"I Just Like the Stock": The Role of Reddit Sentiment in GameStop

Suwan(Cheng) Long[a,e,*], Brian Lucey [a,b,c,g], Ying Xie [f], Larisa Yarovaya [d]

[a] *Trinity Business School, Trinity College Dublin, Dublin 2, Ireland*

[b] *Distinguished Research Fellow, Institute of Business Research, University of Economics Ho Chi Minh City, 59C Nguyen Dinh Chieu, Ward 6, District 3, Ho Chi Minh City, Vietnam*

[c] *Institute for Industrial Economics, Jiangxi University of Economics and Finance, 169, East Shuanggang Road, Xialuo, Changbei District 330013 Nanchang, Jiangxi, China*

[d] *Southampton Business School, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, United Kingdom*

[e] *Judge Business School, University of Cambridge, Trumpington St, Cambridge CB2 1AG, United Kingdom*

[f] *Cranfield School of Management, Cranfield University, College Road, Cranfield, MK43 0AL, United Kingdom*

[g] *Distinguished Visiting Research Professor, Abu Dhabi University, Zayed City - Abu Dhabi - United Arab Emirates*

[*] *Email address: longch@tcd.ie(Suwan Long)*

Abstract

This paper investigates the role, if any, played by the social media platform Reddit in the events around the GameStop(GME) share rally in early 2021. In particular, we analyse the impact of discussions on the r/WallStreetBets subreddit on the price dynamics of the American online retailer GameStop. We customise a sentiment analysis dictionary for Reddit platform users based on the VADER sentiment analysis package and perform textual analysis on 10.8 million comments. The analysis of the relationships between Reddit sentiments and 1-minute, 5-minute, 10-minute, and 30-minute GameStop returns contribute to the growing body of literature on 'meme stocks' and the role of the discussions on investment forums on intraday stock price movements.

Keywords: Gamestop; Reddit; Robinhood; media sentiments; meme stocks; herding; textual analysis; sentiment analysis
*JEL Code:*D91, G41

1. Introduction

In early 2021 shares of the American video game retailer GameStop surged more than 700% in one week following the speculative involvement of individual investors, a move touted as investors piling in to buy the stock that they like. The *"to the moon"* movement coordinated on r/WallStreetBets subreddit during the GameStop share rally provides an opportunity to add to the literature on the market impact of online discussion at investment forums (Renault [2017], Cookson and Niessner [2019], Corbet et al. [2021]). The impact of investment sentiments extracted from Reddit on stock prices is particularly interesting topic for textual analysis since investors are directly creating the content of the forums and influencing each other rather than reacting to any published news. In this paper, we aim to add to the media-sentiment literature by designing a Reddit-specific investment lexicon and testing it using GameStop intraday data.

The GameStop and related social media moderated investment "pile-on" was extensively discussed in broadcast and print media, perhaps both reporting on and fuelling the perception of small investors. Media sentiments have previously been analysed using news-aggregation databases (e.g., Chahine et al. [2015]; Ahmad et al. [2016]; An et al. [2020]), printed news papers (Bajo and Raimondo [2017]), Yahoo Finance and Raging Bull message boards (Antweiler and Frank [2004]) and social media platforms, such as Twitter (Behrendt and Schmidt [2018], Al Guindy [2021]), Facebook (Danbolt et al. [2015]), Weibo (Feng and Johansson [2019]), Stock Twits, (Cookson and Niessner [2019]), amongst others. In our research, we extract investment discussion related sentiments from Reddit, and analyse how the tone and timing of the r/WallStreetBets subreddit's posts affect the share price movements. We collect 10.8 million comments from the Reddit platform and analyse the impact of the message board sentiment on intraday GameStop stock prices for the period from January 1st to 28th February 2021.

The results of previous studies show diverse effects of tone and timing of media coverage on corporate performance and stock markets returns. The impact of media is found to be especially pronounced during periods of stock price explosivity, i.e. asset prices bubble (Campbell et al. [2012]), and during the stressed market conditions of a recession (Garcia [2013]), while the impact of positive and negative sentiments can be time varying and asymmetric. Furthermore, it is evident that media coverage affects the performance of highly speculative assets, such as cryptocurrencies (e.g., Corbet et al. [2020]; Guegan and Renault [2021]). Cioroianu et al. [2021] considered the impact of social media "hype" on short-term profitability of the firm around blockchain related announcements. Their results show that investors were subject to a very sophisticated form of asymmetric information which in turn led to market euphoria, contributing to the findings of Akyildirim et al. [2020]. Danbolt et al. [2015] analysed the impact of sentiments extracted from Facebook on bidder announcement returns and showed that uninformed traders would be the most susceptible of all investors to social media sentiments. This is in line with the behavioural finance literature on investment overconfidence and other cognitive biases in capital markets (Daniel et al.

[2002]).

There have already been some analyses on the Robinhood/Gamestop experience that are related to ours. Corbet et al. [2021] examines Reddit social media posts with a background of the GameStop situation to examine possible stock price manipulation. Umar et al. [2021b] examine fundamentals sentiment for Gamestop but use twitter data as a proxy for sentiment. In our paper, we directly measure sentiment in the (widely agreed) main social media platform influencing the GameStop price. While Eaton et al. [2021] consider the causal effects of the Robinhood trading platform on financial markets, in this paper, we focus on the impact of meme-driven culture and extract investors sentiments from the r/WallStreetBets subreddit posts. Hu et al. [2021] analysed the impact of Reddit on the 'Robinhood 50' stocks, including GameStop, and demonstrated strong and significant impacts of positive comments on daily stock returns. In our paper, we focus on GameStop stock performance only, considering evidence from high-frequency data.

The GameStop share rally attracted significant attention of academic scholars to analysis of investors' behaviour on online trading forums, such as Reddit. However, there is no paper available to date that offer a specific lexicon for Reddit forums that sufficiently explain the underlying mechanism behind the observed impacts on stock markets. The novelty of this paper is twofold. First, we introduce a unique Reddit lexicon that can be used by scholars to extract the investment sentiments from this platform. Before GameStop, Reddit was not widely considered as a platform that can influence the stock markets. In meme culture, emotions are often expressed with images, videos, emojis, while the text messages in Reddit often contain offensive terms, slang, as well as significant larger volume of data that would be highly challenging to analyse in standard academic software (e.g., Das and Chen [2007]) or using existing lexicons employed by popular textual analysis packages. In order to extract sentiments from Reddit messages, we design a unique lexicon containing specific terms that are frequently used on r/WallStreetBets subreddit. This offers a better tool for sentiment analysis on Reddit than any existing approaches. Thus, in this paper, we provide a first attempt to capture these sentiments using a unique Reddit lexicon, establishing a foundation for future research in this area. Second, this paper contributes to the previous literature on impact of social media on high-frequency intraday stock prices. Our results indicate that Reddit sentiments affect GME intraday returns, and demonstrate the relationships between specific tone and number of comments and 1-min GME returns, confirming the (limited but real) role of r/WallStreetBet's discussion in GameStop's price rally.

The remainder of this paper is organised as follows. Section 2 specifies the research hypotheses. Section 3 explains data and methodology. Section 4 discusses the empirical results, while Section 5 concludes providing some directions for future research.

