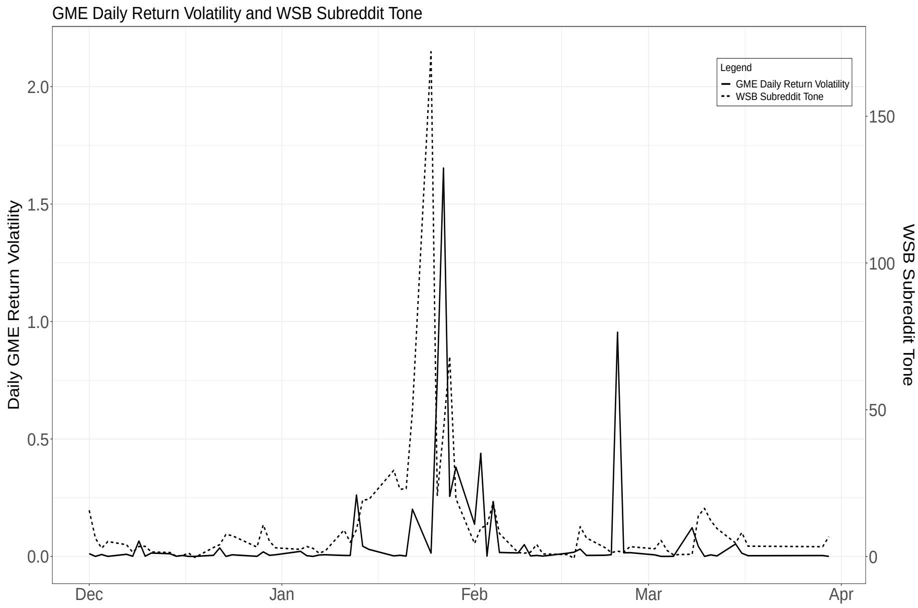

Figure 5: Comovement of GME return volatility (solid line) and the WSB subreddit tone (dotted line) from December 2020 to April 2021.

Table 1: Descriptive Statistics

	Min	Mean	Median	Max	SD	IQR
GME Return	−0.600	0.066	0.001	1.348	0.288	0.115
WSB Tone	−0.707	10.631	4.043	171.893	22.255	8.20
Num Threads	1	100.77	33.5	1512	215	62.75
Num Comment	46	7,530	1929.5	126,889	17,603	5539
Num Upvotes	4	7,030	71.5	166,279	24,790	931.75
GSI	4.357	9.783	6.96	62.286	9.570	2.82
Twitter Sent	−0.027	−0.006	−0.008	0.015	0.010	0.013
Twitter Count	131	9,651.081	3483	74,211	15,773.150	9245.5

Note: Summary statistics for all variables at the daily frequency.

able ‘Twitter Sentiment’, all other variables feature fat right tails on account of a large dispersion between the lower median values and the higher mean values. This is also reflected in the notable discrepancies between the larger standard deviations and the lower inter-quartile ranges.

To examine more formally, the role played by the WSB subreddit on GME stock returns outlined in the plots above, we investigate the impact of the subreddit's tone on GME's return as well as its volatility. We employ ordinary least squares with HAC errors and conduct the analysis on daily returns as well as on intraday 1-min interval returns.

5.1 WSB subreddit's impact on GME returns

5.1.1 Impact on daily returns

In table 2 we regress daily GME returns on daily WSB subreddit tone in the presence of the following controls as specified in regression specification (1): lag 1 daily GME return; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent); as well as the day of the week and month dummies. Inclusion of the first lag of daily GME returns ensures that we account for any short-term momentum based effects; GSI and Twitter Sentiment are included as controls because Google and Twitter are among the most popular venues for retail (as well as institutional) investors' queries regarding trends in stock prices; and finally the number of upvotes and number of comments are thread-level proxies for the popularity of the thread on which a discussion regarding GME stock prices is centered.

Insert table 2 here.

The results are as follows: the WSB subreddit coefficient displays significance at daily lags 1, 2, 3, 4 and 5. In other words, from December 2020 to March 2021 even after controlling for a variety of variables, the subreddit's tone has significant predictive association with daily GME returns for five successive future days.

Among the controls, we observe that GME lagged return exhibits significance at lag 4; the number of upvotes on a thread exhibits contemporaneous significance; the number of comments are significant at daily lags 4 and 5; and the Google Search Index shows significance at daily lags 3, 4 and 5. Twitter Sentiment does not display any significance.

