

"I Just Like the Stock" versus "Fear and Loathing on Main Street" : The Role of Reddit Sentiment in the GameStop Short Squeeze

Cheng Long^a, Brian Lucey ^{a,b,c,d}, Larisa Yarovaya ^e

^a Trinity Business School, Trinity College Dublin, Dublin 2, Ireland

^b University of Sydney Business School, University of Sydney, Sydney, New South Wales, Australia

^c 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

^d Institute for Industrial Economics, Jiangxi University of Economics and Finance, 169, East Shuanggang Road,
Xialuo, Changbei District 330013 Nanchang, Jiangxi, China

^e Southampton Business School, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK.

* Corresponding Author: blucey@tcd.ie

Abstract

This paper investigates the role, if any, played by the social media platform Reddit, in the events around the GameStop short squeeze 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 perform textual analysis on 10.8m comments and surface the relationships between the comment sentiments and 1-min GameStop returns. Results indicate that both *tone* and *number of comments* influence GME intraday returns. Sentiments extracted from longer threads have a greater influence. "*Fear*" is the dominant sentiment in all comments, while comments that express a "*Sad*" sentiment show the most significant impact. While investors may just like the stock, it appears that fear and loathing also are important.

Keywords: Gamestop; Reddit; Robinhood; media sentiments; short squeeze; Herding; textual 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*. Wall Street experts, finance scholars, and financial regulators warned amateur investors that they cannot in the long run outplay experienced investors, including hedge funds, since retail investors do not have sufficient skills or investments experience to actually understand the complex trading strategies performed by the large institutional investors. However, these attempts at dampening down the excitement in the online investors' community may have only amplified the investor mania and strengthen the so-called "*to the moon*" movement. Apart from GameStop, Reddit's investors were coordinately buying other assets, such as AMC Entertainment and the DogeCoin cryptocurrency, adding further volatility and uncertainty to the financial markets. Media and online discussion suggests that the social media platform Reddit, and particularly discussion on the r/WallStreetBets subreddit, a part of the platform where members can discuss trading strategies, was identified as especially influential. Therein lies the motivation for this paper.

The GameStop short squeeze attracted substantial media attention. 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 taking on larger in a "revenge of the little guy" scenario¹. 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]), amongst others. In this paper, however, for the first time, we aim to 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 1-min GameStop stock prices for the period from January 1st to 28th February 2021.

The results of previous studies show the 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 and short squeezes (e.g., 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 and tone affects the performance of highly speculative

¹Good summaries of this can be found in the collated papers on the phenomenon at <https://theconversation.com/global/topics/gamestop-99332>

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]).

GameStop's experience has been portrayed in the media as a war between amateur investors and the Wall Street giants. While there are many possible cultural, financial, and environmental causes that can explain the GME phenomenon, we can distinguish three commonly discussed in news: (1) Financial Technologies (Fintech) enabled easy access to financial markets by amateur investors via online trading platforms, such as Robinhood; (2) the meme culture and herding behaviour of (younger) investors coordinated via social media platforms and online forums, like r/WallStreetBets subreddit; (3) increased feelings of social injustice and anger against the older generation and established centralised financial institutions, amplified by the ongoing COVID-19 pandemic².

While Eaton et al. [2021] consider the causal effects of the Robinhood trading on financial markets, we focus here on the impact of meme-driven culture and extract investors sentiments from the r/WallStreetBets subreddit posts. In contrast to previous studies that mainly focused on positive and negative sentiments (e.g., Apergis et al. [2016]; Nisar and Yeung [2018]) we attempt to go further and surface the specific emotions that were commonly expressed in this subreddit's posts, such as "Angry", "Fear", "Happy", "Sad", and "Surprise". Thus, our paper helps to uncover whether the feelings such as Anger and Fear amplified by the social distancing and isolation during the COVID-19 encouraged amateur investors to rise against Melvin Capital hedge funds and others short sellers. Furthermore, we assess the impact of the most influential and viral Reddit's threads, those that had more than 2k comments in each thread (long threads), and compare it with the impact of (short) threads that had less than 2k comments. We also hypothesise that the impact of the shorter and longer threads could be different. We perform thorough data cleansing procedure, and perform a number models using sentiments derived from GME-only and All GME and non-GME comments published on r/WallStreetBets during the first and second GameStop's stock rallies.

The novelty of this paper is twofold. First, this is, we believe, the first paper that analyses the

²See Yarovaya, L. (2021) "GameStop: WallStreetBets trader army is back for a second share rally – here's how to make sense of it" <https://theconversation.com/gamestop-wallstreetbets-trader-army-is-back-for-a-second-share-rally-heres-how-to-make-sense-of-it-156083>

effect of social media sentiments derived from Reddit on stock price movements. In contrast to other popular social media platforms and forums, the recent GameStop stock rally was driven by discussion on the influential social media site Reddit. Before GameStop, Reddit was not considered as a platform that can influence the stock markets, and has generally been ignored by finance scholars. This is despite some evidence which suggests that it is in fact the most influential social media platform (see <https://knowledge.wharton.upenn.edu/article/history-of-reddit/>) and 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, and would require powerful text and messy data analytical tools (e.g. Das and Chen [2007]). However, these are the messages that could explain the “*to the moon*” movement, to give it one of the common terms used in the subreddit, potentially better than any traditional approaches. Thus, in this paper we provide a first attempt to capture these sentiments, establishing a foundation for future research in this area. Second, this paper contributes to the previous literature on asset price bubbles and short squeezes utilising high-frequency intraday data. In summary, 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 role of social media in stock price formation is increasingly discussed . Social media platforms can act 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]). Furthermore, social media sentiment is found to be a powerful tool for forecasting stock market returns. For example, Gu and Kurov [2020] report 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. Motivated by the literature on the impact of social media sentiments on stock price movements and considering the evidence on the importance of both timing and tone of the sentiments (e.g.,[Ahmad et al. \[2016\]](#); [You et al. \[2017\]](#); [Liu and Han \[2020\]](#)), we hypothesise that the extent and intensity of sentiments extracted from r/WallStreetBets subreddit will affect intraday share prices. We formulate two related hypotheses:

H1a: *The tone of Reddit's sentiments affects GME intraday returns.*

H1b: *The number of Reddit's comments affects GME intraday returns.*

Considering the fluid and dynamic nature of Reddit's discussion fora, intraday changes of tone can be also expected. Furthermore, there are strong reasons to believe that the linkages between investors sentiments and GME stock could be bidirectional, and the tone of sentiments can change from positive to negative following intraday changes in GME prices. As was shown by [Tetlock \[2007\]](#) media pessimism can predict downwards trends on stock markets and high trading volume, while low market returns also leads to high media pessimism. Therefore, in addition to *H1*, we further specify two related hypotheses that can shed a light on intraday changes in mood of the Reddit's investors, and account for intensity of discussion on trading forum.

H2a: *Tone of the Reddit's comments changes within the day.*

H2b: *The tone effect is intensified when more comments are posted.*

Thus, *H2a* will examine how the tone of investor's sentiments vary within one trading day and outside the trading hours. [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. Additionally, *H2b* will account for the number of comments posted per hour on r/WallStreetBets during and outside the trading hours. Therefore, we assume that more active discussion on the investment forum will happen during the trading hours and will lead to stronger media effect on GME's shares. [Umar et al. \[2021\]](#) used Twitter publication count as a proxy of media sentiments in analysis of GameStop short squeeze 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 tone, timing, and number of 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 scores from forum's participants. Thus the number of comments in thread can be used a determinant of the popularity of this discussion among redditors, while comments's scores additionally can indicate the popularity of the specific comment. Thus, we compare the impacts of sentiments extracted from Long and *Short* threads on GME returns to account for the intensity of the discussion, attention, and influential power of the threads, and we hypothesise:

H3: *Sentiments derived from Long threads have a stronger impact on GME intraday returns than sentiments derived from Short threads.*

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 increased effect of social media "hype" 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 2020 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. In order to compare the tone of the sentiments contained in most popular and less popular threads we divided the data into two different groups. The first group *LONG* contains messages from threads containing more than 2k comments, while the second group *SHORT* contains data collected from the threads with less than 2k comments in each thread ³. This grouping will help to compare the influential power of the most popular and less popular Reddit threads and test H3. We further separate all comments into *GME-only* - containing comments that mention GameStop (e.g. Gamestop/gamestop/GME/gme), and *Non-GME* - containing all other comments in the threads that do not explicitly mention GameStop. As it was pointed earlier, the Reddit comments often contains slang, offensive words, allusion and elliptical commentary, therefore we assume that the

³The 2000 comment limit was determined heuristically

comments that do not contain GameStop could still affect GME stock prices. Table 1 shows number of comments collected for each group out of total 10.8 million comments.