2. Why might sentiment matter?

The analysis of behaviour of subgroups of investors and their speculative activity on financial markets has a long history. From (Shiller [1990]; Shiller [2014]), many scholars have focused on this research question. For example, the impact of retail traders on stock price formation has been previously explored by Barber and Odean [2002], Kumar and Lee [2006] (2006), Dorn et al. [2008], among others. Barber and Odean [2002] explains the overconfidence and higher trading activity of online investors by self-attribution bias and illusions of knowledge and control, while Kumar and Lee [2006] further highlighted the key role of investor sentiment. Dorn et al. [2008] distinguished between two different types of retail traders: speculative and other traders, showed that retail speculators as a group behave as positive feedback traders. The impacts of feedback trading and herding behaviour have been further considered by Nofsinger and Sias [1999], who reported that institutional investors a play more influential role in financial markets. These papers are in line with early ideas expressed by Shleifer and Summers [1990] on irrational behaviour of noise traders and the limits of arbitrage, as being superior to the efficient market hypothesis. (Fama [1970]).

The role of social media in stock price formation is increasingly of interest to researchers. Rantala [2019] examined how investment ideas transmit among retail investors via social interactions considering the inviter-invitee relationships as a Ponzi scheme, showing the power of the word-of-mouth. Eliaz and Spiegler [2020] using Bayesian networks empirically showed that people are drawn to hopeful narratives. Cookson and Niessner [2019] explored an investment disagreement using >18 million messages from StockTwits website, and demonstrated that more than half of investor disagreement have been driven by differences in investment philosophies. In contrast to above mentioned studies, the large body of literature on investment sentiments consider social media platforms as a channel of dissemination of corporate information, which might be especially beneficial for small firms that have lower analyst coverage (e.g., Feng and Johansson [2019], Al Guindy [2021]). Therefore, the majority of empirical papers on investment sentiments do not distinguish between different groups of investor publishing their posts on social media, and consider the impact of sentiments extracted from social media as any other variable or factor affecting share price. This approach is closely related to increasingly popular news-based indices that have been constructed using big data collected from various news aggregated platforms (e.g.Lucey et al. [2021]), and then being used as a predictor of financial market performances.

Social media sentiment is found to be a powerful tool in forecasting stock market returns. For example, Gu and Kurov [2020] reports that Twitter contains information that is not fully reflected in the share price, showing the ability of the Twitter sentiment index constructed by Bloomberg to predict Russell 3000 returns. Liang et al. [2020] compared the predictive ability of three sentiment indexes using positive and negative social media posts, newspaper news and Internet media news, and shows that while traditional newspapers have no impact on Chinese stock markets, both social media and Internet news have strong predictive power. Furthermore, the findings by Dong and Gil-Bazo [2020], who constructed media sentiment measures using more than 58 million social media messages in China, suggests that stock returns are mainly driven by positive sentiment and amateur investors. The impact of social media sentiments on intraday stock returns has been examined by Broadstock and Zhang [2019] using Twitter data, Sun et al. [2016] using Thomson Reuters MarkPsych sentiment data, and Renault [2017] using the StockTwits microblogging platform, among others. It was found that investor sentiment can predict intraday market return throughout the day and especially during the last two trading hours (Sun et al. [2016]), where this predictive power cannot be explained by lagged macroeconomic fundamentals and news. Renault [2017] constructed a lexicon of words used by online investors on StockTwits, and shows that the first half-hour change in investors sentiment can be used to forecast the last half-hour market returns.

Considering the nature and main features of Reddit platform, we have limited opportunity to split investors into specific groups, therefore we utilise an alternative approach to Dorn et al. [2008], Rantala [2019] or Cookson and Niessner [2019], and make an assumption that all r/WallStreetBets participants are speculative retail traders. Some existing evidence suggests the importance of both *timing* and *tone* of the media-extracted investment sentiments (e.g., Ahmad et al. [2016]; You et al. [2017]; Liu and Han [2020]). For example, Ahmad et al. [2016] examined the relationships between media-expressed firm-specific tone and firm-level returns, showing that the effect of negative media tone varies from significant effect to no effect at all. Therefore there are strong reasons to believe that the tone of the Reddit sentiments will be time-varying as will be any predictive power of the sentiment. Umar et al. [2021a] used Twitter publication count as a proxy of media sentiments in analysis of GameStop share rally and the role of Reddit amateur in it. However, Twitter Count will be not an accurate proxy to capture Reddit's sentiments. Therefore, in this paper we consider comments published specifically on the r/WallStreetBets forum.

Reddit's comments are organised in *Threads* with titles and some of those threads become more popular and receive higher levels of up-votes and higher numbers of comments from the forum's participants. Thus the number of comments in a thread can be used a determinant of the popularity of this discussion among redditors, while comments' scores additionally can indicate the popularity of the specific comment. Reddit has a very unique design that allows to better understand the mechanism of "hype" creation on investment forums. Thus, our empirical investigation will help to shed a light on the role of Reddit sentiments in asset price dynamics (Campbell et al. [2012]). While Garcia [2013] report an effect of negative media tone during the recession, Sun et al. [2016] report stronger predictive power of high-frequency investment sentiment during the economic expansion, which is consistent with several studies which highlighted the role of investment mania during periods of stock market explosivity.

3. Data and methodology

3.1. Data

This paper utilises high-frequency stock prices data for GameStop (GME) from the 1st January 2021 to 28th February 2021. The GME data and Russell 2000 index (of which GME is a constituent) are collected at a 1-min intervals from Bloomberg.

Shown in Figure 1 are the price dynamics of GameStop and the R2000 index, of which it is a constituent, over the two months of the analysis.

[Figure 1 here]

For sentiment analysis, we collected 10.8 million comments from the r/WallStreetBets subreddit for the same observation period, i.e. first two calendar months of 2021.

Table 1 shows number of comments collected for each group out of total 10.8 million comments.

[Table 1 here]

3.2. Methods

We begin our analysis by extracting sentiments from text messages scrapped from r/WallStreetBets subreddit. Text Sentiment Analysis is a trending field with substantial amount of academic research behind it. There are two main approaches employed in the existing literature: the lexical approach and the machine learning approach. Lexical approaches aim to map words to sentiment by building a lexicon or a 'dictionary of sentiment'. We can use this dictionary to assess the sentiment of phrases and sentences, without the need of looking at anything else. Lexical approaches look at the sentiment category or score of each word in the sentence and decide what the sentiment category or score of the whole sentence is. This approach has been utilized by Loughran and McDonald [2011]; Renault [2017], among others. Machine learning approaches, on the other hand, look at previously labelled data in order to determine the sentiment of never-before-seen sentences. The machine learning approach involves training a model using previously seen text to predict/classify the sentiment of some new input text.

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a sentiment package that is explicitly sensitive to feelings expressed in social media. VADER relies on a lexicon and five general rules to map lexical features to sentiment scores (Hutto and Gilbert 2014). When compared with feature-based machine learning methods, VADER carries a few advantages: 1) VADER was designed with a focus on social media data, with an emphasis on the rules that captured the essential meaning of social media texts. The lexicon and rules used by VADER are directly accessible and can be easily inspected and updated in discipline context. 2) VADER does not require any training data, and it utilizes a human-validated sentiment lexicon and general rules that are related to grammar and syntax. Threads and comments posted in Reddit are usually short sentences with emojis and rich use of punctuation, hence, VADER was selected to conduct sentiment analysis. However, the lexicon developed by Hutto and Gilbert [2014] for VADER is not specific for finance discipline, thus some key financial words are excluded, such as "Bear" and "Bull". Key phrases and hashtags unique to financial social media platform are also missing from the VADER lexicon, such as "Diamond Hands". Directly applying VADER to analyses sentiment of r/WallStreetBets suReddit may misclassify words when gauging tone in financial applications.