On a similar note, in table 3 we conduction regression analysis of daily GME returns on WSB thread count in line with the specification in equation (2).

Insert table 3 here.

The results are as follows: the number of WSB threads shows high significance uniformly from lags 0–5. In other words, from December 2020 to March 2021, even after controlling for all relevant variables, the WSB subreddit's thread count has significant predictive association with GME daily returns contemporaneously as well as for five successive future days.

Among the controls, we observe that GME lagged return exhibits significance at lag 4; the number of upvotes on a thread exhibits contemporaneous significance; the number of comments are significant at lags 4 and 5; the Google Search Index displays significance at lag 4; and the number of Tweets has significant predictive associations contemporaneously as well as at lag 5.

5.1.2 Impact on intraday returns

In table 4, we regress intraday one-minute interval GME returns on intraday one-minute interval WSB subreddit tone in the presence of the following controls as specified in regression equation (1): lag 1 intraday one-minute GME return; number of upvotes; and the number of comments. We are not able to include other daily frequency control variables such as the Google Search Index and the Twitter Sentiment because the highest frequency at which they are available is at the daily frequency.

Insert table 4 here.

The results are as follows: the intraday WSB subreddit tone coefficient displays significance at the third one-minute interval, while the one-minute lagged GME return exhibits significance uniformly for all lags 0–10. The other variables—number of upvotes and the number of comments—show no significance at any lag.

On a similar note, in table 5 we conduct intraday one-minute interval GME returns on intraday one-minute interval WSB subreddit thread count in the presence of the following controls as specified in equation (2): lag 1 intraday one-minute GME return; number of upvotes; and the number of comments. We have to exclude GSI and Twitter Sentiment for the reasons explained above.

Insert table 5 here.

The results are as follows: the intraday WSB subreddit thread count coefficient displays significance at the fourth one-minute interval, while the one-minute lagged GME returns exhibit significance uniformly for all lags 0–10. The other variables—number of upvotes and the number of comments—show no significance at any lag

5.2 WSB subreddit's impact on GME volatility

5.2.1 Impact on daily volatility

We calculate the daily realised return volatility by demeaning the squared residual returns and then calculating the mean of the demeaned residual over five days [Tetlock, 2007].

The results are compiled in table 6 and are as follows: the WSB subreddit coefficient displays significance at daily lags 1, 2 and 3. In other words, from December 2020 to March 2021 even after controlling for a variety of variables, the subreddit's tone has significant predictive association with daily GME volatility for three consecutive future days.

Insert table 6 here.

Among the controls, we observe that the number of upvotes on a thread has no significance; the number of comments are significant at daily lags 1, 2, 4 and 5; and the Google Search Index shows significance at daily lags 0–4. Twitter Sentiment does not display any significance in being associated with GME daily volatility.

On a similar note in table 7, we test the impact of the number of WSB subreddit's threads on GME daily volatility in the presence of our standard controls. The results are as follows: the number of threads shows significant predictive association with daily GME volatility 2 days in advance.

Among other variables, the number of upvotes shows contemporaneous significance, the number of comments shows significance at lags 2, 4 and 5; while the number of Twitter comments shows contemporaneous significance with daily GME volatility. The Google Search Index has no significant association with GME volatility.

Insert table 7 here.

5.2.2 Impact on intraday volatility

Similar to the analysis of daily volatility with respect to the WSB subreddit tone and number of threads, we also examine the impact of both the subreddit variables on intraday realized volatility. The intraday realized volatility is calculated using the methodology specified in Andersen et al. [2007] using the Heterogeneous Autoregressive (HAR) model which also takes into consideration the jump component of the high frequency return volatility. The results are presented in tables 8 and 9. We find that both WSB subreddit thread tone and the thread count are significantly associated with intraday realized volatility.

In table 8, we regress intraday one-minute interval GME realized volatility on intraday one-minute interval WSB subreddit tone in the presence of the controls as specified in regression equation (3): the number of upvotes; and the number of comments.

Insert table 8 here.

The results are as follows: the intraday WSB subreddit tone coefficient displays significant predictive association at lags 1 and 4—i.e., up to 4 minutes in advance; the number of upvotes shows contemporaneous significance and also from lags 6–10 minutes uniformly; while the number of comments has no significant association with intraday GME volatility.