[Table 1 here]

A point to note is that despite what a casual reader or consumer of media might have concluded, the subreddit was not consumed by a discussion of GME to all else. We note that in fact less than 10% of all comments mentioned GME explicitly. This is not to say that the site was un-influential - merely that it contains a wide range of discussions, around a variety of assets, and could therefore contain in aggregate useful information on overall sentiment towards the markets.

3.2. Methods

We begin our analysis by extracting sentiments from text messages scrapped from r/WallStreetBets subreddit. To extract emotions from the comments we utilise the Python 3.0 *Text2emotions* package⁴, that categorises emotions displayed in text into 5 groups "Angry", "Fear", "Happy", "Sad" and "Surprise". We also use *SentimentIntensityAnalyser* object from the VADER package⁵, the *polarity_scores* provide the overall sentiment metrics(compound score) for the comments. 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. Thus, *OVERALL* variable takes value of +1 for positive, -1 for negative and 0 for neutral tones respectively.

We investigate the impact of five specific tones and overall sentiments on the intraday GME prices and price changes in a number of ways. First, we produce a baseline result on the impact of Reddit sentiments on 1-min GME open-to-open returns. For this empirical investigation we do not formally "clean" the data with respect to market movements, however, we do account for overall market activity by adding Russell2000 changes.

We investigate the relationship between the top one minute changes in Gamestop returns the Reddit comments and posts, by adding OPENING variable to our regressions and accounting for overnight returns. To test *H2a* and *H2b* we further examine how the selected five tones and overall sentiment in messages vary during our observation period, and how the number of comments published affect the analysed impacts.

To further assess the relationships between Reddit investors sentiments and GME price changes we employ the wavelet coherence framework following [Torrence and Compo \[1998\]](#) specifications. This method is particularly powerful for analysis of shorter observation periods (see [Sharif et al. \[2020\]](#)), and can help to identify time-frequency co-movements between selected variables adding to

⁴See for details: <https://pypi.org/project/text2emotion/>

⁵See for details: <https://www.nltk.org/api/nltk.sentiment.html>

the regression analysis results. It has already been used in an analysis of GME price movements in Umar et al. [2021], albeit with a different perspective to here. In addition, to assess causal linkages between Reddit's sentiments and GME intraday price changes we employ a Granger causality test. Considering that all models used are well-known and used widely in the finance literature, for space consideration we do not include detailed specifications of the methods utilised in this paper, and the details are available upon request.

4. Results

4.1. The GameStop short squeeze 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 GameStop short squeeze. The outcome of this short squeeze was particularly bad for Melvin Capital, which required nearly \$3b in additional capital, with overall short squeeze losses nearing \$25b⁶. 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 rollercoaster ride in January-February 2021.

The GameStop short squeeze received enormous public attention and was widely discussed in the media. The gamification of trading and increased ease of access to financial markets by amateur investors via online trading platforms such as Robinhood, spiked extensive debates and was followed by a sequence of congressional hearings and lawsuits. Indeed, trading forums like r/WallStreetBets in Reddit were accused of market manipulation. This case showed the growth of the decentralised financial system and technology and its potential to destabilise financial markets, therefore the GameStop case became 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

⁶see <https://www.ai-cio.com/news/gamestops-robinhood-boasters-clobbering-hedge-funds/> for a comprehensive overview

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]) for example, the content of which, however, is still highly correlated with 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. Therefore a better understanding of investors' lexicon used by redditors can help to improve a quality of sentiment analysis of other social media platforms.

Thus, without going in-depth in semantic analysis, we show the key examples of messages and hashtags used on r/WallStreetBets subreddit to demonstrate a typical lexicon of words used by Reddit's amateur investors during the GameStop's share rally. Table 2 below combines the most popular and typical phrases included in 10.8 million comments collected, and their financial interpretations.

[Table 2 here]

As it is evident from Table 2, redditors clearly used hashtags and standard phrases to express trading signals which provides further support to our hypotheses, since these messages can correlate with bullish and bearish market trends. Figure 2 below shows the most popular words used in GME related comments and all comments.

[Figure 2 here]

4.2. The impact of Reddit sentiments on GME

We analyse the impact of the investment sentiments r/WallStreetBets subreddit on 1-min Open-to-Open GME returns. Table 3 presents the results of three regression models and shows the impact of specific tones "Angry", "Fear", "Happy", "Sad" and "Surprise" on GME 1-min Open-to-Open returns. We ran three regressions using sentiments extracted from GME-only comments in SHORT and LONG threads containing GME-only comments, and ALL threads containing also non-GME comments. We also used *Opening* variable, which is a dummy variable of opening price used to capture overnight effect, as well *Russell2000 Delta* to account for 1-min change in Russell 2000 index. Table 4 further shows the results of these regression models performed using lagged values of each variables.

[Tables 3 and 4 here]

The results display significant relationships between *Opening* price and Russell 2000 1-min changes and GME 1-min open-to-open return which is evident from all six regressions performed.

These results show the importance of overnight returns on GME price dynamics. **Table 4 further revealed the impact of tones on GME intraday returns, and among all emotions displayed by redditors in short threads only *Sad* plays significant role in GME price formation.** Similarly, analysis of sentiments extracted from GME-only comments in Short threads demonstrated the significant impact of *Sad* sentiment, as well as its lagged value, however, for Long threads we results also displayed the significant impact of lagged "*Angry*" sentiment on GME intraday returns. In contrast, analysis of the sentiments extracted from all comments, including non-GME comments, demonstrated that only lagged values of *Surprise* sentiment affect the GME 1-min open-to-open returns.

Acknowledging the fact that extracting the investors sentiments and emotions from Reddit's discussion is a difficult semantic problem, we further use the common alternative approach to extract the sentiments and construct the *OVERALL* variable, which is a dummy variable that takes value of -1 for negative, 0 for neutral, and +1 for positive tone in the comments. Figure 3 shows the value of overall sentiments in GME-only Short, Long, and All threads.

[Figure 3 here]

We can clearly see that in both Short and Long threads the value of positive sentiments is the highest. However, while in Short threads the value of negative sentiments is higher than neutral, this is *opposite* in Long threads. The results for all threads show the dominance of neutral sentiments followed by positive and negative sentiments respectively.

Thus we analyse the impact of overall Reddit sentiment retrieved from GME-only comments in Short and Long threads, and All r/WallStreetBets comments on GME 1-min returns, and the results are presented in Table 5. Our findings indicate a lack of significant relationship between Positive, Negative or Neutral sentiments extracted from Short GME-only comments and ALL comments, however, show a significant impact of *OVERALL* sentiments extracted from LONG threads. **This provides supporting evidence that longer threads are more influential than shorter threads, and sentiments extracted from comments in longer threads play a more significant role in GME intraday price formation.**

[Table 5 here]

We further examine all instances where there was a rise or fall greater than 10% on a one minute basis. For rises this gives us 24 instances, 12 of which happen at the open, while for falls we have 26, with only 2 at the open. Shown in Table 6 are the emotions (for all) associated with these rises and falls. The *Overall* tone is lower, more negative, for both rises and falls, than in general. There is no difference in *happiness* or *anger*, but we do see marked changes in *Surprise* (above average for rises) *Sadness* (below average for rises, much above average for falls) and *fear* (below average for both).

