NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS
International College of Economics and Finance

BACHELOR'S GRADUATE QUALIFICATION WORK
BACHELOR GRADUATION PAPER

ASSET PRICING WITH SOCIAL NETWORKS

Degree programme: Double degree programme in Economics of the National Research University Higher School of Economics and the University of London

by

Lena Yagafarova

Supervisor: Assistant Professor Vincent Fardeau

2020

ABSTRACT

In early 2021, stocks listed on New York Stock Exchange like GameStop Corp. (GME) and BlackBerry (BB) saw dramatic price rises, influenced by retail traders on Reddit's online community *Wallstreetbets*. This study finds a negative correlation between *Wallstreetbets'* prevalent negative opinions and the 5-minute log-returns of GME and BB. While the prevalence of negative sentiments predicts lower GME log-returns, a higher volume of marketable retail orders under such sentiments indicates larger positive log-returns shortly after. The study also shows that increased traffic boosts retail investor bullishness and deters short sellers, with a lag effect. Post-trading restrictions, the impact of *Wallstreetbets* opinions on GME's in-the-money call option interest dynamics significantly diminishes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In late January to early February 2021, the United States stock market experienced significant volatility, with several stocks, including GME, AMC, and BB, experiencing sharp price increases. This surge is widely attributed to the collective actions of retail traders in the Reddit *Wallstreetbets* online investing community, reminiscent of the "Occupy Wall Street" movements following the 2007-2008 Global Financial Crisis and the economic downturn caused by the COVID-19 pandemic. This period marked an accumulation of frustration among young retail investors towards traditional financial institutions.

The past decade has seen an increase in retail investor participation in U.S. financial markets, facilitated by the advent of commission-free trading apps like Robinhood. These factors converged, leading to a significant conflict between Reddit users and Wall Street.

A key moment occurred when Redditors noticed that hedge funds had heavily shorted GME, betting on a price decline. In response, they coordinated to drive up GME's stock price. On January 27, 2021, GME's stock price surged by 1,500% compared to its price on January 11, 2021. However, following the imposition of trading restrictions by retail brokerages, the stock price plummeted.

The influence of investment advice on social media has gained prominence in recent years. The SEC, in early 2012, acknowledged social media's "landscape-shifting" impact on financial markets (U.S. Securities and Exchange Commission 2012). Numerous studies have explored the relationship between social media opinions and financial markets. For instance, Chen et al. 2014 found that opinions shared on the online investment community 'strongly predict future stock returns and earnings surprises'. Similarly, the January 2021 surge in meme stock prices has been the subject of various studies. Allen et al. 2021 examined the short squeeze episodes, retail trading, and social media activities related to the 13 stocks affected by the trading restrictions imposed by retail brokerage firms. Hu et al. 2021 explored the connections between Reddit *Wallstreetbets* activity, short selling, and retail order flow.

This research enhances existing studies on the effects of coordinated trading through social media platforms, extending the literature in several ways. Allen et al. 2021 used daily social media activity, measured by the number of mentions of a company's ticker, as a proxy, but this approach may not fully capture the impact of investor coordination. In contrast, Hu et al. 2021 expanded this framework by showing that a more positive tone in Reddit posts correlates with increased retail order flow and higher stock returns. Both studies suggest a link between heightened Reddit activity and positive stock price movements.

The current study seeks to determine whether interactions among investors on social media provide useful information for predicting stock returns. Specifically, it investigates whether popular opinions on *Wallstreetbets* can offer valuable insights for short-horizon stock return forecasting. To the best of my knowledge, no prior studies have simultaneously considered the sentiment and popularity of opinions shared on social media, especially in the context of the late January 2021 meme stock volatility. The focus on popular opinions from Reddit *Wallstreetbets* represents a novel aspect of this research. The choice of *Wallstreetbets* is motivated by its significant role in the GME and other meme stock rallies in January 2021.

The study also examines the contemporaneous relationship between popular Reddit opinions and last trade price log-returns, allowing for the number of *Wallstreetbets* mentions to be endogenously determined based on stock price volatility, retail trading, short selling, and total trading activity. Additionally, it explores the connection between the options market and *Wallstreetbets* opinions. Analyzing stock returns at 5-minute intervals, the study addresses intraday seasonality and market peculiarities. High-frequency data enables the identification of a strong real-time link between GME mentions on *Wallstreetbets* and trading activity, as well as stock price variations (Gârleanu et al. 2021).

The study differentiates between opinions shared during NYSE trading hours (09:30 – 16:00 EST) and non-trading hours. It focuses on the period from January 11 to February 4, 2021, encompassing significant events from the announcement of GameStop's board refreshment to the lifting of trading restrictions by retail brokerages.

In summary, the main research questions are: What is the effect of popular *Wallstreetbets* opinions on short-horizon stock returns? What contemporaneous relationship exists between these opinions and stock log-returns? How does conditional stock return volatility, retail order flow, and short selling flow react to a shock in *Wallstreetbets* traffic, considering that this traffic can be endogenously determined? How long does the effect of a Reddit traffic shock on these variables persist? And what is the contemporaneous relationship between the options market and *Wallstreetbets* opinions?

The research utilizes the Pushshift Reddit dataset. The granularity of this data allows for a thorough examination of both interday and intraday evolution of Reddit users' opinions about specific stocks, particularly focusing on the popularity of these opinions. The unique features of the Reddit website enable quantifying each opinion's contribution to the conversation through a voting system. This effectively illustrates the coordination effect, showing how Reddit users interact to guide the stock-related agenda.