On a similar note, in table 9 we conduct regressions on intraday one-minute interval GME volatility on intraday one-minute interval WSB subreddit thread count in the presence of the usual intraday controls in line with equation (4): the number of upvotes and the number of comments.

Insert table 9 here.

The results are as follows: the intraday WSB subreddit thread count coefficient displays significance at the sixth one-minute interval—i.e., six minutes in advance; and the one-minute number of upvotes is significant contemporaneously and from minutes 6–10 in advance; while the number of comments shows no significance at any lag.

6 Robustness

We conduct a sequence of extensive robustness exercises in order to confirm whether our results—which suggest a significant predictive ability of WSB subreddit's tone and its thread count on GME returns as well as its volatility—are reliable.

These are primarily of the three following types: i) out of sample and sub-sample testing for the putative impact of WSB subreddit tone and thread count on GME returns, ii) inclusion of further controls which could be correlated with GME returns, and iii) testing the impact of WSB subreddit on other GME variables such as the put-call ratio.

In the following discussion we only elaborate on results regarding the WSB subreddit's tone on the dependent variables while noting that the exact same set of results is also obtained for the WSB subreddit's thread count which we do not display for brevity.

6.1 Out-of-sample analysis

We conduct out-of-sample analysis for the month of November 2020 when there was some interest but no major discussion on WSB subreddit about the GME stock yet. Owing to the very few daily trading days in November 2020, we resort to intraday analysis and regress the intraday GME stock returns on the intraday WSB subreddit tone in November 2020 along with the controls lagged return, number of upvotes and the number of comments.

Insert table 10 here.

The results show that even during the out-of-sample month of November 2020, there was a significant predictive association between the WSB subreddit's tone and the intraday GME return. Except for the number of comments, none of the other control variables had any noteworthy association in the month of November 2020.

6.2 Subsample Analysis

Our original results are based on the sample from December 1, 2020 to March 31, 2021. However, the most important period from the perspective of WSB subreddit activity as well as the short squeeze is during January and February 2021. Thus we select the period from January 1 2021 to February 28 2021 and test if the WSB subreddit's tone is still significantly predictive for the GME daily return.

Insert table 11 here.

The results are outlined in table 11 and are as follows: during the subsample of January 1, 2021–Februray 28, 2021, the WSB subreddit's tone displayed significant predictive association with GME daily returns up to

5 days in advance. This is in presence of the usual daily controls such as the number of upvotes, comments, GSI, Twitter Sentiment, one-day lagged return etc. in line with (1).

Among other variables, lagged one-day GME returns are significant at lag 4, the number of comments is significant at lags 4 and 5; and the GSI is significant at lag 4.

6.3 Additional control variables

6.3.1 Russell 2000 index returns

In order to forestall concerns that the benchmark stock market index returns are not controlled for, we include the Russell 2000 index returns—one of whose components is the GME stock—as an additional control. The results are discussed in table 12 and indicate that the WSB subreddit tone continues to be of predictive significance for the daily GME return for up to 5 days in advance.

Insert table 12 here.

The Russell index return itself is not significantly associated with movements in GME stock. Among other controls, lagged GME returns, the number of upvotes and comments; and the Google Search Index are those which exhibit predictive association with the GME daily return.

We do not include the detailed results for intraday GME returns' association with intraday WSB subreddit's tone but note that the results continue to indicate the importance of WSB subreddit's tone in predicting GME's intraday stock returns.

6.3.2 Price-to-book ratio

So far, none of the controls employed in this study have featured any variable based on the company's fundamentals. Since there are no fundamental variables at the intraday frequency, we resort of testing the price-to-book ratio for which we have observations at the daily frequency. We add the GME P/B ratio to the list of controls specified in equation (1). The results are collected in table 13.

Insert table 13 here.

The results are as follows: even in the presence of a fundamental variable, the GME P/B ratio, the daily WSB subreddit tone exhibits significant predictive association—4 to 5 days in advance—with the GME daily return. The P/B ratio is itself highly significant at all lags, while the Google Search Index also displays extensive predictive association with the daily GME return.