Using LDA modeling, firstly we classified the threads and comments into 49 topics, then we created a list of 130 most common words used on r/WallStreetBets by analysing the content and key information in the 49 topics. The number of LDA modelling topics was determined in such as way that the topic coherence score was maximised. To construct and validate the new VADER lexicon, a human-centered approach was adopted and 10 annotators were hired to manually estimate the sentiment valence (intensity) of each keyword. This approach is consistent with the methods used by Hutto and Gilbert [2014] in developing the original VADER lexicon. We updated the lexicon in VADER by 1) adding new r/WallStreetBets subreddit words and corresponding valence scores to the original lexicon; and 2) replacing the original valence scores with new valence scores if some words already exist in VADER. The updated valence scores are in the range of [-4, +4], with [-4] being Extremely Negative, [0] being Neutral, and [+4] being Extremely Positive. The results of this are shown in Table 2 and Table 3

The *SentimentIntensityAnalyser* object from the VADER package[1] is used to extract the *polarity_scores*. *polarity_scores* provide the overall sentiment metrics(compound score) for the comments. The compound score is computed by taking the sum of the valence scores of each word in the lexicon, adjusted according to the five general rules defined by Hutto and Gilbert [2014], and then normalized to be between -1 (most negative) and +1 (most positive). When the compound score is greater than 0.05, it denotes a positive sentiment. When the compound score is less than -0.05, it denotes a negative sentiment. For the compound score lies between 0.05 and -0.05, it denotes a neutral sentiment. In addition to the compound score, VADER *SentimentIntensityAnalyser* also returns Positive, Negative and Neutral scores for a text. These scores are calculated as the sum of Positive, Negative and Neutral valence scores in the lexicon, respectively. Taking the difference between the Positive and Negative measures, we obtain the net sentiment score of a text. A new variable NET is defined as the sum of the net sentiment scores of the texts posted in the time period, for example, within a 5-minute period. We also introduce the variable $AVERAGE$ that is the mean value of the net sentiment scores of all the texts posted in the time period. VADER sentiment analysis package has been employed widely across different disciplines, see Pano and Kashef [2020], Shelar and Huang [2018], Sivasangari et al. [2018], and Oliveira et al. [2016].

After sentiment variables have been constructed, we analyse the impact of NET and $AVERAGE$

[1] See for details: https://www.nltk.org/api/nltk.sentiment.html

sentiments on the GME intraday prices using the standard Granger Causality approach (Granger [1969])[2].

4. Results

4.1. GameStop overview

Short-selling strategies are commonly used by large institutional investors and hedge funds, that have both substantial investment and human capital to affect the market. In anticipation of a stock price decline, a large hedge fund Melvin Capital opened short positions against GameStop shares, a game and gaming retailer that was one of the many suffering during the COVID-19 induced economic disruption. On average GameStop shares were at that time traded around $7 per share, however, the company experienced a noticeable increase in share price creating sufficient volatility to perform short selling strategy by the hedge fund. The prevailing argument then runs that a coordinated movement led by retail investors on r/WallStreetBets forum caused a rapid increase in GME's share price. The outcome was particularly bad for Melvin Capital, which required nearly $3b in additional capital, with overall short squeeze lost nearing $25b[3]. Figure 1 shows the dynamics of GME shares in comparison to Russell 2000 for the period from January 2020 to March 2021, showing the GME roller-coaster ride in January-February 2021.

[Figure 1 here]

The GameStop case received enormous public attention and was widely discussed in the media. The gamification of trading and increased ease of access to financial markets by retail investors via online trading platforms such as Robinhood, spiked extensive debates and was followed by a sequence of congregational hearings and lawsuits. This case showed the growth of the decentralised financial system and technology and its potential to destabilise financial markets, therefore the GameStop case become significant from a policy perspective.

Apart from GameStop other *"meme stocks"* and assets were targeted by Reddit's amateurs, which suggest that the GameStop phenomenon uncovered a new channel for potential market manipulation - Reddit. In comparison to Twitter, Reddit is much more chaotic platform, and extracting sentiments from the subreddits is a challenging semantic problem. As it was pointed earlier, the majority of papers on media sentiments use aggregated news or social media platforms, while only a few papers have actually targeted some micro blogging trading platforms, such as StockTwits (Oliveira et al. [2016]; Renault [2017]), the content of which, however, is still highly correlated with

[2]Considering that this approach is well-known and used widely in the finance literature, for space consideration we do not include detailed specifications of the method, and the details are available upon request.

[3]see https://www.ai-cio.com/news/gamestops-robinhood-boosters-clobbering-hedge-funds/ for a comprehensive overview

Twitter due to high degree of integration between two platforms. Considering the platform design, Reddit might contain unique sentiments which were not captured by Twitter. Furthermore, the lexicon of redditors differs from other forums, since Reddit in itself is a manifestation of meme culture where messages are absurd and often offensive, as demonstrated bu Figure 2. Therefore a better understanding of investors' lexicon used by redditors can help to improve a quality of sentiment analysis of other social media platforms.

[Figure 2 here]

4.2. A new lexicon and sentiments in r/WallStreetBets

A unique Reddit-specific lexicon is designed following three-step process. First, all threads from r/WallStreetbets subbreading have been classified into 49 topics using LDA modelling. Second, a list of 130 most commonly used words has been created, and 10 annotators have been asked to manually rank the sentiment valence of each word. Table 2 shows the list of words in the new lexicon and their respective scores. The human-central approach to sentiment valence ranking has been particularly helpful to provide score for jargon terms, for example, 'to the moon' received [+3.5] score,'yolo' dimond hand' [+2.4] and [+3] respectively. To express negative sentiments various curse words have been used most often in addition to standard negative words, such as 'loss' [-2.5], 'wrong' [-1.8] and 'fake' [-2.3].

[Table 2 here]

Third, the existing lexicon in VADER has been updated by adding new Reddit-specific words with their corresponding valence scores as well as updating the scores of the words in the original VADER lexicon. Table 3 demonstrate an example of the updating process. New scores not only show the differences in the intensity of the sentiment, but also often display the change in tone of the sentiment from negative to positive, for example, 'crazy' from [-1.4] to [0.7]. Shown in Figures 3 and 4 are some overall sentiment measures. These plot the frequency and intensity of sentiment over the time-frame of the sample, across different sampling frequencies. We see a major burst of sentiment (net positive as per Figure 4) in the middle and a slightly smaller burst of net positive at the end of the sample period, As can be seen from Figures 5 and 6 a simple overlay of the net sentiment value against GME price suggests a relationship with a weaker against returns.

[Table 3 here]

[Figure 3 here]

[Figure 4 here]

4.3. The impact of Reddit net sentiment on GME

We begin our analysis of the impact of Reddit sentiments on GameStop by plotting the dynamics of the *NET* Sentiment (Positive - Negative), open and closing GME prices, Open-to-Open and Close-to-Close returns, and trading voulme at 1-min, 5-min, 10-min anf 30-min frequencies. Figure 5 displays the results for 1-min frequency. According to Figure 5(a) the relationships between the *NET* Sentiments and Opening and Closing GME prices at 1-min frequency are much stronger during the up market days, i.e. from 20 January 2021 to 27 January 2021. The engagement of Redditors with the discussion forum affect opening and closing prices specifically during the bullish market, however, when GME stock turned bearish, we can see that 1-min sentiments and opening and closing prices are decoupled from each other and moved in opposite directions. It is particularly visible on Figure 5(a) during the period after 31st of January 2021, where the spike in *NET* sentiments occurred (Positive>Negative) but GME price continued falling. This implies that positive comments and encouragements to hold GME stocks on Reddit were not able to stop or prevent this downturn movement of its price. Similar patterns are identified for Open-to-Open and Close-to-Close at 1-min frequency, as shown by Figure 5(b). There is a second period of stronger positive relationships between the *NET* Sentiment and GME returns during the up market movement at the end of February 2021, however, similarly, the linkages are weakened during the down market days. These are in line with the results obtained for 1-min total trading volumes, showing that the discussions on Reddit forum during the down market movements is only weakly linked to trading volume of GME stocks.