[Table 6 here]

Finally, we employ the Granger Causality test to evaluate the causal linkages between investors sentiments and GME 1-min returns. Results reported in Table 7 confirm our previous findings shows that *Sad* sentiments extracted from Short GME-only comment Granger cause 1-min GME returns. Furthermore, these linkages are bidirectional, and changes in 1-min GME returns also Granger cause *Sad* sentiment. Furthermore, among all emotions expressed by Redditors on r/WallStreetBets forum Granger causality test revealed that causal linkages between *Happy* sentiment and 1-min GME returns. However, Granger causality test failed to display causal linkages between tones and GME intraday returns for sentiments extracted from Long and All threads, and confirmed only causal linkages between Russell 2000 1-min changes and GME intraday returns.

[Table 7 here]

4.3. Tone and number of comments

We further analyse how the tone and number of comments vary within one trading day during our observation period. Figure 4 below displays how the investor sentiments change throughout a day in Short and Long GME-only comments, and in all comments.

[Figure 4 here]

The results display a variability of tones for 1-min data, with however, limited variability of tones for 1-hours data. We can observe that *Fear* is the most dominant emotions expressed by Redditors in Short and Long Threads, which is also evident for All comments, following by *Sad* and *Surprise*. Notably, both *Happy* and *Anger* emotions are less present in the comments in all groups analysed, therefore we can conclude that in contrary to popular belief that amateur investors were driven by anger against large hedge funds, injustice etc, we cannot observe it in our data.

Figure 5 demonstrates the relationships between number of comments posted per hours and Sentiments and GME hourly prices. The engagement of Redditors with the discussion forum varies within 24-hour, and number of comments posted increase after the opening of the US stock market, reaching its peak at 15:00 GMT and decline after 20:00 GMT. The lowest number of comments is around 10:00 GMT (03:00 Pacific Time). While we cannot observe significant relationships between number of posts and tones in all groups of comments analysed (Figure 5 i, iii and v), we can see a strong relationship between number of comments posted per hour and opening and closing GME prices (Figure 5 ii, iv and vi).

[Figure 5 here]

It is particularly interesting to observe that the number of comments peaked already after the GME price fell following the first short squeeze at the beginning of February. The decline in GME prices led to intensified discussion on r/WallStreetBets. Moreover, later in February 2021 prior

to the second GME stock rally, we clearly can see an increase in number of comments before the increase in GME prices. This is evident for comments in all groups analysed.

To further understand the role of tone and number of comments in GME short squeeze, we utilise daily opening and closing prices for GME to compute open-to-close and close-to-open returns, as $OTC = LN(P_{C_t}) - LN(P_{O_t})$ and $CTO = LN(P_{O_t}) - LN(P_{C_{t-1}})$ respectively. [Kim and Suh \[2021\]](#) used intraday and overnight returns to construct firm-level sentiment indices as alternative measure of investors sentiments. We also assume that discussion on Reddit outside the trading hours can also affect the returns of GME shares, therefore we plot the relationships between tone and number of comments and GME OTC and CTO daily returns, as it shown in Figure 6.

[Figure 6 here]

Figure 7 also presents the results for Short and Long threads combined capturing all GME-related comments posted on r/WallStreetBets, while Figure 8 demonstrates the impact of Scores per comments and emotions channel by scores to further highlight that the most popular comments on Reddit are posted during same hours when the highest number of comments have been posted, which demonstrates the engagement of r/WallStreetBets investors with the platform content.

[Figures 7 and 8 here]

At the final stage of our analysis, we verify our results by using the wavelet coherence framework. Figure 9 demonstrates results obtained for sentiments derived from GME-only comment in Short threads, while Figure 10 for Long threads, and Figure 11 from all GME and non-GME comments posted on r/WallStreetBets.

[Figures 9-11 here]

The results show very low levels of coherence between the specific tones extracted from all comments on 1-min GME returns. Areas of high coherency are only evident for low frequencies and most pronounced for "Surprise" sentiment in all comments. The lack of coherency, of significant relationships, stands in contrast to the results in [Umar et al. \[2021\]](#) for overall volume of posts in twitter/news articles and GME returns. This highlights the importance of not just examining how loud the commentary might be around an asset but also the intensity of the discussion.

5. Conclusions

This paper investigates the impact of sentiments extracted from Reddit on GameStop's intraday returns. We textually analyse 10.8 million comments from the r/WallStreetBets subreddit for the period from the 1st of January to 28th February 2021 and analyse the dynamics of tones and

number of GME-related comments posted in Short and Long threads. We compare the results with all GME and non-GME comments posted on r/WallStreetBets during the observation period.

We report several important pieces of evidence.

First, we confirm the impact of tone and number of comments posted in r/WallStreetBets on GME 1-min returns. The results show that Redditors are more actively engaged with the discussions on r/WallStreetBets forum during the opening hours of the US stock exchange, reducing outside the trading hours, which is also confirmed by analysis of scores received by each comment. This stands in contrast to the stereotypical view of social media activists being economically unengaged. While number of comments are not related to tones, the results display a strong relationship between the number of comments posted per hour and GME prices during the first and the second GameStop short squeezes.

Second, among all emotions expressed by investors on r/WallStreetBets, *Fear* and *Sad* seem to be most common tones, while in contrast to popular beliefs, *Anger* towards hedge funds and large institutional investors is the least popular emotion expressed by amateur investors on Reddit. This is evident for Short, Long and all threads analysed. However, only a few tones of Reddit's sentiments demonstrated significant impact on GME 1-min returns. These are *Sad*, *Anger* and *Surprise*, while impact of *Happy* and *Fear* tones are insignificant according to regression analysis performed. However, the results of Granger causality test revealed bidirectional causal linkages between "Happy" sentiment and GME intraday returns. Third, we report that opening price and Russell 2000 returns are strongly related to GME 1-minute returns, which is consistent across all models.

Finally, we report that longer threads are more influential than shorter threads, since the overall (positive/negative/neutral) sentiments affect GME intraday returns only if extracted from Long threads, while overall sentiments from short threads demonstrate no impact on 1-min GME returns.

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 and academics, help to understand the role of Reddit platform in the GameStop short squeeze. We confirm the impact of Reddit sentiments on stock market, therefore, it is possible that any other asset will become a subject of excess volatility without any fundamental reasons for it following discussions of r/WallStreetBets forum, which shows the growing power of retail investors.

For future research, we can recommend considering alternative algorithms to extract tones and sentiments from Reddit's comments, as well as data at alternative frequencies. This study can be also extended by showing the impact of Reddit's sentiments on other stocks and cryptocurrencies that are popular among retail investors. Finally, future papers can differentiate between the impact of title and comments on stock markets returns to further verify the results reported in this paper.

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.
- Apergis, N., C. Lau, and L. Yarovaya (2016). Media sentiment and cds spread spillovers: Evidence from the giips countries. *International Review of Financial Analysis* 47, 50–59.
- Bajo, E. and C. Raimondo (2017). Media sentiment and ipo underpricing. *Journal of Corporate Finance* 46, 139–153.
- 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.
- 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.

- 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>*.
- 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.
- 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.
- Kim, B. and S. Suh (2021). Overnight stock returns, intraday returns, and firm-specific investor sentiment. *North American Journal of Economics and Finance* 5.
- 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.
- Nisar, T. and M. Yeung (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. *The Journal of Finance and Data Science* 4(2), 101–119.
- 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.
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. stock market. *Journal of Banking and Finance* 84, 25–40.
- Sharif, A., C. Aloui, and L. Yarovaya (2020). Covid-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the us economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis* 70.
- 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.
- Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139–1168.
- Torrence, C. and G. Compo (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79(1), 61–78.
- Umar, Z., M. Gubareva, I. Yousaf, and S. Ali (2021). Sentiment during recessions. *Journal of Behavioural and Experimental Finance* forthcoming.
- 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.