6.4 Impact of WSB tone on GME's put-call ratio

Apart from the equity markets, the derivatives markets also reacted strongly during the January-February short squeeze episodes [Allen et al., 2021]. In order to confirm that such developments in the derivatives markets can also be predicted by WSB subreddit discussions, we test if the thread tone affects GME's put-call ratio. If the put-call ratio is numerically high, it suggests a larger fraction of put options reflecting a more bearish scenario. On the other hand, a smaller numerical value of the put-call ratio suggests a higher fraction of call options reflecting a more bullish sentiment regarding the stock.

Insert table 14 here.

The results are compiled in table 14. As can be seen, even for the put-call ratio, the WSB subreddit displays significant predictive associations up to 2 days in advance. Among other variables, the number of upvotes and comments show mild significance at 5 lags, while the GSI shows strong significance uniformly from lags 0 to 4.

7 Concluding Remarks

We examine the unprecedented short-squeeze episode triggered by mass-coordinated buying of GME stock and call options by countless retail traders who pooled their efforts over the internet and discussed their trading strategies quite openly on several internet fora—most prominently on the subreddit `r/wallstreetbets`. We show that the tone as well as the volume of discussions on the subreddit have significant predictive association with the GME stock return and the GME stock volatility during the period December 2020–March 2021. Since app-based investing has lowered the barrier for entry for retail investors, one cannot discount such coordinated campaigns from influencing stock prices and volatilities in the future. Our paper offers among the first systematic investigations which confirms the popular narrative that activity on the WSB subreddit was of critical importance in influence the fortunes of GME in early 2021.

References

David Y Aharon, Renatas Kizys, Zaghum Umar, and Adam Zaremba. Did David win a battle or the war against Goliath? Dynamic return and volatility connectedness between the GameStop stock and the high short interest indices, 2021. Working Paper.

Khurshid Ahmad, JingGuang Han, Elaine Hutson, Colm Kearney, and Sha Liu. Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, 37:152–172, 2016.

Franklin Allen, Eric Nowak, Matteo Pirovano, and Angel Tengulov. Squeezing shorts through social news platforms, 2021. Swiss Finance Institute Research Paper Series No 21-31.

Abhinav Anand, Sankarshan Basu, Jalaj Pathak, and Ashok Thampy. Who moved the market? Analyzing the role of central bank speeches, 2021a. IIM Bangalore Research Paper No 622.

Abhinav Anand, Sankarshan Basu, Jalaj Pathak, and Ashok Thampy. Whose speeches impact European markets: ECBs' or the national central banks'?, 2021b. IIM Bangalore Research Paper No 640.

Torben G Andersen, Tim Bollerslev, and Francis X Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89(4):701–720, 2007.

Alina Andreevskaia and Sabine Bergler. When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. In *Proceedings of ACL-08: HLT*, pages 290–298, 2008.

Werner Antweiler and Murray Z Frank. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294, 2004.

Emanuele Bajo and Carlo Raimondo. Media sentiment and IPO underpricing. *Journal of Corporate Finance*, 46:139–153, 2017.

Simon Behrendt and Alexander Schmidt. The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, 96:355–367, 2018.

Gareth Campbell, John D Turner, and Clive B Walker. The role of the media in a bubble. *Explorations in Economic History*, 49(4):461–481, 2012.

Salim Chahine, Sattar Mansi, and Mohamad Mazboudi. Media news and earnings management prior to equity offerings. *Journal of Corporate Finance*, 35:177–195, 2015.

Iulia Cioroianu, Shaen Corbet, and Charles Larkin. The differential impact of corporate blockchain-development as conditioned by sentiment and financial desperation. *Journal of Corporate Finance*, 66:101814, 2021.

Jo Danbolt, Antonios Siganos, and Evangelos Vagenas-Nanos. Investor sentiment and bidder announcement abnormal returns. *Journal of Corporate Finance*, 33:164–179, 2015.

Sanjiv R Das and Mike Y Chen. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388, 2007.

Gregory W Eaton, T Clifton Green, Brian Roseman, and Yanbin Wu. Zero-commission individual investors, high frequency traders, and stock market quality, 2021. Working Paper.

Xunan Feng and Anders C Johansson. Top executives on social media and information in the capital market: Evidence from China. *Journal of Corporate Finance*, 58:824–857, 2019.

Henry L Friedman and Zitong Zeng. Retail investor trading and market reactions to earnings announcements, 2021. Working Paper.

Diego Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3): 1267–1300, 2013.

Manoochehr Ghiassi and S Lee. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, 106:197–216, 2018.