Submissions or comments with a higher net number of votes are prioritized by Reddit's algorithms and are more visible at the top of the subreddit or comment section, hence reaching a larger audience. I have algorithmically scraped submissions and comments from the *Wallstreetbets* subreddit from January 11, 2021, to February 4, 2021, focusing on mentions of GME, AMC, BB, and NOK. This was done using the `R` programming language and the Pushshift Reddit dataset. The selection of GME, AMC, BB, and NOK was based on the volume of mentions, which exceeded 100,000 during this period, providing a robust dataset for analyzing *Wallstreetbets* popular opinions at high frequency.

However, the Pushshift Reddit dataset had missing values, particularly for AMC and NOK, which were more significant than for GME and BB. Consequently, AMC and NOK were excluded from the analysis. For GME and BB, I developed a sophisticated univariate time series imputation approach using Python. This method algorithmically imputes missing *Wallstreetbets* variables by employing other highly correlated variables, such as stock closing prices and the number of retail purchases. The results of this data imputation procedure are presented in a dedicated section.

It is worth noting that while the issue of missing data in the Pushshift Reddit dataset is overlooked by Allen et al. 2021; Hu et al. 2021 attempt to address it with missing Reddit variable indicators. However, the effectiveness of this approach is questionable, given that the main cause of missing data is the imperfections in Reddit data transfer to Pushshift, as identified in related literature.

The sentiment of comments was analyzed based on the frequency of negative words, following previous literature. The popularity of comments was determined by the Reddit score, or the number of upvotes minus the number of downvotes received by a post. Indicators for the prevalence of popular negative and non-negative *Wallstreetbets* opinions mentioning the stock were created for both regular trading and non-trading hours.

The research utilizes the Pushshift Reddit dataset alongside various financial data sources. Trading volume and transaction numbers at a 5-minute frequency are sourced from the Bloomberg database. Additionally, trade and quote millisecond data, constituting around 30 million observations for each company, were obtained for the period January 11 – February 4, 2021, from https://www.tickdatamarket.com. Using advanced Excel tools and Power-Query, I extracted 5-minute last trade prices and the preceding stock quote. From these, 5-minute last trade price stock returns were computed. This approach was chosen to control for market microstructure effects crucial for high-frequency stock return prediction. The same dataset was used to calculate retail trading activity in terms of volume and number of transactions, employing methodologies similar to Boehmer, Charles Jones, et al. 2021 for identifying retail transactions.

High-frequency data relating to shorting flow, both in volume and transaction numbers, were obtained from https://www.cboe.com. The options market data were also sourced from https://www.cboe.com. A meticulous matching process was conducted to align the financial data with the Reddit *Wallstreetbets* data, ensuring both datasets were computed at a 5-minute frequency.

Before the main analysis, I calculated conditional 5-minute stock return volatility for

GME and BB using the ARIMA-GARCH model. I accounted for intraday seasonality and made appropriate parametric assumptions regarding error distribution, choosing a suitable GARCH specification. The results were verified using the Mincer-Zarnowitz procedure. Intraday seasonality in trading volume and transaction numbers, as well as short trading volume and transaction numbers, were detected and adjusted using a pure diurnal dummy model.

The analysis revealed a contemporaneous negative association between GME's last trade price log-return and the prevalence of *Wallstreetbets*' popular negative opinions at a 5-minute frequency. This finding was robust even when retail order flow and shorting flow variables were included. The market microstructure variable showed a bid-ask bounce effect for GME. A positive association was found between the marketable retail order imbalance in terms of volume and the last trade price log-return for both GME and BB. Consistent with prior literature, shorting flow negatively predicted GME's 5-minute ahead last trade price log-return, while an excess of retail purchase volume over retail sell volume positively predicted it. The study also found that *Wallstreetbets*' popular opinions had no predictive power beyond a 5-minute horizon.

The study introduces a novel framework where the number of *Wallstreetbets* mentions is determined internally based. Innovations in *Wallstreetbets* traffic increased bullishness among GME and BB retail investors, with the impact on retail transactions dissipating after approximately 3.5 hours. The positive contemporaneous response in the number of short transactions reversed sharply after about 30 minutes for GME and 50 minutes for BB, becoming negative. This supports evidence that Reddit activity deters GME short sellers as it induces higher positive stock returns, which are unfavorable for short sellers (Allen et al. 2021; Hu et al. 2021; Pedersen 2022). An increase in informed trading (proxied by short trading) positively affected conditional return volatility for GME following a contemporaneous negative response. For BB, the response in conditional return volatility to an increase in informed trading was positive at all observed horizons, aligning with existing literature

(Glosten and Milgrom 1985a). Concerning the options market, both BB and GME showed an association between retail brokerage firm restrictions and increased GME in-the-money call and in-the-money put option open interest, consistent with the existing literature (Allen et al. 2021).

The paper is structured to first provide background on GME and describe the Reddit social media website. A literature review on related research is then presented, followed by a discussion on the peculiarities of modeling high-frequency financial data. The proxies for retail trading volume and transaction numbers, measures of short selling, and the predictability of short-horizon stock returns with a focus on market microstructure theories are examined. The variables and their computation methodologies are then described, followed by data visualizations and a correlation analysis. Finally, the empirical results are presented.