[Figure 5 here]

These patterns could be observed more clearly using 5-min data as it shown by Figure 6. There are clear break points in relationships between *NET* Sentiment and GME prices and returns, however, during the bearish market the positive linkages are weakened. These results shed a light on the mechanism of sentiment formation at r/WallStreetBets and its impact on GME share rally, and show that Reddit discussion helped to spike the interest in buying GME stock, but incapable to maintain it when price going back to its fair value. For example, Figure 5(c) shows linkages between trading volume and *NET* Sentiments at the beginning of the estimation period, but not in the period around 31st of January when market already turned bearish.

[Figure 6 here]

These results are consistent even for lower frequencies, and Figures 7 and 8 display similar patterns. At 10-min and 30-min frequencies we can clearly observe how the relationships intensify specifically during two short periods of GME price increase. However, outside of these two periods, the growth in *NET* Sentiments, i.e. increase in number of positive comments over negative comments in r/WallStreetBets, the relationships were weaker.

[Figures 7 and 8 here]

To further examine the causal linkages between NET Sentiments and GME returns we employ the standard Granger causality test (Granger [1969]), and the results are presented in Table 4. The findings show that at 1-min frequency GME Open-to-Open returns affect *NET* sentiments and *AVERAGE* sentiments, but not the other way around. Neither *NET* nor *AVERAGE* sentiments caused any affect on GME 1-min returns. The impact of *NET* sentiments on GME return become significant only at lower frequencies, and we can see clear evidence of positive impact for 5-min, 10-min, and 30-min data. However, the causal linkages between *AVERAGE* sentiment and GME returns have not been identified at these frequencies.

[Table 4 here]

5. Conclusions

This paper investigates the impact of sentiments extracted from Reddit on GameStop's intraday returns. Using 10.8 million comments from the r/WallStreetBets subreddit for the period from the 1st of January to 28th February 2021, this paper introduces a unique Reddit-specific lexicon with the corresponding sentiments scores obtained using human-centred approach. The lexicon's aim is to support researchers conducting textual analyses of investment sentiments derived from Reddit. The power of the lexicon to capture investment sentiments on social media platform has been examined using the GameStop share rally in early 2021.

The impact of *NET* Investment sentiment on 5-minute, 10-minute and 30-minute GME open-to-open returns have been identified. However, there was no impact of *AVERAGE* Sentiments on GME performance. For 1-minute data, our results suggest the opposite direction of causality, GME returns Granger cause both *NET* and *AVERAGE* sentiments. Analysis of the dynamics of the *NET* Senitments and GME prices, returns and trading volume, further show that linkages tend to be stronger during the bullish market in comparison to bearish. This implies the role of Reddit in transmitting positive sentiments during up market movements, however, a weak power of social media to influence markets during down market movements and stop share's downfall. Hence, this paper uncovers limited, but real, role of Reddit in GameStop share share rally.

This paper contributes to the literature on the impact of social media sentiments on stock prices providing a novel empirical evidence from Reddit. Our results will be of interest to institutional and retail investors, as well policy makers, academics, and media, since they shed a light on 'meme stocks' phenomenon. While the impact of Reddit sentiments on stock market is confirmed, our results issue a warning to individual investors that social media discussions might not be able to protect their investments when the market turns bearish. Even if r/WallStreetBets and other investments forums shows the growing power of retail investors to act in an organised manner, the influence is still not strong enough to protect individual investors from losses at the highly speculative 'meme' stock markets.

References

Ahmad, K., J. Han, E. Hutson, C. Kearney, and S. Liu (2016). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance 37*, 152–172.

Akyildirim, E., S. Corbet, D. Cumming, B. Lucey, and A. Sensoy (2020). Riding the wave of crypto-exuberance: The potential misusage of corporate blockchain announcements. *Technological Forecasting and Social Change 159*, 120191.

Al Guindy, M. (2021). Corporate twitter use and cost of equity capital. *Journal of Corporate Finance (forthcoming)*.

An, Z., C. Chen, V. Naiker, and J. Wang (2020). Does media coverage deter firms from withholding bad news? evidence from stock price crash risk. *Journal of Corporate Finance 64*.

Antweiler, W. and M. Frank (2004). Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance 59*(3), 1259–1294.

Bajo, E. and C. Raimondo (2017). Media sentiment and ipo underpricing. *Journal of Corporate Finance 46*, 139–153.

Barber, B. and T. Odean (2002). Online investors: Do the slow die. *Review of Financial Studies 15*, 455–487.

Behrendt, S. and A. Schmidt (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. *Journal of Banking and Finance 96*, 355–367.

Broadstock, D. and D. Zhang (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters 30*, 116–123.

Campbell, G., J. Turner, and C. Walker (2012). The role of the media in a bubble. *Explorations in Economic History 49*, 461–481.

Chahine, S., S. Mansi, and M. Mazboudi (2015). Media news and earnings management prior to equity offerings. *Journal of Corporate Finance 35*, 177–195.

Cioroianu, I., S. Corbet, and C. Larkin (2021). The differential impact of corporate blockchain-development as conditioned by sentiment and financial desperation. *Journal of Corporate Finance 66*.

Cookson, J. and M. Niessner (2019). Why don't we agree? evidence from a social networkof investor. *Journal of Finance 75*(1), 173–228.

Corbet, S., G. Hou, Y. Hu, and L. Oxley (2021). We reddit in a forum: The influence of messaging boards on firm stability. *Available at SSRN 3776445*.

Corbet, S., C. Larkin, B. Lucey, A. Meegan, and L. Yarovaya (2020). The impact of macroeconomic news on bitcoin returns. *The European Journal of Finance 26*, 1396–1416.

Danbolt, J., A. Siganos, and E. Vagenas-Nanos (2015). Investor sentiment and bidder announcement abnormal returns. *Journal of Corporate Finance 33*, 164–179.

Daniel, K., D. Hirshleifer, and S. Teoh (2002). Investor psychology in capital markets: evidence and policy implications. *Journal of Monetary Economics 49*, 139–209.

Das, S. R. and M. Chen (2007). Yahoo! for amazon: sentiment extraction from small talk on the web. *Management Science 53*(9), 1375–1388.

Dong, H. and J. Gil-Bazo (2020). Sentiment stocks. *International Review of Financial Analysis 72*.

Dorn, D., G. Huberman, and P. Sengmueller (2008). Correlated trading and returns. *Journal of Finance 63*, 885–920.

Eaton, G., T. Green, B. Roseman, and Y. Wu (2021). Zero-commission individual investors, high frequency traders, and stock market quality. *Working paper, Available at SSRN: https://ssrn.com/abstract=3776874*.

Eliaz, K. and R. Spiegler (2020). A model of competing narratives. *American Economic Review 110*, 3786–3816.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance 25*(2), 383–417.

Feng, X. and A. Johansson (2019). Top executives on social media and information in the capital market: Evidence from china. *Journal of Corporate Finance 58*, 824–857.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance 68*(3), 199–242.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 44–438.

Gu, C. and A. Kurov (2020). Informational role of social media: Evidence from twitter sentiment. *Journal of Banking and Finance 121*.

Guegan, D. and T. Renault (2021). Does investor sentiment on social media provide robust information for bitcoin returns predictability? *Finance Research Letters 38*, 116–123.

Hu, D., C. M. Jones, V. Zhang, and X. Zhang (2021). he rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery. *Available at SSRN: https://ssrn.com/abstract=3807655*.