Matthew Jockers. *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*, 2017. R package version 1.0.6.

Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125, 2006.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.

Cheng Long, Brian M Lucey, and Larisa Yarovaya. "I just like the stock" versus "Fear and loathing on main street": The role of Reddit sentiment in the GameStop short squeeze, 2021. Working Paper.

Tim Loughran and Bill McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65, 2011.

Štefan Lyócsa, Eduard Baumöhl, and Tomáš Výrost. YOLO trading: Riding with the herd during the GameStop episode, 2021. Working Paper, Leibniz Information Centre for Economics, Kiel, Hamburg.

Philippe Massicotte and Dirk Eddelbuettel. *gtrendsR: Perform and display Google trends queries*, 2021. URL https://cran.r-project.org/web/packages/gtrendsR/index.html. R package version 1.4.8.

Saif Mohammad and Peter Turney. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA, jun 2010. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W10-0204.

Livia Polanyi and Annie Zaenen. Contextual valence shifters. In *Computing Attitude and Affect in Text: Theory and Applications*, pages 1–10. Springer, 2006.

Marc Schulder, Michael Wiegand, Josef Ruppenhofer, and Stephanie Köser. Introducing a lexicon of verbal polarity shifters for English. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*, volume 1, pages 1393–1397, 2018.

Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007.

Table 2: Impact of WSB subreddit on GME daily returns

	Dependent variable: GME daily returns					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	-0.0005	0.005^{**}	0.009^{***}	-0.006^{*}	0.011^{***}	-0.016^{***}
	(0.008)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
Ret Lag 1	-0.118	-0.137	-0.151	0.154	0.421^{**}	-0.119
	(0.198)	(0.172)	(0.157)	(0.195)	(0.173)	(0.155)
Num Upvotes	-0.00000^{**}	-0.00000	-0.00000	-0.00000	0.00000	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	-0.00001	-0.00000	0.00001	0.00000^{**}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	0.008	0.003	-0.004	0.012^{*}	-0.021^{***}	0.012^{*}
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.006)
Twitter Sent	5.367	6.056	5.466	2.379	2.031	3.632
	(4.640)	(4.384)	(4.048)	(4.842)	(4.063)	(4.220)
Constant	-0.039	0.013	0.046	-0.014	0.086	-0.012
	(0.118)	(0.104)	(0.098)	(0.113)	(0.095)	(0.105)
Observations	59	59	59	56	56	57

Note: This table presents the results from the regression of GME daily returns on WSB daily tone from December 2020 to March 2021 in line with equation (1). Controls include lag 1 GME daily return; num of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 3: Impact of WSB Thread Count on GME daily returns

	Dependent variable: GME daily returns					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Num Threads	-0.001^{*}	0.0005^{**}	0.001^{***}	-0.001^{*}	0.001^{***}	-0.001^{***}
	(0.001)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0003)
Ret Lag 1	0.140	−0.017	−0.036	0.183	0.497^{***}	−0.097
	(0.182)	(0.160)	(0.145)	(0.174)	(0.155)	(0.130)
Num Upvotes	-0.00000^{**}	−0.00000	−0.00000	−0.00000	0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00001	−0.00001	−0.00000	0.00001	0.00000^{**}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	0.002	−0.012	−0.008	0.009	-0.016^{*}	−0.008
	(0.012)	(0.009)	(0.009)	(0.014)	(0.009)	(0.009)
Twitter Count	0.00001^{*}	0.00001	0.00000	0.00000	−0.00000	0.00001^{*}
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Constant	−0.038	0.060	0.043	−0.016	0.056	0.036
	(0.114)	(0.108)	(0.108)	(0.128)	(0.099)	(0.103)
Observations	59	59	59	56	56	57

Note: This table presents the results from the regression of GME daily returns on WSB daily thread count from December 2020 to March 2021 in line with equation (2). Controls include lag 1 GME daily return; num of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 4: Impact of intraday WSB tone on GME intraday returns

	Dependent variable: GME intraday returns										
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
WSB Tone	0.002	0.002	0.001	0.002*	−0.0001	−0.002	0.002	0.002	0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ret Lag 1	0.138***	0.139***	0.140***	0.139***	0.140***	0.140***	0.138***	0.138***	0.138***	0.138***	0.138***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Num Upvotes	−0.00000	−0.00000	0.00000	0.00000	0.000	0.00000	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	0.00000	−0.00000	0.000	0.00000	−0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	−0.0003	−0.0002	0.0001	−0.0002	0.0001	0.0004	−0.0003	−0.0003	−0.0003	−0.0003	−0.0003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498