BACKGROUND

GameStop Corp. background

GME is a leading specialty retailer offering games and entertainment products through its e-commerce platforms and numerous stores across the United States, Canada, Australia, New Zealand, and several European countries including Austria, France, Germany, Ireland, Italy, and Switzerland (GameStop Corp. 2021b). During the first pandemic year, GME continued to shift its business online, resulting in the closure of 693 physical stores (GameStop Corp. 2021b). The company reported a 191% increase in global e-commerce sales in 2020, helping to mitigate the losses due to COVID-19 lockdowns (GameStop Corp. 2021b). George E. Sherman, Chief Executive Officer of GME, in the 2020 Letter to Shareholders, highlighted the company's commitment to evolving into "a customer-obsessed technology company that delights gamers", aligning with GME's "multi-year transformation initiative" named *GME Reboot* (GameStop Corp. 2021b).

GME stock experienced a 1,500% price surge in January 2021. The stock had previously shown volatility in 2020, with a standard deviation of inter-daily log-returns at 7.13%, compared to the S&P 500's 2.19%, based on data sourced from Yahoo Finance `https://uk.finance.yahoo.com/`. The GME 2020 Annual Report acknowledges that the early 2021 stock price and trading volume fluctuations were disproportionate to the company's operating performance or prospects (GameStop Corp. 2021b). The report attributes the volatility to factors like short squeezes, comments by analysts and on various media platforms, large stockholders altering their positions, and changes in short interest in GME Class A Common Stock, among others (GameStop Corp. 2021b).

Discussions on social media platforms, particularly by retail traders, have been linked to GME's dramatic price movements. These discussions often touted GME potential as an undervalued investment or its capacity to transition to e-commerce (U.S. Securities and Exchange Commission 2021). Some discourse also revolved around the potential for a short squeeze due to high short interest in GME. The media has portrayed this trading activity as a form of rebellion against professional short-sellers. John Authers 2021, a senior Bloomberg editor, described the situation as fueled by "righteous anger about generational injustice," terming it an "angry bubble".

GME discusses the possibility of selling shares or issuing equity securities, acknowledging that such actions could affect the market price of its Class A Common Stock (GameStop Corp. 2021b). However, the SEC granted GME the flexibility to sell equity securities. Meanwhile, the indenture for the company's 2023 Senior Notes limits the issuance of certain preferred stock or similar equity securities. In 2021, GME completed two at-the-money equity offerings, capitalizing on the elevated stock price, generating significant gross proceeds.

The 2020 Annual Report also addresses various risk factors associated with GME operations and financial performance, including restrictive covenants in the indenture governing its 2023 Senior Notes and its revolving credit facility (GameStop Corp. 2021b). These covenants potentially affect GME strategic transformation. On May 3, 2021, the company redeemed all of its outstanding 10% Senior Notes, effectively clearing its long-term debt (GameStop Corp. 2021b).

Furthermore, the report mentions an ongoing evaluation of the executive leadership team to meet changing business needs, hinting at potential changes in senior executive positions (GameStop Corp. 2021b). Changes in the Board of Directors since June 2020 also have introduced additional operational and strategic risks (GameStop Corp. 2021b).

According to the Schedule 13D (required by the SEC to be filed by the investor in the United States 10 days after the investor acquires more than 5% of any voting class of the company's equity) filed by Ryan Cohen and RC Ventures LLC on August 28, 2020, Ryan Cohen, a managing member of RC Ventures LLC, started new purchases of GME Class A Common Stock on August 13, 2020 (Securities and Exchange Comission 2020). Over August 13 – 18, 2020, Ryan Cohen purchased 1,466,100 shares, which is lower than 5% of around 65 mln GME shares of Common Stock outstanding in the second and third quarter of 2020, being approximately 3,250,000 shares (Securities and Exchange Comission 2020). It is implied, but not explicitly stated, that Ryan Cohen was GME Corp.'s shareholder prior to August 13, 2020. Assuming this, Cohen owned 2,050,047 (approximately 3.15% of the GME shares of Common Stock outstanding) GME shares before August 13, 2020, which can be inferred from the difference between 4,834,607 GME shares transferred and his new GME share purchases over August 13 – 25, 2020, amounting to 2,784,560. The date when Ryan Cohen's GME Class A Common Stock shareholding exceeded 5% was August 18, 2020, which is indicated in the Schedule (Securities and Exchange Comission 2020). As of August 18, 2020, Ryan Cohen owned 3,516,147 GME shares, which indeed slightly exceeded 5% of approximately 65 mln GME Class A Common Stock outstanding, or 3,250,000, and triggered the Schedule 13D filing. As of August 28, 2020, Ryan Cohen's GME stock holding reached 4,834,607, which he transferred to RC Ventures LLC (Securities and Exchange Comission 2020). RC Ventures LLC purchased 433,697 GME shares on August 27, 2020, received an internal transfer of 4,834,607 GME shares from Ryan Cohen, and additionally purchased 531,696 GME shares on August 28, 2020 (Securities and Exchange Comission 2020). In total, as of August 28, 2020, RC Ventures LLC owned 5,800,000 shares of GME Class A Common Stock, or 9% of the corresponding share class (Securities and Exchange Comission 2020). Over August 13 – 28, 2020, GME Common Stock adjusted close price rose from \$4.64 to \$5.39, which amounts to a 16% increase. On August 31, 2020, intra-daily GME share price increase amounted to 24%, which is the largest intraday gain since mid-April 2020. Looking back at the GME stock price dynamics over 2020, it can be concluded that the purchase of approximately 6% of the GME shares of Common Stock outstanding by RC Ventures LLC and Ryan Cohen over the second half of August 2020 was at the origin of the steady GME price increase over the remaining 2020 – beginning of 2021 (see Figure 1). In addition, there is evidence, as suggested by Jensen and Ruback 1983, that stock return of a target firm tends to increase around the M&A deal announcement. While there was no deal announcement regarding GME, the Schedule 13D filing has been a powerful market signal that triggered a response on behalf of the investors, and, consequently, affected the stock price.