Hutto, C. and E. Gilbert (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media 8*(1), 216–225.

Kumar, A. and C. Lee (2006). Retail investor sentiment and return comovements. *The Journal of Finance 61*, 2451–2486.

Liang, C., L. Tanga, Y. Lia, and Y. Wei (2020). Which sentiment index is more informative to forecast stock market volatility? evidence from china. *International Review of Financial Analysis 71*.

Liu, S. and J. Han (2020). Media tone and expected stock returns. *International Review of Financial Analysis 70*.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis,dictionaries, and 10-ks. *The Journal of Finance 66*, 35–65.

Lucey, B. M., S. A. Vigne, L. Yarovaya, and Y. Wang (2021). The cryptocurrency uncertainty index. *Finance Research Letters*, 102147.

Nofsinger, J. R. and R. Sias (1999). Herding and feedback trading by institutional and individual investors. *The Journal of Finance 54*, 2263–2295.

Oliveira, N., P. Cortez, and N. Areal (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems 85*, 62–73.

Pano, T. and R. Kashef (2020). A complete vader-based sentiment analysis of bitcoin (btc) tweets during the era of covid-19. *Big Data and Cognitive Computing 4*(4).

Rantala, V. (2019). How do investment ideas spread through social interaction? evidence from a ponzi scheme. *The Journal of Finance 74*, 2349–2389.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. stock market. *Journal of Bankinga and Finance 84*, 25–40.

Shelar, A. and C.-Y. Huang (2018). Sentiment analysis of twitter data. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1301–1302.

Shiller, R. J. (1990). Market volatility and investor behavior. *The American Economic Review 80*, 58–62.

Shiller, R. J. (2014). Speculative asset prices. *The American Economic Review 104*, 1486–1517.

Shleifer, A. and L. Summers (1990). The noise trader approach to finance. *The Journal of Economic Perspectives 4*, 19–33.

Sivasangari, V., A. K. Mohan, K. Suthendran, and M. Sethumadhavan (2018). Isolating rumors using sentiment analysis. *Journal of Cyber Security and Mobility*.

Sun, L., M. Najand, and J. Shen (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking and Finance 73*, 147–164.

Umar, Z., M. Gubareva, I. Yousaf, and S. Ali (2021a). Sentiment during recessions. *Journal of Behavioural and Experimental Finance forthcoming*.

Umar, Z., M. Gubareva, I. Yousaf, and S. Ali (2021b). A tale of company fundamentals vs sentiment driven pricing: The case of gamestop. *Journal of Behavioral and Experimental Finance*, 100501.

You, W., Y. Guo, and C. Peng (2017). Twitter's daily happiness sentiment and the predictability of stock returns. *Finance Research Letters 23*, 58–64.

Figure 1: GME and Russell 2000 index, one minute frequency

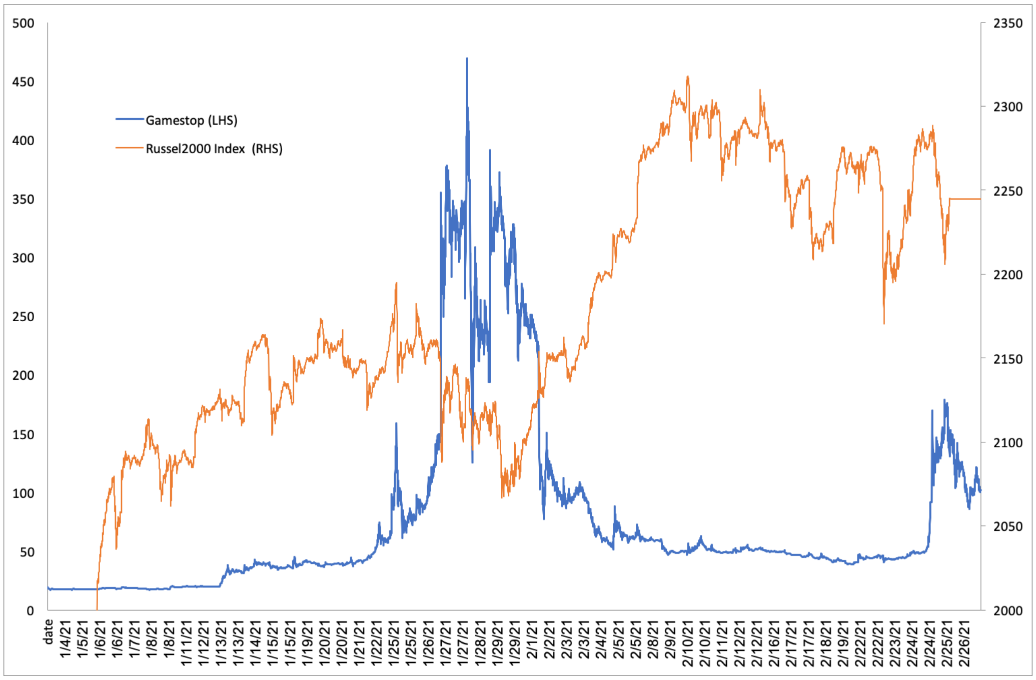


Table 1: Number of r/WallStreetBets threads and comments

Total Threads	**846,628**
Total Comments	**10,024,617**
Total	**10,871,245**

Table 2: New Words Dictionary and its Valence Scores

Words	Scores	Words	Scores	Words	Scores	Words	Scores
available	0.8	diamond_hand	3	cash	0.6	advice	1.3
awesome	3.7	dip	-0.4	concern	-1.3	alternative	0.9
baby	1.2	dumb	-1.9	crash	-3.2	amazing	3.2
bad	-2.7	earning	1.8	crazy	0.7	ass	-1.9
ball	0.4	easy	1.6	crypto	0.5	attack	-1.9
bull	2.8	end	-0.8	damn	-1.7	capital	1
bullshit	-2.4	enough	0.1	diamond	2.9	fact	0.3
buy	1.9	hype	1.2	hard	-1.1	fake	-2.3
call	0.9	idiot	-2.6	hedge	0.5	fight	-1.2
future	1.1	illegal	-3.2	hell	-2.5	fine	1.3
gain	2.2	interest	1.1	high	2.4	flair	1.4
gamma	0	issue	-1.1	hodl	2.8	fuck	-2.8
gang	-0.3	joke	-0.5	hold	1.5	fucking	-2.7
gold	2	jump	1.4	holding	1.6	fun	1.9
good	2.5	least	-0.4	hope	1.5	funny	1.9
great	3.1	legal	1.9	limit	-0.4	problem	-2.3
green	2	manipulation	-2.3	lmao	2.6	profit	2.5
hand	0.1	margin	-0.1	lol	1	proud	2.1
party	0.8	moment	0.7	long	1.8	pump	-0.5
penny	-0.2	moon	2.1	loss	-2.5	purchase	1.3
poor	-1.9	movement	0.9	love	2.3	push	0.5
possible	0.8	naked	-1.1	low	-1.7	quick	0.8
potential	1.4	nice	2	luck	2.1	retard	-2.2
power	2.2	order	0.4	revolution	2	share	0.8
pretty	2.3	panic	-3	rich	2.5	shit	-2.6
probably	0.4	straight	1	ride	1	short	-1.8
top	2.4	strong	2.1	rocket	2.8	silver	-0.2
trade	0.6	stupid	-2.1	sale	-0.7	small	-0.3
value	1.3	support	2.2	scare	-2.3	squeeze	-1.6
win	2.7	target	1.3	scared	-2.6	star	2.4
worth	1.9	tendie	1.7	sell	-1.8	stonk	1.5
wrong	-1.8	to_the_moon	3.5	seller	-1.3	stop	-0.8
yolo	2.4			selling	-1.9		height

This table shows the list of words used in the new lexicon with their associated valence scores .