Note: This table presents the results from the regression of GME intraday 1-min interval returns on WSB intraday tone from December 2020 to March 2021 in line with equation (1). Controls include lag 1 GME intraday return; num of upvotes; number of comments; and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 5: Impact of intraday WSB thread count on GME intraday returns

	Dependent variable: GME intraday returns										
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
Num Threads	−0.00000	−0.00000	−0.00000	−0.00000	0.00000*	0.00000	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Ret Lag 1	0.139***	0.139***	0.140***	0.140***	0.141***	0.139***	0.139***	0.139***	0.139***	0.139***	0.139***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Num Upvotes	−0.00000	−0.00000	0.00000	0.00000	0.000	0.00000	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	0.00000	−0.00000	−0.000	0.00000	−0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	0.0005	0.001	0.001	0.0003	−0.001	−0.0004	0.0005	0.0005	0.0005	0.0005	0.0005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498	1,498

Note: This table presents the results from the regression of GME intraday 1-min interval returns on WSB intraday tone from December 2020 to March 2021 in line with equation (2). Controls include lag 1 GME intraday return; num of upvotes; number of comments; and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 6: Impact of WSB subreddit tone on GME daily volatility

	Dependent variable: GME daily volatility					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	−0.0004	0.002*	0.008***	−0.003*	0.002	−0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Num Upvotes	−0.00000	0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	−0.00000	−0.00000*	−0.00000*	0.00000	0.00001***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	0.014***	0.011***	0.005**	0.017***	0.008**	0.005
	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
Twitter Sent	0.754	0.756	−0.509	0.678	2.141	0.230
	(2.775)	(2.724)	(2.110)	(2.825)	(2.612)	(2.157)
Constant	−0.133*	−0.118	−0.080	−0.150*	−0.080	−0.041
	(0.078)	(0.072)	(0.056)	(0.076)	(0.069)	(0.062)
Observations	74	74	72	71	71	73

Note: This table presents the results from the regression of GME daily volatility on WSB daily tone from December 2020 to March 2021 in line with equation (3). Controls include number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 7: Impact of WSB subreddit thread count on GME daily volatility

	Dependent variable: GME daily volatility					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Num Threads	-0.0001	0.0002	0.001^{***}	-0.0004	0.0001	-0.00004
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Num Upvotes	-0.00000^{*}	0.00000	-0.00000	-0.00000	-0.00000	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comment	-0.00000	-0.00000	-0.00000^{*}	0.00000	0.00000^{***}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	0.002	0.002	0.004	0.012	0.006	-0.0003
	(0.006)	(0.006)	(0.005)	(0.008)	(0.006)	(0.005)
Twitter Count	0.00001^{**}	0.00001	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	-0.102	-0.086	-0.072	-0.135	-0.078	-0.023
	(0.076)	(0.074)	(0.060)	(0.081)	(0.072)	(0.062)
Observations	74	74	72	71	71	73

Note: This table presents the results from the regression of GME daily volatility on WSB daily tone from December 2020 to March 2021 in line with equation (4). Controls include number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 8: Impact of intraday WSB tone on GME intraday volatility

	Dependent variable: GME intraday volatility										
	Lag0	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
WSB Tone	0.0002	0.0004*	0.0003	0.0001	0.0004*	0.00002	0.0002	0.0002	0.0002	0.0002	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Num Upvotes	−0.00000**	−0.00000	−0.000	0.00000	0.00000	0.000	−0.00000**	−0.00000**	−0.00000**	−0.00000**	−0.00000**
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	0.00000	0.00000	−0.00000	−0.000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480

Note: This table presents the results from the regression of GME intraday 1-min interval volatility on WSB intraday tone from December 2020 to March 2021 in line with equation (3). Controls include the num of upvotes; number of comments; and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 9: Impact of intraday WSB thread count on intraday GME volatility