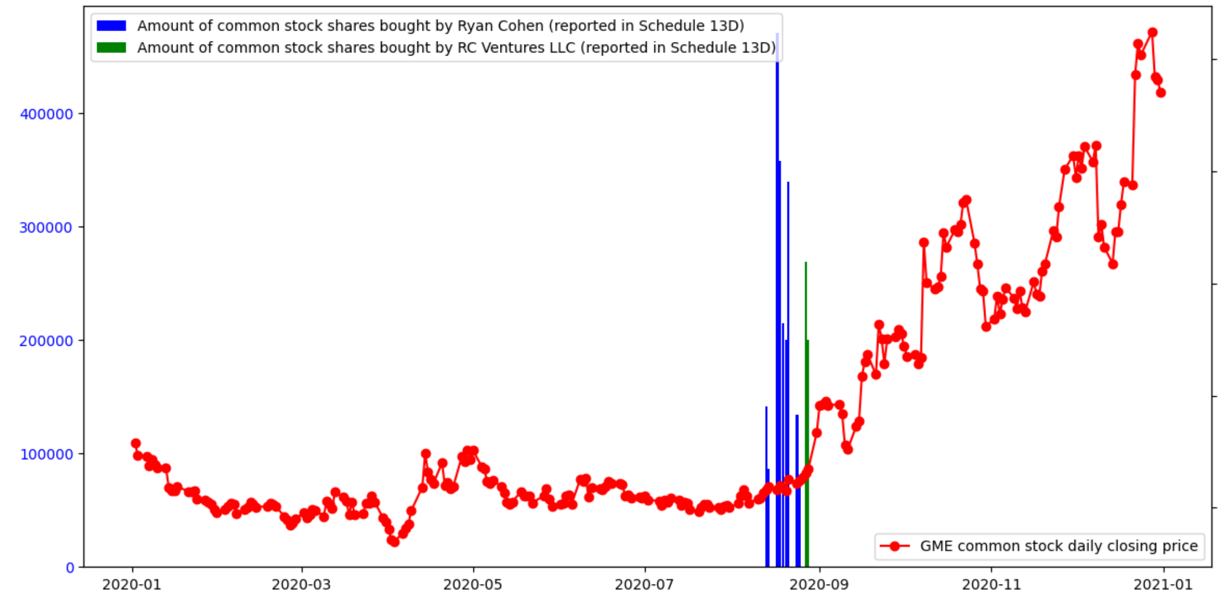

Figure 1: The figure displays daily closing price trend of GME common stock (in red) starting from January 2020 and extending into January 2021. On the right y-axis two types of GME common stock purchase transactions are represented by bars: one set by Ryan Cohen and another by RC Ventures LLC, distinguished by blue and green colors. The information about GME common stock purchase transactions is sourced from Schedule 13D filed by Ryan Cohen and RC Ventures LLC on August 28, 2020 (Securities and Exchange Comission 2020). Daily closing price data is sourced from Yahoo Finance https://uk.finance.yahoo.com/.

On November 16, 2020, Ryan Cohen, on behalf of RC Ventures LLC and as a major stockholder of GME at the time, communicated with GME's board of directors through a letter (Cohen 2020). He strongly suggested that GME's leadership must urgently review their strategy and present a plan to leverage the growing opportunities in the gaming sector. Cohen highlighted GME's reluctance to adapt to market changes, leading to missed revenues and a decline in shareholder value (Cohen 2020). He specifically critiqued GME's Chief Executive Officer George Sherman for sticking to an outdated business model focused on physical stores, in contrast to the digital trend in the industry (Cohen 2020). Cohen, leveraging his experience as the former Chief Executive Officer of Chewy Inc., pointed out the successful strategies at Chewy that GME lacked and indicated possible areas for internal investment (Cohen 2020).

By January 10, 2021, Cohen's stake in GME had increased to 9,001,000 shares, or 12.9% of the Class A Common Stock (GameStop Corp. 2021a). Subsequently, on January 11, 2021, GME broadened its board by including Alan Attal, former Chief Marketing Officer of Chewy Inc., along with Cohen and Jim Grube, the former Chief Financial Officer of Chewy, Inc. (GameStop Corp. 2021a). This addition was aimed at enhancing GME's capabilities in e-commerce, online marketing, finance, and strategic planning, in line with their goals to improve core operations and create a dynamic ecosystem for games and entertainment (GameStop Corp. 2021a).

Ryan Cohen's appointment to GME board of directors highlighted the company's potential for a promising future. This optimism was rooted in Cohen's proven track record, notably his success with Chewy Inc., an online pet supply retailer he co-founded with Michael Day in 2012. In 2017 Chewy Inc. was sold to PetSmart for $3.35 billion, marking it as the largest e-commerce deal at the time. The enthusiasm surrounding Cohen's joining GME resonated with retail investors, who actively discussed the company's prospects online. According to Allen et al., many user posts on online platforms revealed that they had taken long positions in GME, encouraging others to do the same (Allen et al. 2021). A significant portion of this discussion took place within the Reddit community *Wallstreetbets*, a forum known for its members' engagement in conversations about stock and options trading.

Reddit social media

Reddit is a social media platform that, as of 2021, had more than 100,000 active communities connecting individuals with shared interests (Reddit Inc. 2021). In October 2020, Reddit had 52 million daily active users, marking a 44% increase from October 2019 (Patel 2020). In 2020, Reddit ranked as the tenth most popular social networking site in the United States (Pew Research Center 2021), with users spending an average of about 10 minutes per visit (SimilarWeb 2020).