Table 1: Number of r/WallStreetBets comments

| | LONG | SHORT | GME-only | Non-GME |
|--------------------------------|-------------------|-----------|----------|-----------|
| Number of Comments | 7,420,086 | 2,604,531 | 830,696 | 9,193,921 |
| % from Total Comments | (74.02%) | (25.98%) | (8.29%) | (91.71%) |
| Full Titles/threads | 653 | 845,975 | 140,200 | 706,428 |
| % from Total Titles/Threads | (0.08%) | (99.92%) | (16.56%) | (83.44%) |
| GME titles only in L/S Threads | 182 | 140,018 | | |
| % from Full Titles/Threads | (27.87%) | (16.55%) | | |
| Total Comments | 10,024,617 | | | |
| Total Threads | 846,628 | | | |
| Total Data units | 10,871,245 | | | |

Note: *LONG* and *SHORT* refer to comments retrieved from threads with >2k and <2k comments respectively; *GME-only* includes comments that mention GME, while *Non-GME* all other comments. All comments are collected from r/WallStreetBets.

Table 2: Reddit - r/WallStreetBets lexicon

| Word/Phrase/Hashtag | Signal | Interpretation |
|---------------------------|---------|--|
| I just like the stock | bullish | Used to justify purchasing the stock or holding it |
| We like the stock | bullish | Used to justify purchasing the stock or holding it |
| To the moon | bullish | Used to justify purchasing the stock or holding it |
| Mooning | bullish | The stock goes up |
| Diamond Hands | hold | Used to encourage community to not sell the stock |
| Paper Hands | hold | Used as an offence to discourage community from selling the stock |
| YOLO | buy | "You only live once" used to encourage to invest more in the stock |
| Buy High Sell low | buy | Used to encourage to invest more in the stock |
| Buy the Dip/BTFD | buy | Used to encourage to buy the stock has dropped in price |
| Guh | bearish | Used when share price fell caused losses |
| Drilling | bearish | The stock price goes down |
| Bear Gang/Gay Bears | bearish | People who gets happy when stocks go down |
| Bull Gang/Big Dongus Crew | bullish | People who gets happy when stocks go up |
| Kang Gang | hold | People who gets happy when stocks go way up and then way down |
| Theta Gang | hold | People who gets happy when stocks go sideways |

Note: For more detailed r/WallStreetBets glossary follow <https://www.reddit.com/r/wallstreetbets/wiki/glossary>

Table 3: Impact of SHORT, LONG and ALL comments on GME 1-min Open-to-Open returns.

| Variable | SHORT | | LONG | | ALL | |
|--------------------|-------------|--------|-------------|--------|-------------|--------|
| | Coefficient | Prob. | Coefficient | Prob. | Coefficient | Prob. |
| GME Tick | 0.0000*** | 0.0006 | 0.0000*** | 0.0006 | 0.0000*** | 0.0006 |
| Opening | 0.0762** | 0.0486 | 0.0762** | 0.0485 | 0.0762** | 0.0484 |
| Russell 2000 Delta | 5.0061*** | 0.0000 | 5.0014*** | 0.0000 | 5.0038*** | 0.0000 |
| Anger | -0.0003 | 0.7745 | -0.0003 | 0.7607 | -0.0005 | 0.7074 |
| Fear | -0.0002 | 0.7098 | 0.0001 | 0.8945 | -0.0002 | 0.6330 |
| Happy | 0.0002 | 0.7680 | 0.0000 | 0.9656 | -0.0006 | 0.3919 |
| Sad | -0.0010 | 0.1057 | 0.0004 | 0.4849 | 0.0000 | 0.9405 |
| Surprise | -0.0006 | 0.2627 | -0.0006 | 0.4718 | -0.0009 | 0.2056 |
| Constant | -0.0003 | 0.2834 | -0.0007* | 0.0907 | -0.0004 | 0.1378 |
| R-squared | 0.0719 | | 0.0717 | | 0.0717 | |
| Obs | 15389 | | | | | |

Note: Dependent variable: GME 1-min open-to-open returns; Method: Least Squares; Number of Observations: 15389. OPENING is a dummy variable of opening price used to capture overnight effect. Russell2000 is 1-min change in Russell 2000 index. *, **, *** Statistics is significant at 10%, 5% and 1% levels.

lscape

Table 4: Impact of SHORT, LONG and ALL comments on GME 1-min Open-to-Open returns with lags.

| Variable | SHORT Coefficient | Prob. | LONG Coefficient | Prob. | ALL Coefficient | Prob. |
|-------------|----------------------|---------|---------------------|---------|--------------------|---------|
| GMETICK | 0.00000** | 0.01400 | 0.00000** | 0.01310 | 0.00000** | 0.01300 |
| GMETICK(1) | 0.00000 | 0.24900 | 0.00000 | 0.24740 | 0.00000 | 0.26520 |
| OPENING | 0.08373** | 0.02970 | 0.08371** | 0.02970 | 0.08392** | 0.02940 |
| R2ODELTA | 5.91625*** | 0.00000 | 5.91042*** | 0.00000 | 5.92207*** | 0.00000 |
| R2ODELTA(1) | -5.52021 | 0.14700 | -5.52494 | 0.14670 | -5.52126 | 0.14690 |
| ANGRY | -0.00104 | 0.42520 | 0.00083 | 0.49050 | -0.00073 | 0.62030 |
| ANGRY(1) | 0.00030 | 0.77190 | -0.00335*** | 0.00460 | 0.00042 | 0.75680 |
| FEAR | -0.00089 | 0.17120 | -0.00079 | 0.21870 | -0.00067 | 0.31590 |
| FEAR(1) | 0.00032 | 0.65020 | 0.00031 | 0.65120 | 0.00042 | 0.51250 |
| HAPPY | -0.00082 | 0.34190 | -0.00074 | 0.39060 | -0.00031 | 0.76570 |
| HAPPY(1) | 0.00089 | 0.29550 | 0.00085 | 0.31200 | -0.00057 | 0.57390 |
| SAD | -0.00229*** | 0.00390 | -0.00208*** | 0.00930 | 0.00023 | 0.74660 |
| SAD(1) | 0.00186** | 0.01240 | 0.00193** | 0.01240 | -0.00102 | 0.14010 |
| SURPRISE | -0.00106 | 0.12570 | -0.00094 | 0.17110 | 0.00024 | 0.78290 |
| SURPRISE(1) | -0.00005 | 0.93440 | -0.00012 | 0.85840 | -0.00237** | 0.01090 |
| Constant | -0.00030*** | 0.00040 | -0.00030*** | 0.00020 | -0.00008*** | 0.00010 |
| R-squared | 0.10275 | | 0.10303 | | 0.10244 | |
| Obs | 15388 | | | | | |

Dependent variable: GME 1-min open-to-open returns; Method: Least Squares; Number of Observations: 15389. OPENING is a dummy variable of opening price used to capture overnight effect. Russell2000 is 1-min change in Russell 2000 index. *, **, *** Statistics is significant at 10%, 5% and 1% levels.

Table 5: Impact of Positive, Neutral and Negative on GME 1-min Open-to-Open returns

| Variable | SHORT Coefficient | Prob. | LONG Coefficient | Prob. | ALL Coefficient | Prob. |
|--------------|----------------------|--------|---------------------|--------|--------------------|--------|
| GMEODELTA(1) | -0.0394** | 0.0272 | -0.0394** | 0.0271 | -0.0394** | 0.0272 |
| GMETICK | 0.0000*** | 0.0004 | 0.0000*** | 0.0004 | 0.0000*** | 0.0004 |
| OPENING | 0.0835** | 0.0295 | 0.0835** | 0.0295 | 0.0835** | 0.0295 |
| R2ODELTA | 5.8891*** | 0.0000 | 5.8929*** | 0.0000 | 5.8882*** | 0.0000 |
| R2ODELTA(1) | -5.3242 | 0.1584 | -5.3277 | 0.1580 | -5.3257 | 0.1582 |
| OVERALL | -0.0001 | 0.4536 | -0.0005** | 0.0379 | -0.0002 | 0.5375 |
| OVERALL (1) | 0.0002 | 0.2832 | 0.0006** | 0.0124 | 0.0002 | 0.5706 |
| Constant | -0.0007*** | 0.0000 | -0.0007*** | 0.0000 | -0.0007*** | 0.0000 |
| Rsquared | 0.1033 | | 0.1034 | | 0.1032 | |
| Obs | 15388 | | | | | |

Note: Dependent variable: GME 1-min open-to-open returns; Method: Least Squares; Number of Observations: 15389. OVERALL is a dummy variable that takes value -1 for negative, 0 for neutral, and 1 for positive tone in the comments. OPENING is a dummy variable of opening price used to capture overnight effect. Russell2000 is 1-min change in Russell 2000 index. *, **, *** Statistics is significant at 10%, 5% and 1% levels.