	Dependent variable: intraday GME volatility										
	Lag0	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
WSB Thread Count	−0.00000	−0.00000	−0.00000	−0.00000	−0.000	0.00000**	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Upvotes	−0.00000**	−0.00000	−0.000	0.00000	0.00000	0.000	−0.00000**	−0.00000**	−0.00000**	−0.00000**	−0.00000**
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	0.00000	0.00000	−0.00000	−0.000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	−0.00003	−0.00001	0.00000	−0.00000	−0.00004	−0.0002	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	1,476	1,476	1,476	1,476	1,476	1,476	1,476	1,476	1,476	1,476	1,476

Note: This table presents the results from the regression of GME intraday 1-min interval volatility on WSB intraday thread count from December 2020 to March 2021 in line with equation (4). Controls include the num of upvotes; number of comments; and the day of the week and month dummy. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 10: Impact of WSB subreddit on GME intraday return: Out-of-sample (Nov)

	Dependent variable: intrday GME returns										
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
WSB Tone	−0.002	−0.001	−0.0001	0.004**	0.0004	−0.001	−0.002	−0.002	−0.002	−0.002	−0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ret Lag 1	0.049	0.032	0.051	0.043	0.083	0.038	0.049	0.049	0.049	0.049	0.049
	(0.129)	(0.128)	(0.133)	(0.122)	(0.135)	(0.134)	(0.129)	(0.129)	(0.129)	(0.129)	(0.129)
Num Upvotes	−0.00001	−0.00001	0.00001	0.00002	−0.00001	0.00001	−0.00001	−0.00001	−0.00001	−0.00001	−0.00001
	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Num Comments	−0.00000	−0.00001	−0.00000	−0.00001	0.00002**	0.00001	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Constant	−0.001	−0.001	−0.002	−0.001	−0.003*	−0.003	−0.001	−0.001	−0.001	−0.001	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
h Observations	66	66	66	66	66	66	66	66	66	66	66

Note: This table presents the results from the regression of GME intraday one-minute interval returns on WSB intraday one-minute interval tone during November 2020 (out-of-sample) in line with equation (1). Controls include lagged returns; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly differentfrom zero at 1%, 5% and 10% respectively.

Table 11: WSB subreddit's impact on GME daily returns: Jan and Feb 2021

	Dependent variable: GME daily returns					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	0.006	0.006**	0.009**	−0.005	0.012***	−0.017**
	(0.013)	(0.003)	(0.003)	(0.005)	(0.004)	(0.008)
Ret Lag 1	−0.213	−0.112	−0.200	0.216	0.504*	−0.112
	(0.346)	(0.268)	(0.268)	(0.327)	(0.249)	(0.272)
Num Upvotes	−0.00000	−0.00000	−0.00000	−0.00000	0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)
Num Comments	−0.00000	−0.00001	−0.00000	0.00001	0.00001*	0.00001**
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00000)	(0.00000)
GSI	0.005	0.001	−0.003	0.010	−0.023**	0.014
	(0.009)	(0.006)	(0.006)	(0.009)	(0.009)	(0.010)
Twitter Sentiment	12.944	9.663	12.891	5.453	3.125	11.394
	(12.036)	(9.973)	(10.154)	(12.378)	(8.973)	(10.877)
Constant	0.122	0.158	0.205	−0.045	0.152	−0.011
	(0.238)	(0.187)	(0.199)	(0.251)	(0.166)	(0.216)
Observations	30	30	29	28	28	28

Note: This table presents the results from the regression of GME daily returns on WSB daily tone from January 2021 to February 2021 in line with equation (1). Controls include lagged returns; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticityand autocorrelation) robust. ***, ** and * indicate coefficients significantly different from zero at 1%, 5% and 10% respectively.

Table 12: WSB impact on GME daily returns: Additional control: Russell index

	Dependent variable: daily GME returns					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	-0.0001	0.005^{**}	0.009^{***}	-0.006^{*}	0.012^{***}	-0.015^{***}
	(0.008)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
Ret Lag 1	-0.132	-0.139	-0.149	0.141	0.474^{**}	-0.135
	(0.203)	(0.175)	(0.159)	(0.200)	(0.181)	(0.157)
Num Upvotes	-0.00000^{**}	-0.00000	-0.00000	-0.00000	0.00000	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	-0.00001	-0.00000	0.00001	0.00001^{**}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	0.007	0.003	-0.004	0.011^{*}	-0.022^{***}	0.011^{*}
	(0.006)	(0.005)	(0.004)	(0.006)	(0.007)	(0.006)
Twitter	5.446	6.084	5.456	2.439	1.719	3.836
	(4.688)	(4.436)	(4.092)	(4.896)	(4.074)	(4.250)
Russell Return	-1.046	-0.387	0.409	-0.957	2.428	-1.798
	(2.755)	(2.585)	(2.405)	(2.897)	(2.411)	(2.401)
Constant	-0.030	0.017	0.042	-0.006	0.073	0.007
	(0.121)	(0.107)	(0.101)	(0.117)	(0.096)	(0.108)
Observations	59	59	59	56	56	57