A Reddit user can engage with an online community through traditional channels such as

- creating a submission,
- posting a comment to a submission, and
- giving upvotes, downvotes, or awards to content.

Consequently, each post receives a score, which is the number of upvotes minus the number of downvotes, visible to all Reddit users. The informal set of principles, known as *Reddiquette*, advises users to upvote content that contributes to the conversation and downvote that which does not or is off-topic within a particular community (Reddit Help 2022). The interpretation of upvoting and downvoting remains a topic of active discussion. Giving an *award* is another way to acknowledge each other's contributions. Posts or comments that have been *awarded* are often highlighted and can only be given in exchange for Reddit Coins – a virtual currency purchasable with real money. According to Reddit, granting an award is a way to express appreciation for an exceptional contribution to the platform, though it does not directly affect the post's visibility.

The publicly available Reddit implementation code reveals that the submissions sorting algorithm favors newer submissions over older ones with the same number of net votes and that the value of votes follows a logarithmic scale where initial votes count more than subsequent ones (Salihefendic 2015). The comments' sorting algorithm, on the other hand, is designed to diminish the importance of the comment's submission time to prevent a feedback loop (Salihefendic 2015). In conclusion, a post's score is the sole metric that directly influences its visibility on Reddit.

The Reddit Pushshift API is a social media data collection, analysis, and archiving platform that has collected Reddit data since 2015 and made it available to researchers. It is a project initiated by Jason Baumgartner that maintains a copy of all comments and submissions posted on Reddit (Baumgartner et al. 2020). In addition, the Reddit Pushshift API provides analytics necessary for analyzing Reddit content engagement. The Pushshift API offers enhanced functionality compared to the official Reddit API (Baumgartner et al. 2020). The Pushshift Reddit dataset is widely used in academic empirical research and is attractive to researchers due to its declared completeness (Baumgartner et al. 2020; Gaffney and Matias 2018).

The Pushshift Reddit output provides the following user-interaction-related parameters for comments:

1. *all_awardings* details all the various types of awards a comment has received;

2. *total_awards_received* indicates the total number of awards a comment has received;

3. *gildings* shows the number of Silver (*gid_1*), Gold (*gid_2*), or Platinum (*gid_3*) awards a comment has received;

4. *score* reveals the net number of votes a comment has received (the number of upvotes minus the number of downvotes).

For submissions, the Pushshift Reddit output includes additional user-interaction-related parameters:

1. *num_comments* indicates the number of comments related to a submission;

2. *num_crossposts* shows the number of times a submission was crossposted (shared in another Reddit community);

3. *upvote_ratio* represents the submission's ratio of upvotes to total votes (upvotes plus downvotes).

It is important to note that the parameters outputted by Pushshift for Reddit are not updated in real time but at timed intervals.

"Gamestop" or "gme" (the search is not case-sensitive) was first mentioned in a submission on the subreddit *Wallstreetbets* on December 10, 2014. The author of the post highlighted "how obsolete GME is becoming as many consoles allow users to download games directly from their online marketplaces," discussed the company's poor financial prospects, noted the stock's relatively high short interest of 35%, and outlined a strategy of going "a long-term short of $GME" (Reddit, subreddit *Wallstreetbets*, posted by u/MerBank on December 10$^{\text{th}}$, 2014). Subsequent mentions of GME on *Wallstreetbets* between 2014 and 2017 were relatively infrequent (111 in total) and mostly bore a bearish sentiment. Meanwhile, "gme" or "gamestop" and "short squeeze" were mentioned in three *Wallstreetbets* submissions in 2016, one in 2017, four in 2018, and seven in 2019. Overall, "gamestop" or "gme" was mentioned more than 200,000 times in submissions and over 1,200,000 times in comments on *Wallstreetbets* as of March 2022.

LITERATURE REVIEW

Relevant previous research

The January to February 2021 meme stock phenomenon, characterized by the extreme volatility of share prices in approximately 12 stocks, captivated media, retail and institutional investors, and regulators. The SEC Staff issued a report on October 14, 2021, analyzing market structure conditions in early 2021 (U.S. Securities and Exchange Commission 2021).

U.S. Securities and Exchange Commission 2021 noted a correlation between the rise in GME prices in January 2021 and an increase in the number of individual accounts trading GME. Additionally, U.S. Securities and Exchange Commission 2021 highlighted significant short interest in GME by some institutional accounts prior to January 2021, influenced by GME 's weak financial performance, making it an attractive target for short sellers. Yahoo Finance's web archive revealed that as of September 7, 2020, 153% of GME shares were held by 287 institutions, a figure that decreased by 22% by November 25, 2020. This high institutional holding percentage of above 100% indicated that many institutions were short sellers in GME, with multiple entities claiming ownership of the same shares. The Yahoo Finance's web archive also showed that GME's short interest was 97.68% on December 14, 2020, and peaked at 109.26% on December 31, 2020. Stocks with such high short interest, considerably exceeding market averages, are typically expected to yield negative excess returns. Going short on an asset is considered worthwhile if the expected transaction benefits outweigh the lending fees.