Table 6: Emotions, Rising and Falling

| | Overall | Rises | Falls |
|----------|---------|-------|-------|
| Overall | 0.14 | 0.08 | 0.12 |
| Happy | 0.10 | 0.10 | 0.10 |
| Angry | 0.07 | 0.06 | 0.05 |
| Surprise | 0.16 | 0.24 | 0.17 |
| Sad | 0.18 | 0.11 | 0.30 |
| Fear | 0.31 | 0.24 | 0.20 |
| Count | 21.01 | 29.54 | 9.00 |

Table shows the average emotion and intensity of posting for the subreddit overall and for those one-minute intervals where the percentage change was greater/lesser than 10%

Table 7: Granger Causality Test results - p-values

| Null Hypothesis: | SHORT | LONG | ALL |
|---|---------------|---------------|---------------|
| GMETICK does not Granger cause GMEODELTA*** | 0.0000 | 0.0000 | 0.0000 |
| GMEODELTA does not Granger cause GMETICK | 0.1338 | 0.1338 | 0.1338 |
| ANGRY does not Granger cause GMEODELTA | 0.6665 | 0.7825 | 0.4695 |
| GMEODelta does not Granger cause ANGRY | 0.9753 | 0.1770 | 0.6702 |
| FEAR does not Granger cause GMEODELTA | 0.2062 | 0.9073 | 0.6770 |
| GMEODELTA does not Granger cause FEAR | 0.6310 | 0.4385 | 0.4611 |
| HAPPY does not Granger cause GMEODELTA*** | 0.0079 | 0.6591 | 0.5749 |
| GMEODELTA not Granger cause HAPPY** | 0.0262 | 0.9851 | 0.9423 |
| SURPRISE does not Granger cause GMEODELTA | 0.6899 | 0.9843 | 0.4206 |
| GMEODELTA does not Granger cause SURPRISE | 0.7867 | 0.5930 | 0.4161 |
| SAD does not Granger cause GMEODELTA** | 0.0197 | 0.6299 | 0.7916 |
| GMEODELTA does not Granger cause SAD** | 0.0349 | 0.8674 | 0.8707 |
| R2ODELTA does not Granger cause GMEODELTA* | 0.0943 | 0.0943 | 0.0943 |
| GMEODELTA does not Granger cause R2ODELTA* | 0.0000 | 0.0000 | 0.0000 |

Note: *, **, *** Rejection of Null, statistics is significant at 10%, 5% and 1% levels. SHORT LONG ALL refers to sentiments extracted from GME-only comments in short threads, GME-only comments in long Threads, and GME- and Non-GME comments in both long and short threads.

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

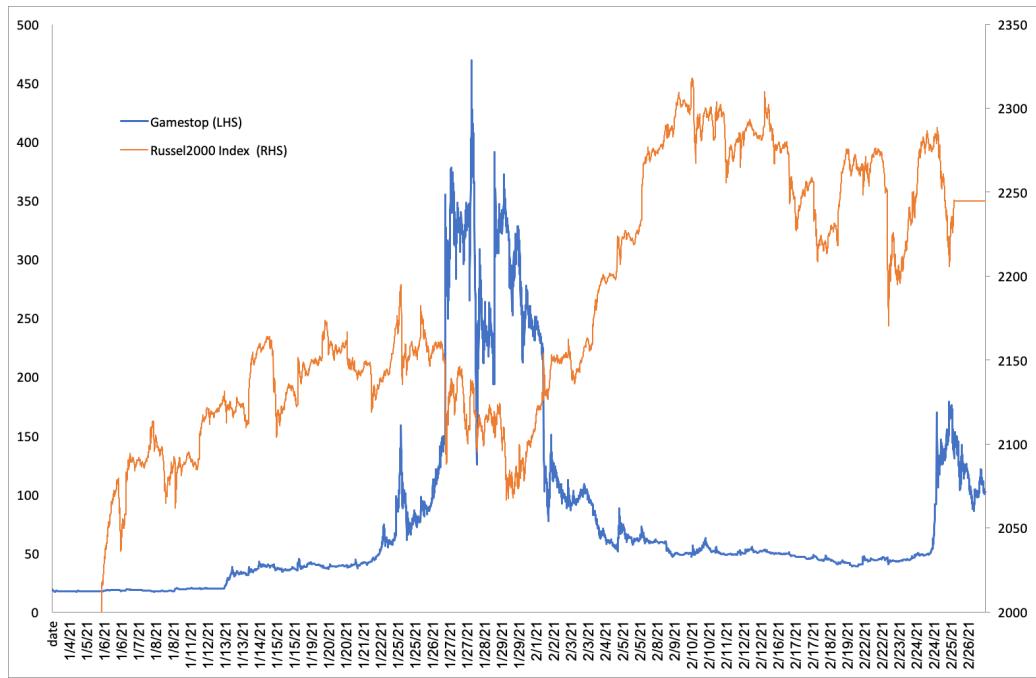


Figure 2: The most frequent words in the title of threads



Figure 3: Overall sentiments in threads

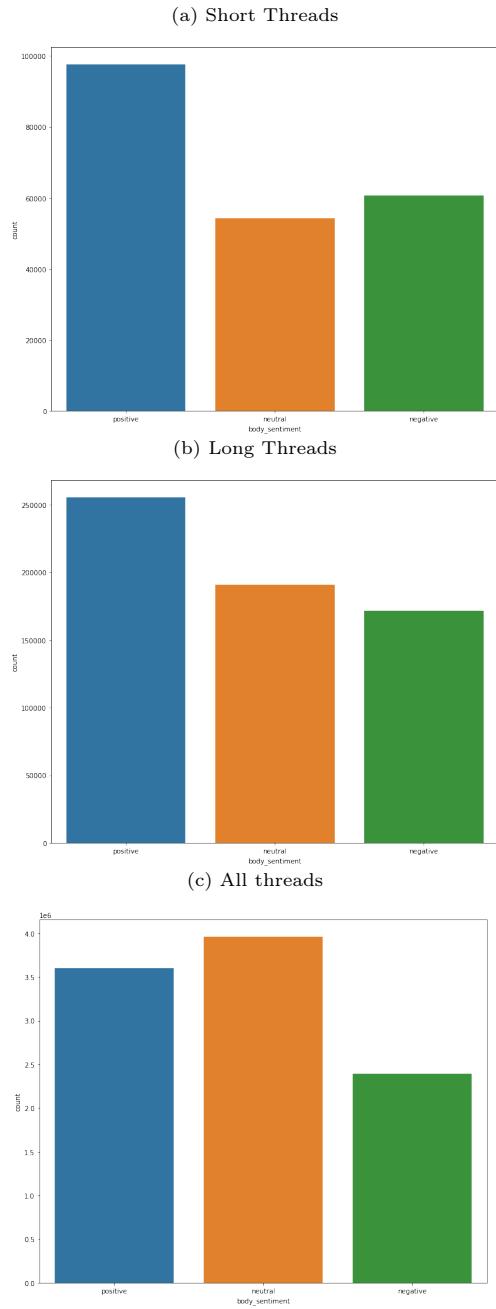
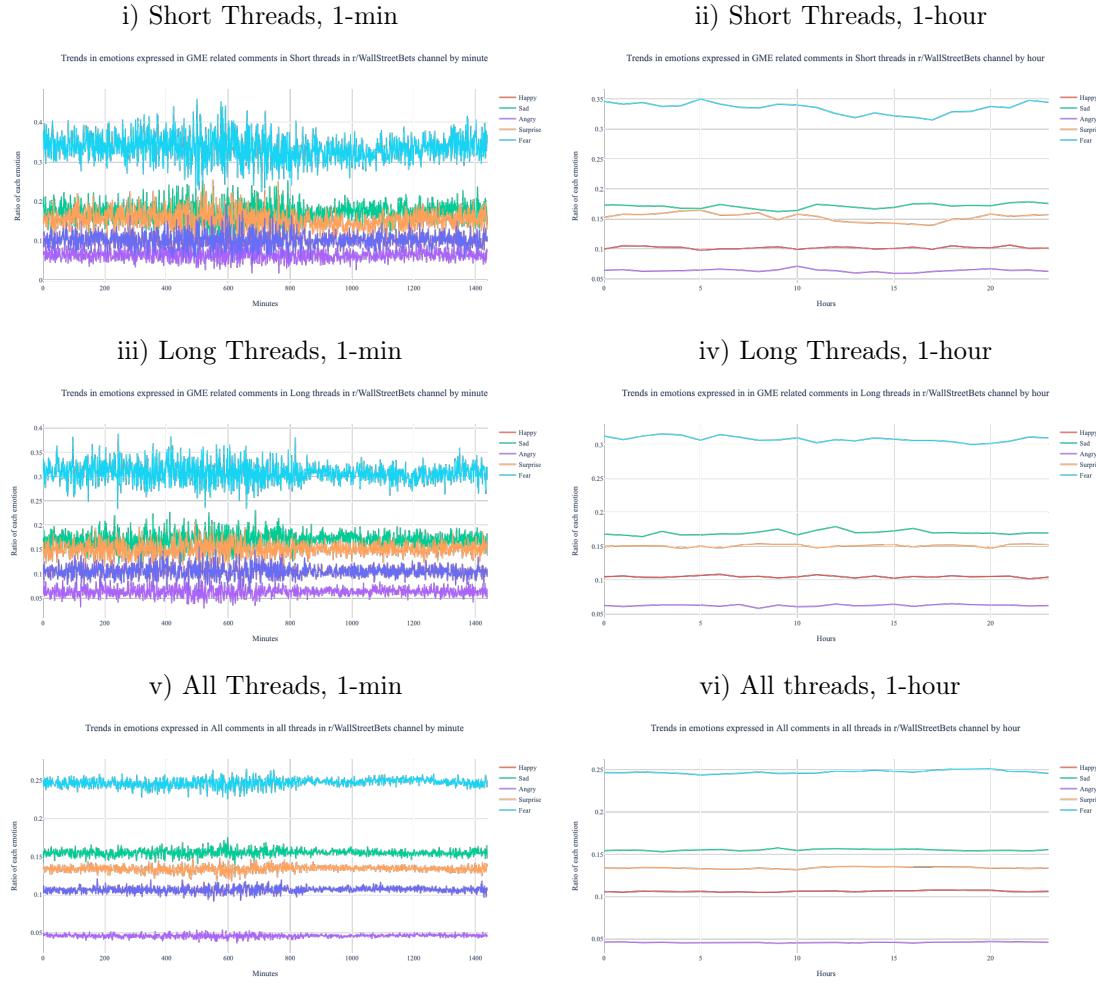