Table 3: Updated VADER keywords valence scores

New words	**New Scores**	**Existing Words**	**Original Scores**	**New Scores**
bull	2.8	crazy	-1.4	0.7
buy	1.9	crash	-1.7	-3.2
diamond_hand	3	interest	2	1.1
tendie	1.7	loss	-1.3	-2.5
to_the_moon	3.5	profit	1.9	2.5

This table shows the valence scores assigned to the *new* words and updated scores assigned to the *existing* words in the lexicon.

Table 4: Granger Causality Results at 1min, 5min, 10min and 30 min frequency

1 minute			
GMEO_R does not Granger Cause NET_SENTIMENT	15227	2.39737*	0.09
NET_SENTIMENT does not Granger Cause GMEO_R	15227	1.12244	0.33
AVERAGE_SENTIMENT does not Granger Cause GMEO_R	15227	1.15125	0.32
GMEO_R does not Granger Cause AVERAGE_SENTIMENT	15227	4.02723**	0.02

5 minutes			
NET_SENTIMENT does not Granger Cause GMEO_R	3074	5.95583*	0.00
GMEO_R does not Granger Cause NET_SENTIMENT	3074	0.43033	0.65
AVERAGE_SENTIMENT does not Granger Cause GMEO_R	3074	0.40559	0.67
GMEO_R does not Granger Cause AVERAGE_SENTIMENT	3074	0.95974	0.38

10 minutes			
NET_SENTIMENT does not Granger Cause GMEO_R	1554	6.6658***	0.0013
GMEO_R does not Granger Cause NET_SENTIMENT	1554	0.00043	1.00
GMEO_R does not Granger Cause AVERAGE_SENTIMENT	1554	1.43995	0.24
AVERAGE_SENTIMENT does not Granger Cause GMEO_R	1554	0.45851	0.63

30 minutes			
NET_SENTIMENT does not Granger Cause GMEO_R	528	3.2623**	0.04
GMEO_R does not Granger Cause NET_SENTIMENT	528	0.22412	0.8
AVERAGE_SENTIMENT does not Granger Cause GMEO_R	528	1.8199	0.16
GMEO_R does not Granger Cause AVERAGE_SENTIMENT	528	1.79278	0.17

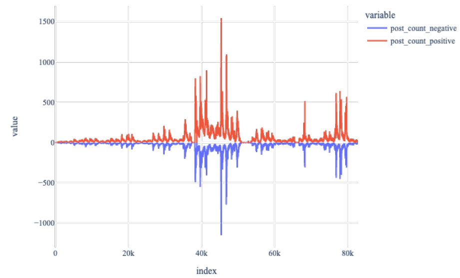
Note: *, **, and *** denote significance at 10%, 5% and 1% levels.

Figure 2: Wordcloud of All threads and comments

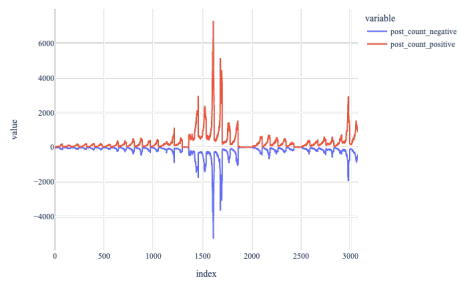
the most frequent words in All comments in all threads.

Figure 3: Positive and Negative Sentiments at different frequencies

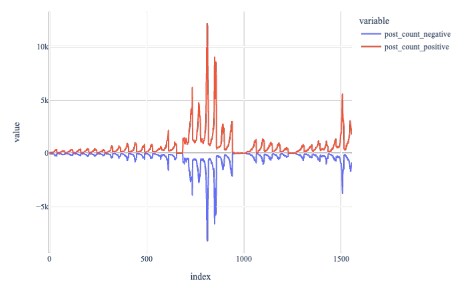
(a) 1 Minute



(b) 5 Minutes



(c) 10 Minutes



(d) 30 Minutes

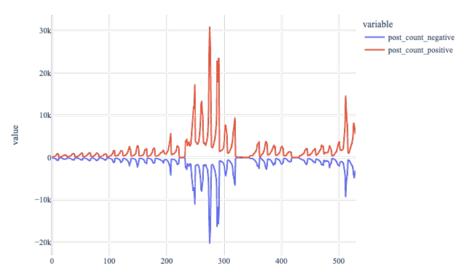
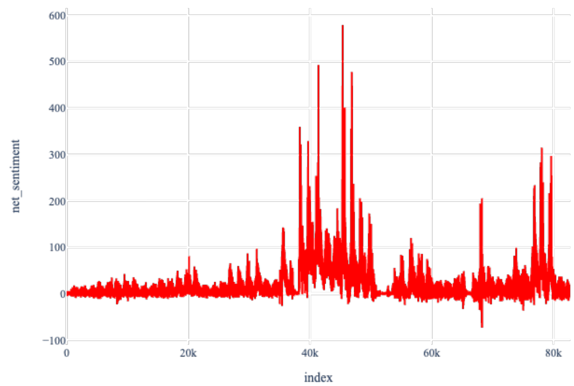
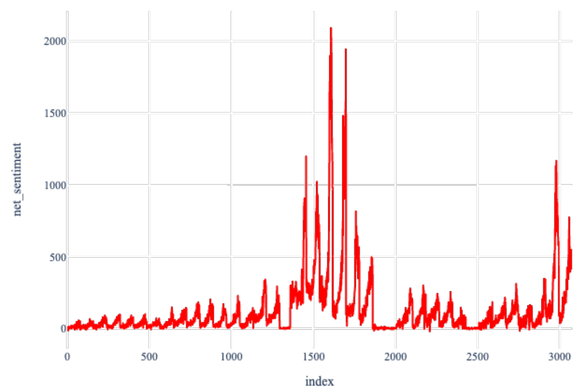