Post-January 11, 2021, online discussions about GME, marked by bullish sentiment and a resolve to purchase and hold the stock, drew an increasing number of retail investors who opted for long positions in GME. This trend forced short sellers of GME to buy back the stock, aiming to minimize losses from the rally, as stated in GameStop Corp. 2021b Annual Report. This buying activity contributed to a further increase in GME's price and prompted purchases to cover short positions. The aggregate short exposure, exceeding available shares, further drove the price increase. GME's short percentage of float fluctuated around 250% in late 2020 to early 2021, dropping to 226% on January 14, 2021. U.S. Securities and Exchange Commission 2021 observed that GME price surges correlated with a decrease in short interest, particularly as major short sellers repurchased stock around January 22, 2021. However, such buying constituted a small cfraction of the total buy volume, and GME prices remained high even after the effects of covering short positions diminished (U.S. Securities and Exchange Commission 2021). U.S. Securities and Exchange Commission 2021 concluded that it was the positive sentiment, rather than covering short positions, that sustained the price appreciation.

U.S. Securities and Exchange Commission 2021 also considered the possibility of a *gamma squeeze* in January 2021, which occurs when market makers buy stock to hedge risks associated with writing call options. While this phenomenon was explored, U.S. Securities and Exchange Commission 2021 found no evidence of a gamma squeeze in GME during this period. Empirical data suggested that increased trading in GME put options led to a rise in individual investors' GME options trading volume, with market-makers primarily purchasing call options (U.S. Securities and Exchange Commission 2021). In January 2021, the increased trading volume was heavily concentrated in call options, many of which were short-dated (U.S. Securities and Exchange Commission 2021). Over 66% of individual customer accounts trading GME options were represented by online retail brokers such as Robinhood, TD Ameritrade, and E*Trade Securities, with Robinhood and TD Ameritrade accounting for over 50% of the total GME options volume traded by individual investors (U.S. Securities and Exchange Commission 2021).

The January to February 2021 meme stock phenomenon, marked by extreme volatility in the share prices of about 12 stocks, drew significant attention from media, retail and institutional investors, and regulators. Contrary to the SEC Staff's assertion that hedge funds were not significantly affected by investments in GME and other meme stocks, the U.S. Securities and Exchange Commission 2021 noted that GME's volatility significantly impacted exchange-traded funds holding GME stock, particularly XRT, an exchange-traded fund comprising equal-weighted U.S. retail company stocks. XRT's holdings in GME shares exceeded its own outstanding shares, rendering XRT's share price extremely sensitive to GME's price fluctuations (U.S. Securities and Exchange Commission 2021). Furthermore, one institutional investor heavily affected by GME's rising stock price was Melvin Capital Management LP (U.S. Securities and Exchange Commission 2021).

U.S. Securities and Exchange Commission 2021 also pointed out that the cost of short selling GME significantly surpassed the average lending fee during 2020 and early 2021. Coupled with the stock's substantial price swings, this made short selling GME less appealing (U.S. Securities and Exchange Commission 2021). The SEC Report drew parallels with a Beber and Pagano 2013 study, which examined the impact of short selling restrictions following the 2007-2009 financial crisis. These restrictions, while intended as a remedial measure, were found to impair market liquidity, reduce market efficiency, and generally not enhance stock price performance (Beber and Pagano 2013).

High volatility in GME's stock price prompted clearing agencies to increase margin requirements to offset heightened credit risk, leading some broker-dealers, particularly those with a large individual investor customer base, to restrict trading in certain stocks, including GME (U.S. Securities and Exchange Commission 2021). U.S. Securities and Exchange Commission 2021 underscored that the end of January 2021 saw dramatic increases in volatility, price, and volume, along with multiple volatility-induced trading disruptions and deterioration in some liquidity measures (U.S. Securities and Exchange Commission 2021).

These events underscored the significant impact of social media on securities markets and the role of social media in potential asset price manipulation. While short squeezes, as described by Allen et al. 2021, are generally illegal, it remains unclear whether the coordination undertaken by investors in these instances falls under existing stock market trading regulations. *Securities Exchange Act of 1934* 1934 delineates two broad categories of market manipulation: action-based and information-based. Information-based manipulation involves the deliberate spread of false information to influence trading activity, whereas action-based manipulation entails transactions that create the appearance of active trading or price manipulation to induce others to trade (*Securities Exchange Act of 1934* 1934). However, as Bloomberg Opinion columnist Matt Levine observed, there was no significant deception about company fundamentals in the GME case (Levine 2021). With extensive media coverage suggesting that Redditors were pushing GME prices up "for fun," investors were generally well-informed about the nature of the stock rally, engaging in activities not expressly prohibited by current U.S. law (Levine 2021).

One of the earliest empirical studies on the January 2021 short squeezes was conducted by Allen et al. 2021, presented in their work "Squeezing Shorts Through Social Media Platforms" in November 2021. The researchers proposed that the sudden spikes in stock prices were due to coordinated actions by traders on social media platforms, marking the first instance of such coordinated efforts targeting stocks with high short interest (Allen et al. 2021). The study emphasized the pivotal role of evolved social media platforms in enabling this type of coordination (Allen et al. 2021). Allen et al. 2021 concluded that this was the first instance of social media being used to orchestrate trading strategies across a large group of traders, with much of this trading attributed to retail traders by the public press.

Both regulators and researchers have noted the growing trend of increased retail investor participation in trading, associated with risks that sudden surges in trading can significantly deviate prices from fundamental values, especially for less liquid securities. Corey Hoffstein, Co-Founder and Chief Investment Officer at Newfound Research, commented on the GME stock price surge, observing an influx of speculative retail traders unconcerned with valuation theories, a sentiment diverging from the traditional Wall Street view that market dynamics are fundamentally value-driven (Hoffstein 2021).