Note: This table presents the results from the regression of GME daily returns on WSB daily tone from January 2021 to February 2021 in line with equation (1). Controls include the Russell 2000 index returns; lagged GME daily returns; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly differentfrom zero at 1%, 5% and 10% respectively.

Table 13: Impact of WSB subreddit on GME daily returns: Additional control: P/B ratio

	Dependent variable: daily GME returns					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	-0.008	0.002	0.001	-0.0002	0.006^{**}	-0.013^{***}
	(0.006)	(0.001)	(0.003)	(0.003)	(0.003)	(0.004)
Lag Ret 1	-0.117	-0.241^{*}	-0.144	-0.126	0.130	-0.135
	(0.151)	(0.136)	(0.141)	(0.163)	(0.175)	(0.128)
Num Upvotes	-0.00000	-0.00000	0.00000	0.000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Num Comments	0.00000	-0.00001	-0.00000	-0.000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
GSI	-0.014^{**}	-0.019^{***}	-0.021^{***}	-0.021^{***}	-0.027^{***}	-0.008
	(0.006)	(0.005)	(0.006)	(0.008)	(0.006)	(0.007)
Twitter Sentiment	4.063	4.120	3.994	3.707	3.412	2.075
	(3.544)	(3.438)	(3.662)	(3.832)	(3.637)	(3.507)
PB Ratio	0.029^{***}	0.027^{***}	0.028^{***}	0.030^{***}	0.022^{***}	0.026^{***}
	(0.005)	(0.005)	(0.008)	(0.006)	(0.006)	(0.006)
Constant	0.001	0.050	0.064	0.065	0.058	-0.014
	(0.090)	(0.081)	(0.088)	(0.091)	(0.085)	(0.087)
Observations	59	59	59	56	56	57

Note: This table presents the results from the regression of GME daily returns on WSB daily tone from January 2021 to February 2021 in line with equation (1). Controls include the GME price-to-book ratio; lagged GME daily returns; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly differentfrom zero at 1%, 5% and 10% respectively.

Table 14: Impact of WSB subreddit on GME daily put-call ratio

	Dependent variable: daily GME put-call ratio					
	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
WSB Tone	−0.003	−0.001	0.009**	−0.008	0.002	0.004
	(0.019)	(0.005)	(0.004)	(0.007)	(0.008)	(0.013)
Num Upvotes	0.00000	0.00001	0.00000	0.00001	0.00002	0.00002**
	(0.00000)	(0.00000)	(0.00000)	(0.00001)	(0.00001)	(0.00001)
Num Comments	0.00000	−0.00000	0.00001	0.00001	0.00000	0.00001*
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
GSI	0.039**	0.037***	0.033***	0.050***	0.036*	0.025
	(0.016)	(0.012)	(0.008)	(0.014)	(0.019)	(0.017)
Twitter Sentiment	−4.995	−7.381	−0.799	−4.243	−2.226	1.156
	(11.920)	(11.954)	(7.541)	(11.597)	(12.320)	(11.345)
Lag Ret 1	0.238	0.312	0.096	0.505	0.286	−0.160
	(0.509)	(0.470)	(0.292)	(0.466)	(0.525)	(0.416)
Constant	0.292	0.309	0.293	0.197	0.261	0.424
	(0.304)	(0.283)	(0.182)	(0.271)	(0.288)	(0.282)
Observations	59	59	59	56	56	57

Note: This table presents the results from the regression of GME daily put-call ratio on WSB daily tone from January 2021 to February 2021 in line with equation (1). Controls include lagged GME daily returns; number of upvotes; number of comments; daily Google Search Index (GSI); daily Twitter Sentiment (Twitter Sent) and the day of the week and month dummies. The standard errors (reported in parentheses) are HAC (heteroskedasticity and autocorrelation) robust. ***, ** and * indicate coefficients significantly differentfrom zero at 1%, 5% and 10% respectively.