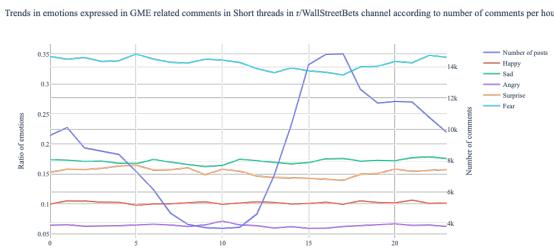
Figure 4: Tone Dynamics in Reddit threads



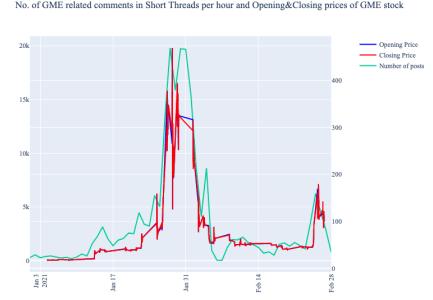
Note: These graphs display the dynamics of sentiments in GME-only comments in Short and Long threads, and in All comments including Non-GME in all threads.

Figure 5: The relationships between number of comments posted per hour and Sentiments and GME hourly prices

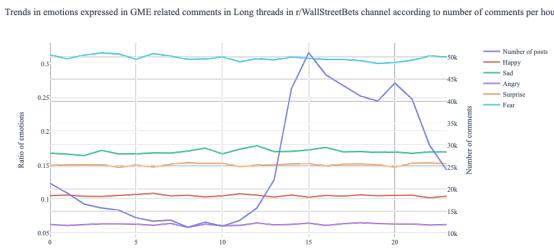
i) Short Threads, Tone VS Number



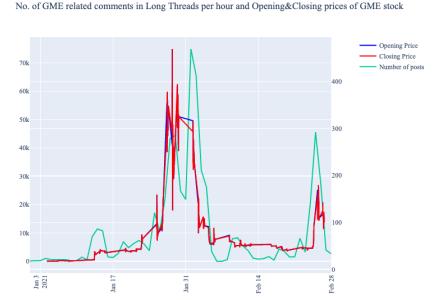
ii) Short Threads, Price VS Number



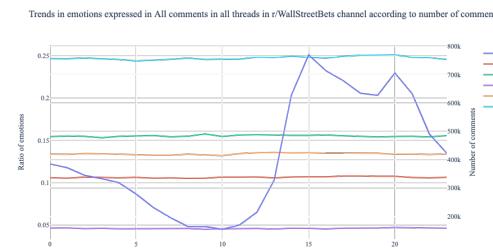
iii) Long Threads, Tone VS Number



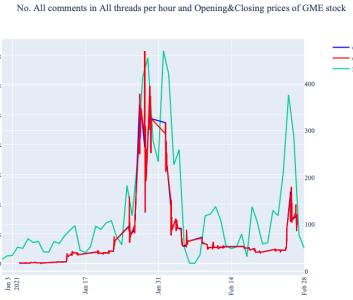
iv) Long Threads, Price VS Number



v) All Threads, Tone VS Number



vi) All threads, Price VS Number

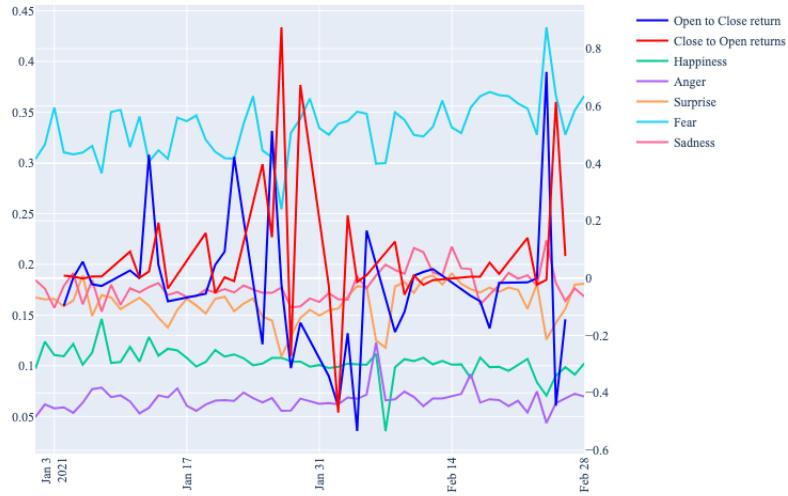


Note: These graphs display the relationship between sentiments and number of comments per hour, and between GME Opening and Closing prices and number of comments per hour, in GME-only comments in Short and Long threads, and in All comments including Non-GME in all threads.