Further, Allen et al. 2021 countered the U.S. Securities and Exchange Commission 2021 Report by providing evidence of a short squeeze in GME and other stocks. Findings of Allen et al. 2021 indicated a significant decrease in short selling activity, an increase in stock average fees, and a shift towards the options market during the price surge, pointing to constraints in the equity lending market indicative of a short squeeze. Allen et al. 2021 also found a strong correlation between retail traders' activities and the returns of stocks during the short squeeze period, suggesting that retail traders contributed to the price increases.

Allen et al. 2021 observed a trend of retail investors moving from stock to options markets, driven by a desire to circumvent trading restrictions imposed by some online retail brokers. It was noted that retail traders relied on call options to express optimistic views about the relevant stocks, exacerbating the squeeze events (referred to as "gamma squeeze", as mentioned previously), while other market participants used put options during and after the squeezes (Allen et al. 2021). Allen et al. 2021 concluded that market quality deteriorated for the stocks in question during the short-squeeze episodes and while retail activity was at its peak.

The influence of coordinated trading on stock markets, based on research analyzing herding behavior, suggests that attention on social media platforms contains valuable information for predicting stock returns and increasing trading volumes (Allen et al. 2021). Herding, characterized by a quick and easy public dissemination of beliefs, particularly through social media, attracts less experienced investors driven not necessarily by rational decision-making (Allen et al. 2021). Theoretical models indicate that both rational and irrational herding behavior can generate contagion, leading to market bubbles or crashes (Allen et al. 2021).

Pedersen 2022 provided a theoretical framework for understanding the anatomy of a bubble, exemplified by the GME stock rally in early 2021. The Pedersen 2022 model predicts that trading behavior spreads through social networks, starting with "fanatic agents" focusing solely on positive aspects of an asset, followed by sophisticated traders anticipating stock price increases. This leads to an "echo chamber effect", where a growing number of people seek wealth without fully understanding the underlying processes, eventually resulting in a bubble that bursts when rational short-term investors reverse their positions (Pedersen 2022).

Hu et al. 2021 conducted a thorough examination of the relationship between Reddit activity and meme stock returns, expanding upon Allen et al. 2021 analysis which primarily focused on the frequency of stock mentions on Reddit. Hu et al. 2021 incorporated additional variables such as the tone of Reddit posts, the dispersion of social media post tones, and the connectedness of Reddit users participating in stock discussions. The findings indicate that social media activity variables, including the tone and traffic of Reddit posts, have a positive and statistically significant impact on next-day stock returns (Hu et al. 2021). Furthermore, Hu et al. 2021 observed that increased Reddit traffic amplifies the predictive power of retail order flow, aligning with the spillover effect identified by Hu et al. 2021; Pedersen 2022. Interestingly, Reddit traffic was found to have a more sustained predictive power than variables related to post tone and user connectedness (Hu et al. 2021).

During the short squeeze period from January 26 to February 4, 2021, Allen et al. 2021 noted significant changes in daily option open interest for specific types of call and put options across thirteen meme stocks impacted by trade restrictions from retail brokerage firms. Allen et al. 2021 interpreted the rise in interest for in-the-money call options as a mechanism for expressing optimistic views on these stocks, potentially circumventing trading restrictions. Conversely, the increase in out-of-the-money put options was seen as a way to express pessimistic views or to navigate restricted short-selling opportunities in the equity market (Allen et al. 2021).

Chen et al. 2014 investigated the impact of user-generated opinions on stock market performance. Chen et al. 2014 found that the views could predict future stock market performance, suggesting that social media platforms could either incite naïve investor reactions based on spurious information or provide value-relevant insights not yet reflected in stock prices.

Chen et al. 2014 identified several mechanisms behind the predictability of stock market performance from social media:

- users gain utility from the attention and recognition received for posting opinions that align with market trends;

- social media platforms enable direct interaction and public feedback, allowing users to challenge or correct misguided views, thereby enhancing the informativeness of these platforms;

- informed actors may use social media to publicize their investment ideas and persuade others to adopt their strategies, expediting the convergence of market prices to perceived fundamental values.

Allen et al. 2021 focused their analysis on thirteen stocks central to social media platform discussions, which were subjected to trading restrictions by brokers. These stocks are presented below in Table 1 with the number of mentions in subreddit *Wallstreetbets* over of January 11, 2021 – February 4, 2021. I have manually collected the corresponding numbers using aforementioned Pushshift Reddit dataset.

I suggest considering companies with total mentions exceeding 100,000 for more representative conclusions in further analysis. These companies include GME, AMC, BB, and NOK. Notably, the total mentions of GME or GME surpass those of AMC by more than double.

The following Table 2 illustrates the maximum inter-daily increase in percentage terms in the closing stock prices of these companies. The peak in stock closing prices occurred on January 27, 2021, compared to the closing price on January 11, 2021.

On January 26, 2021, at 16:08 EST, Elon Musk tweeted a link to the Reddit community Wallstreetbets with the word "Gamestonk!!" (Musk 2021). This tweet caused a 60% increase in GME's stock price during after-hours trading (CNBC 2021). By the morning of January 27, 2021, the overnight log-return for GME's stock, calculated from the last trade price on January 26 at 16:00 EST to the last trade price on January 27 at 09:30 EST, was 0.81, indicating an 81% surge in stock price.