Figure 4: Net Sentiments at different frequencies

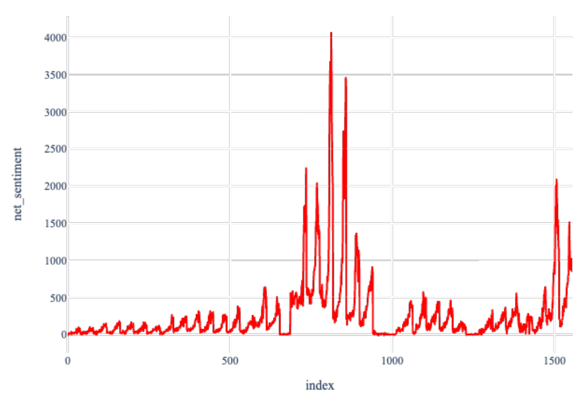
(a) 1 Minute



(b) 5 Minutes



(c) 10 Minutes



(d) 30 Minutes

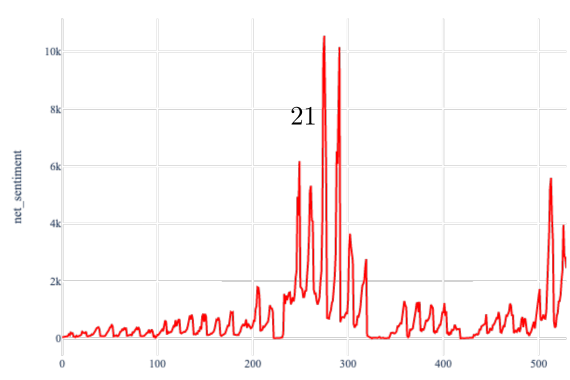
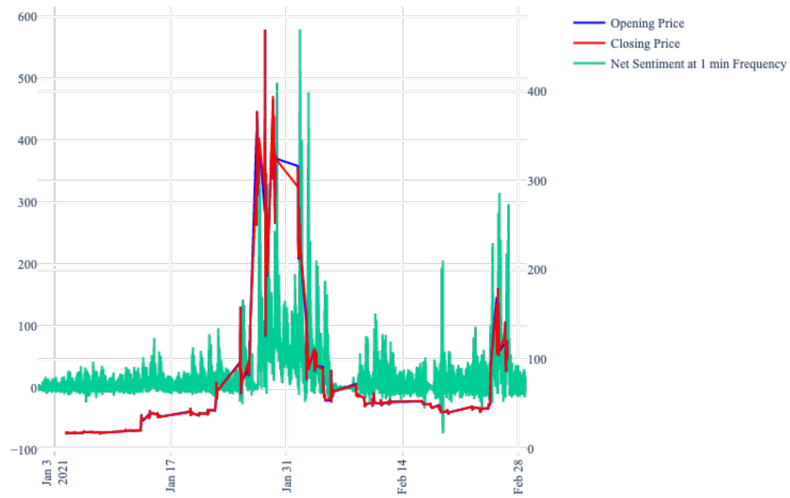
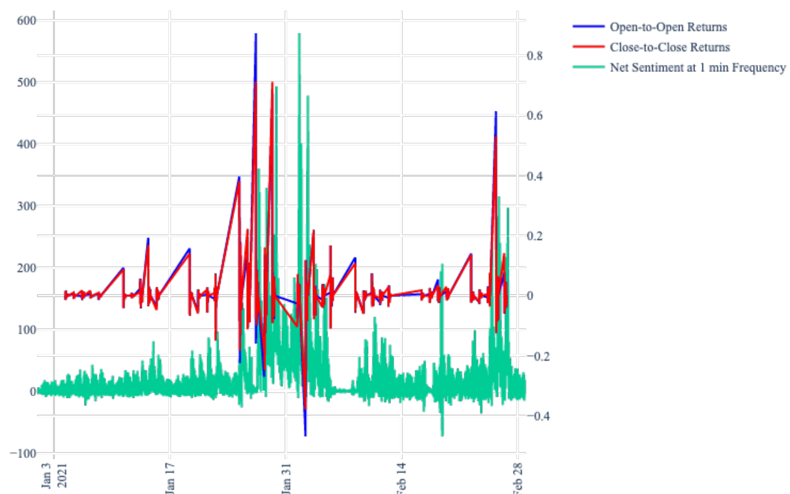


Figure 5: 1 Minute Frequency

(a) 1 Minute net sentiment and GME prices



(b) 1 Minute net sentiment and GME returns



(c) 1 Minute net sentiment and GME trading volume

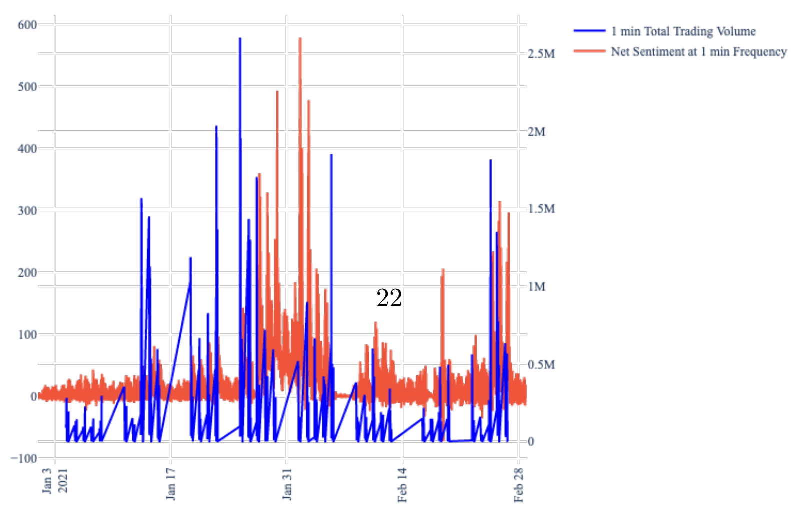
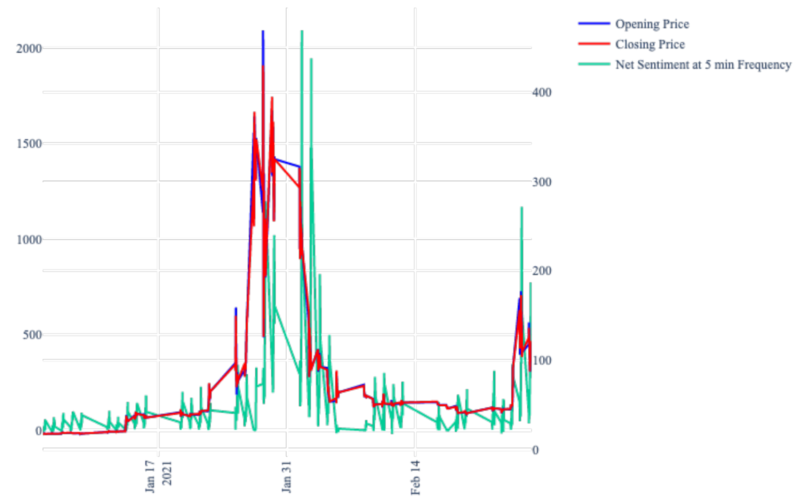
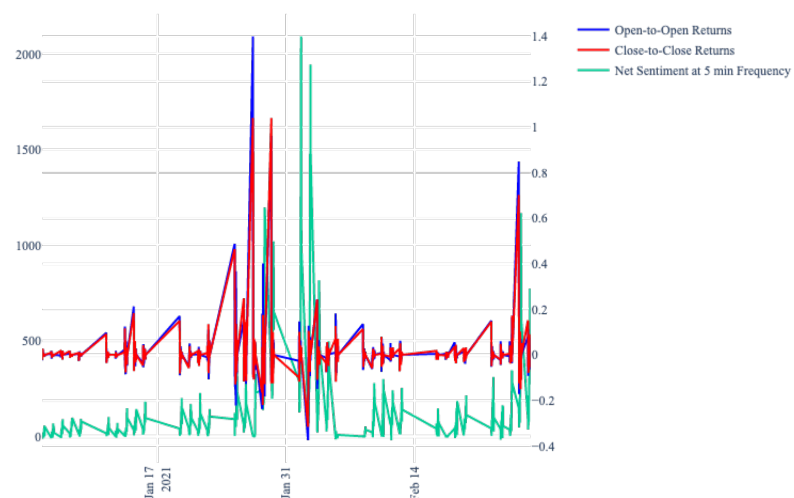