Figure 6: Impact of the number of comments on OTC and CTO GME Returns

(a) Short Threads

No. of GME related comments in Long Threads per hour and Daily Returns of GME stock



(b) Long Threads

No. of GME related comments in Long Threads per hour and Daily Returns of GME stock

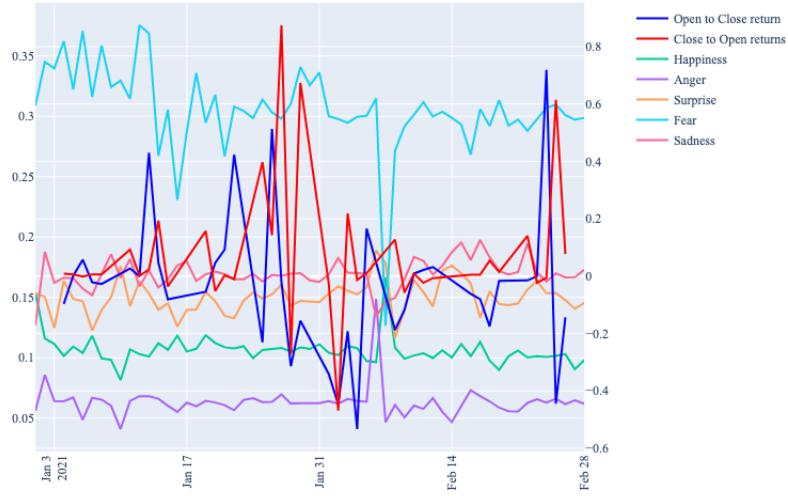
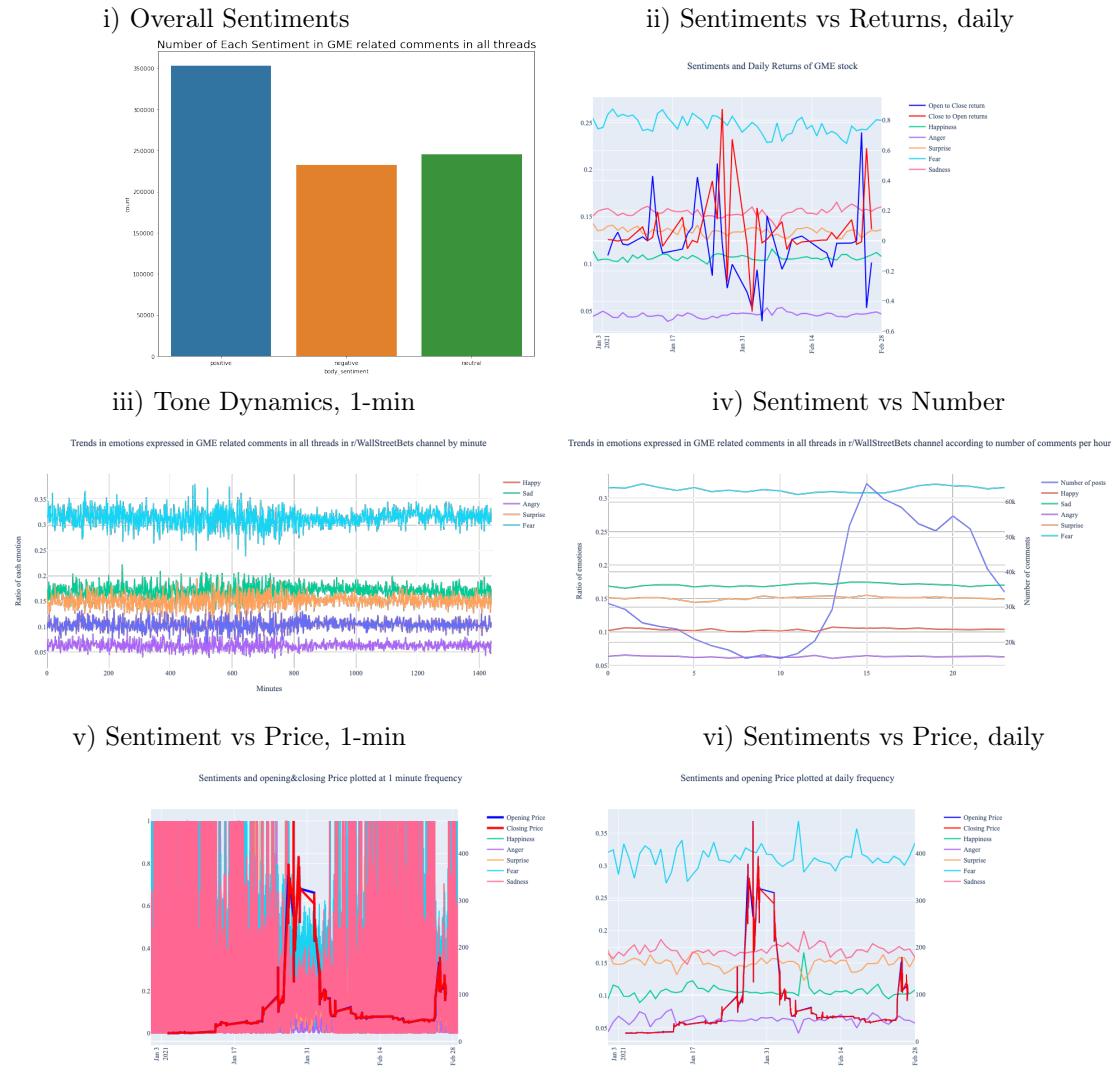


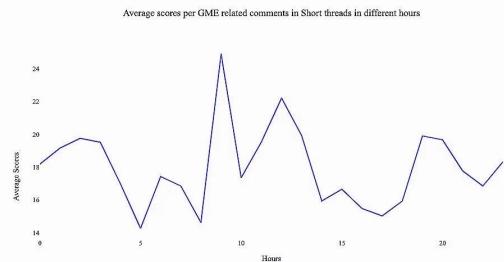
Figure 7: Impact of ALL GME-related comments



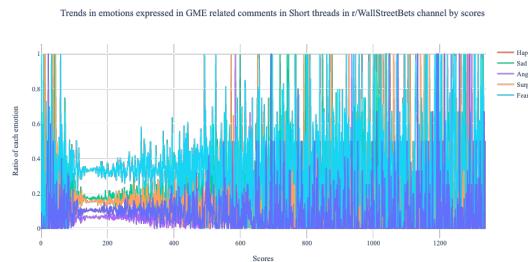
Note: These graphs display the relationships between tone and number of sentiments extracted from all GME-related comments posted on r/WallStreetBets and GME price changes.

Figure 8: Impact of scores

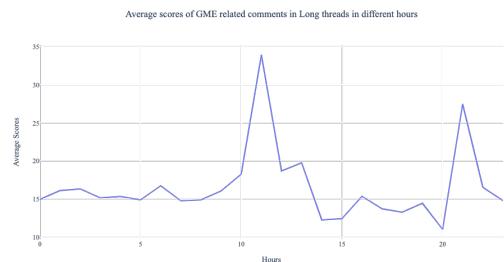
i) Short Threads, Scores per comment



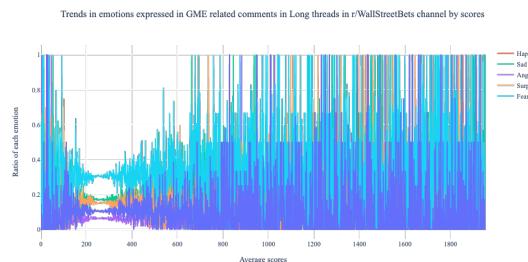
ii) Emotions channel by scores



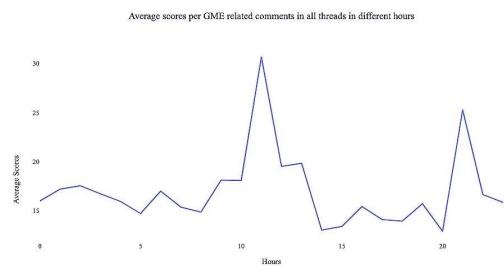
iii) Long Threads, Scores per comment



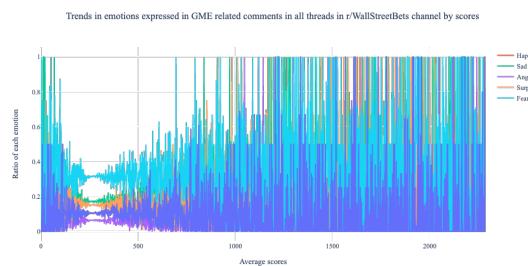
iv) Emotions channel by scores



v) All Threads, Scores per comment

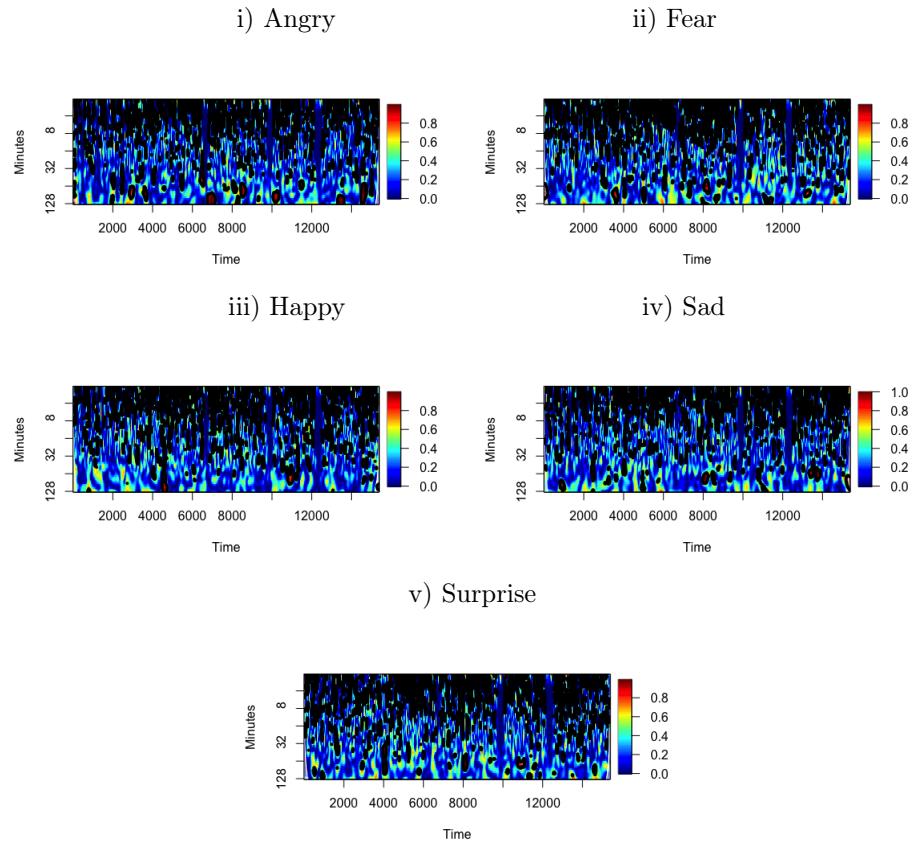


vi) Emotions channel by scores



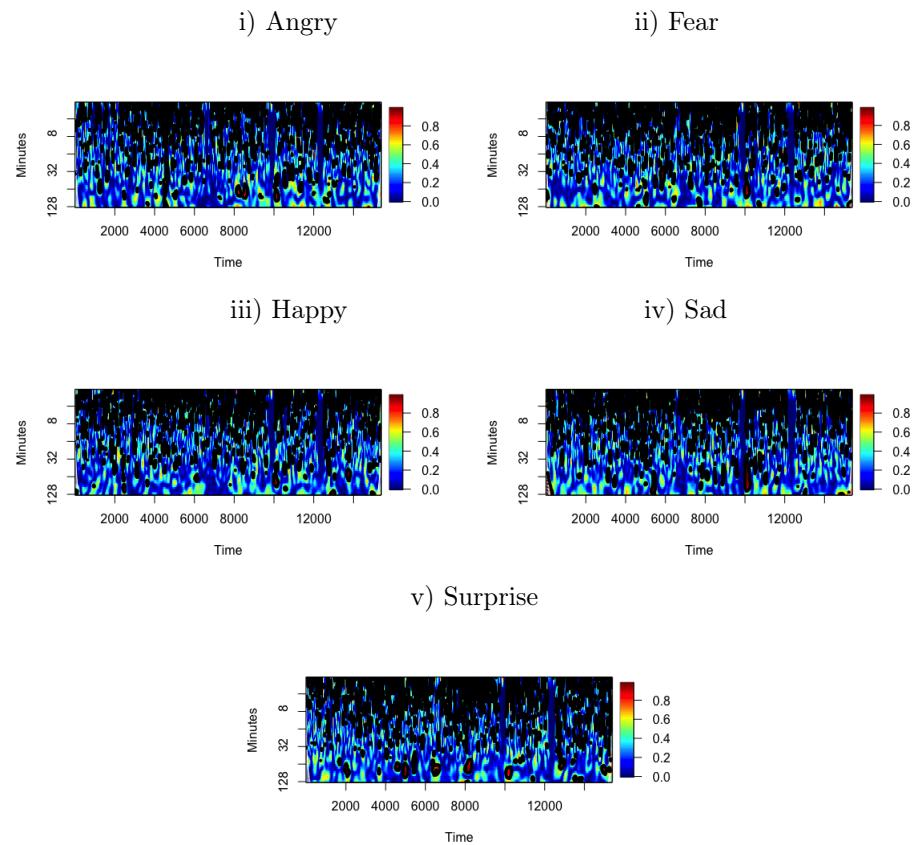
Note: These graphs display the average scores per GME related comments in GME-only comments in Short, Long, and All threads, as well as Trends in emotions expressed in GME related comments in all threads in r/WallStreetBets channelled by scores.

Figure 9: Wavelet Coherence Plots, Short Threads



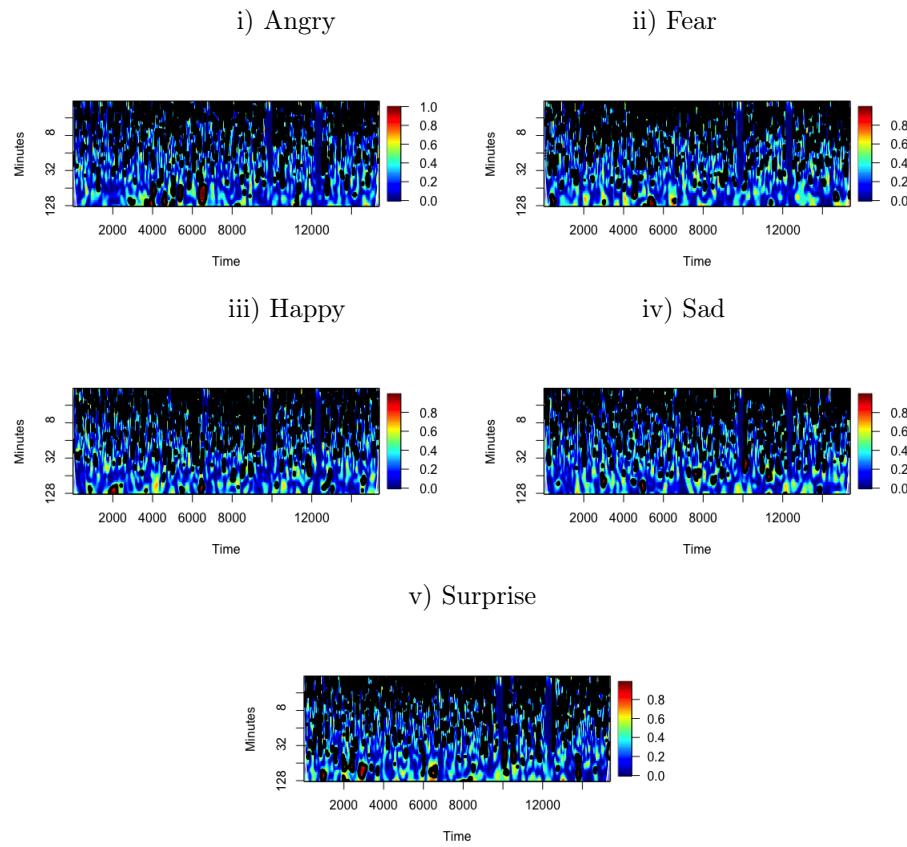
Note: These plots display results for sentiments extracted from GME-only comments in Short threads.

Figure 10: Wavelet Coherence Plots, Long Threads



Note: These plots display results for sentiments extracted from GME-only comments in Long threads.

Figure 11: Wavelet Coherence Plots, All Comments



Note: These plots display results for sentiments extracted from ALL GME-only and Non-GME comments at r/WallStreetBets.