Figure 6: 5 Minutes Frequency

(a) 5 Minutes net sentiment and GME prices



(b) 5 Minutes net sentiment and GME returns



(c) 5 Minutes net sentiment and GME trading volume

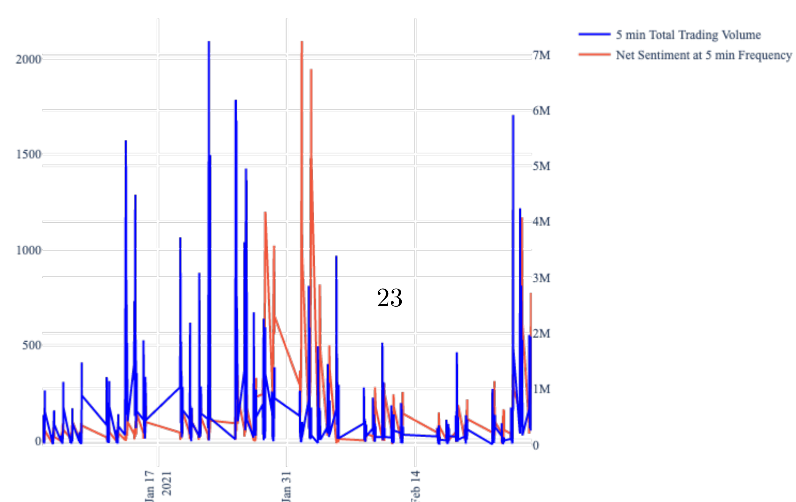
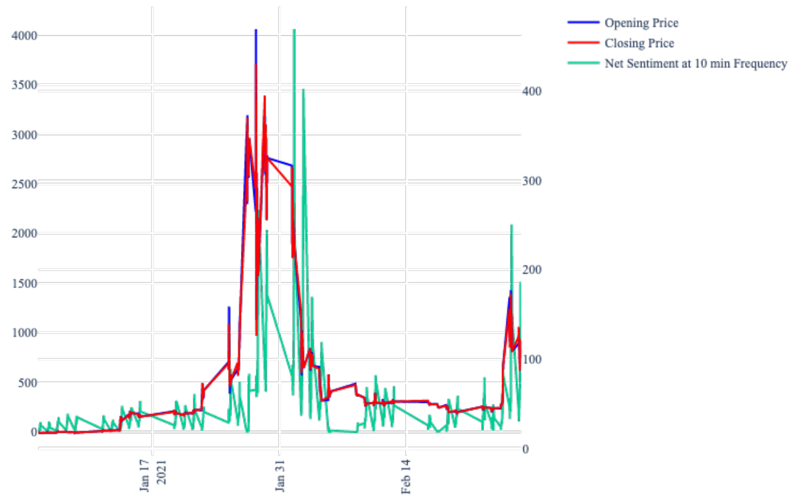
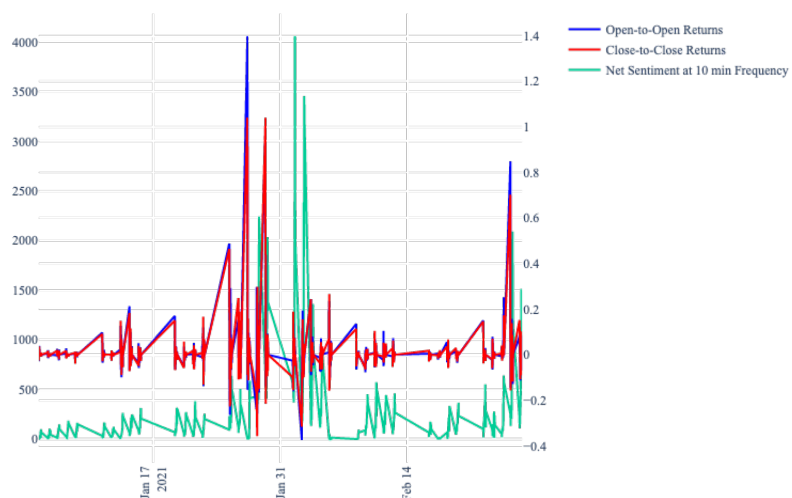


Figure 7: 10 Minutes Frequency

(a) 10 Minutes net sentiment and GME prices



(b) 10 Minutes net sentiment and GME returns



(c) 10 Minutes net sentiment and GME trading volume

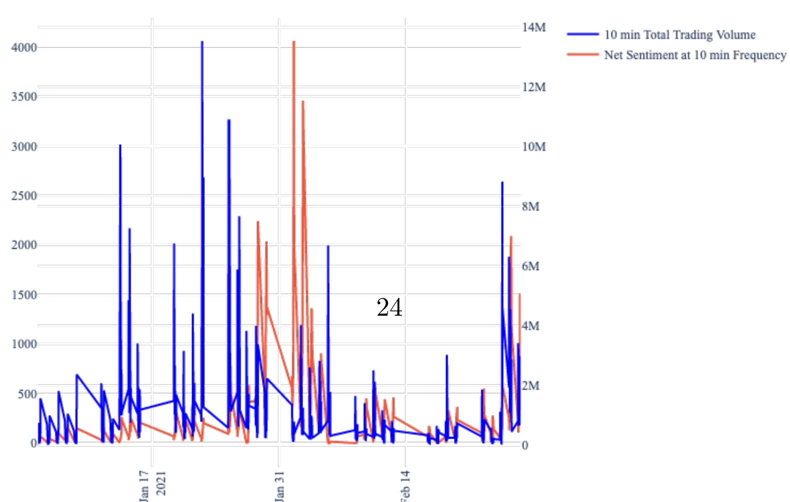
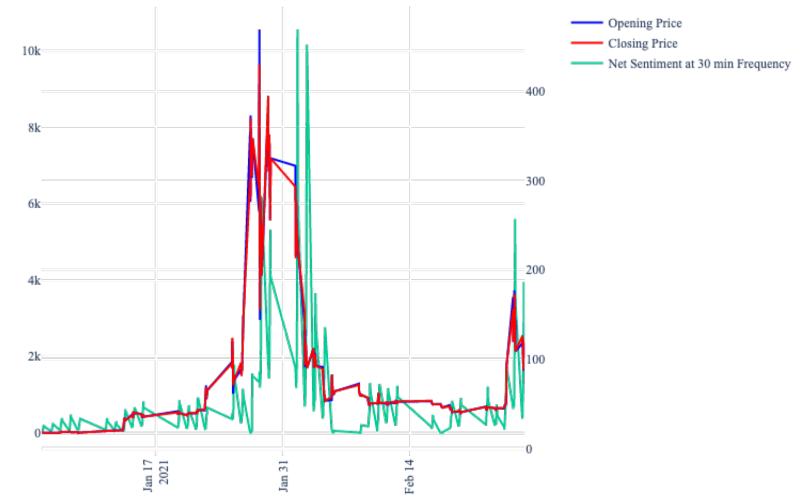
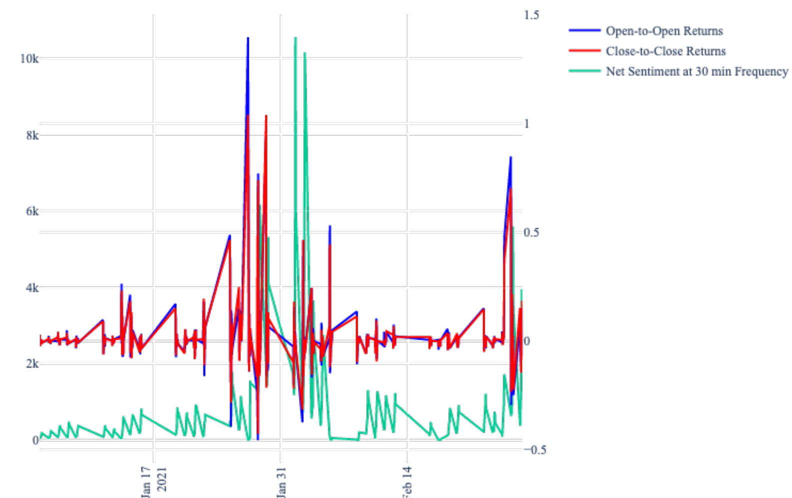


Figure 8: 30 Minutes Frequency

(a) 30 Minutes net sentiment and GME prices



(b) 30 Minutes net sentiment and GME returns



(c) 30 Minutes net sentiment and GME trading volume