Following the peak in inter-daily stock prices for several meme stocks, including GME,

Company (Ticker Symbol)	Query	Mentions in Comments	Mentions in Submissions	Total Mentions
GameStop Corp. (GME)	Gamestop |GME	562,128	143,379	705,507
AMC Entertainment Holdings Inc. (AMC)	AMC	222,114	63,440	285,554
BlackBerry Ltd. (BB)	BlackBerry |BB	134,824	25,374	160,198
Nokia Corp. (NOK)	Nokia |NOK	103,323	26,877	130,200
Sundial Growers Inc. (SNDL)	SNDL	6,019	4,686	10,705
Naked Brand Group Inc. (NAKD)	NAKD	4,246	5,260	9,506
Bed Bath & Beyond Inc. (BBBY)	BBBY	5,560	1,335	6,895
American Airlines Group Inc. (AAL)	AAL	2,010	1,144	3,154
Express Inc. (EXPR)	EXPR	1,275	1,506	2,781
Koss (KOSS)	KOSS	1,149	1,279	2,428
Castor Maritime Inc. (CTRM)	CTRM	551	1,023	1,574
Tootsie Roll Industries Inc. (TR)	TR |TootsieRoll	493	207	700
Trivago N.V. (TRVG)	TRVG |Trivago	181	122	303

Table 1: Number of mentions of various companies in subreddit *Wallstreetbets* sorted in descending order by total mentions. Data covers the period from January 11, 2021, to February 4, 2021, sourced from the Pushshift Reddit dataset.

AMC, BB, and NOK on January 27, 2021, retail brokers imposed restrictions on purchasing thirteen companies' stocks on January 28, 2021. Customers could only sell these stocks. Robinhood announced these restrictions at 09:00:24 EST, just before the start of regular trading hours, citing the inability to meet the surged collateral requirements from clearing houses due to increased market volatility, as explained in a public statement by retail brokers on January 29, 2021. On January 30, 2021, Robinhood expanded the list of companies with trading restrictions (Allen et al. 2021).

Coinciding with the start of trading restrictions on January 28, 2021 (day 13 in the

Company (Ticker Symbol)	**Maximum Day % Close Price Increase**	**Date**
Gamestop Corp. (GME)	1,643%	27.01.2021
AMC Entertainment Holdings Inc. (AMC)	805%	27.01.2021
BlackBerry Ltd. (BB)	228%	27.01.2021
Nokia Corp. (NOK)	69%	27.01.2021

Table 2: Maximum inter-daily % close price increase over January 11 – February 4, 2021 relative to the close price on January 11, 2021. Data come from the Bloomberg database.

sample), the prices of GME, AMC, BB, and NOK began to decline. Despite this decline, some Reddit users rallied to convince others to hold onto their shares, arguing either that the prices would increase in value or that holding would send a political message (Allen et al. 2021). On February 4, after market hours, Robinhood lifted all restrictions on long positions (Allen et al. 2021).

Following the logic of Allen et al. 2021, the short squeeze period is defined as January 26 through February 4, in total nine days. Attention-getting stocks tend to experience mean-reversion in prices after a temporary price increase, within approximately five trading days after the attention-grabbing event (Seasholes and Wu 2007).

Financial Data Modelling

To take advantage of the highly granular textual data available from the Pushshift Reddit dataset, I complement this textual data with high-frequency financial data from Bloomberg database, `tickdatamarket.com`, and `cboe.com`. In the current study, the following Reddit activity-related variables are computed: negative word frequency and tone of submissions and comments, score of submissions and comments, and the number of comments that submissions received. Both Reddit activity-related variables and financial data are collected at 5-minute intervals over regular trading hours, i.e., from 09:30 to 16:00 EST.

The research focuses on the most dramatic phase of the run-up in meme stocks' prices, from January 11, 2021, to February 4, 2021, spanning 18 trading days. The starting date is defined by the positive news regarding GME, the meme stock that attracted the most media attention. The ending date is defined by the end of the short-squeeze period (Allen et al. 2021).

Using high-frequency financial data allows us to avoid the problems associated with a small sample size (18 trading days), such as inflated standard errors and consequently low power of statistical tests. This approach ensures a more robust analysis of the interaction between Reddit activity and stock price movements during this period.

High-frequency financial data, defined by Tsay 2002 as "observations taken daily or at a finer time scale" and "transaction-by-transaction or trade-by-trade data in security markets", has unique characteristics. These include unequally spaced time intervals between transactions, discrete-valued prices, multiple transactions within a single second, and, most importantly, the existence of a daily periodic or diurnal pattern, which is an intraday seasonality. The authors suggest that this diurnal pattern can be illustrated by the empirical observation that on the NYSE transactions are heavier at the beginning and closing of trading hours and thinner during lunch hours, resulting in a "U-shape" transaction intensity. According to Dacorogna et al. 2001, a similar "U-shape" pattern is observed for high-frequency return volatility, with high volatility at the opening, a subsequent decrease, and an increase just before closing. The authors conclude that, in general, return volatility peaks at the beginning of a trading day.

Diurnality introduces a deterministic trend in intraday returns, volatilities, trading volumes, etc., making these time series related over time or non-stationary. Estimation of non-stationary time series can result in spurious regressions, which demonstrate statistically significant regression coefficients in the absence of an actual relationship between the dependent variable and regressors. Additionally, it may lead to inconsistent estimators and invalid statistical tests. Therefore, achieving stationarity of time series is crucial before proceeding with further analysis. Stationarity can be inspected via a sample correlogram (initial exploratory analysis) and tested formally.

Trend stationarity can be tested by regressing the time series directly on time and checking for the significance of the time variable. Random walk type of non-stationarity should be tested using the Dickey-Fuller test or the Augmented Dickey-Fuller test. It is also interesting to note that there is no significant difference between log-returns and simple returns when calculated on the basis of high-frequency financial data.

Due to these peculiarities, traditional methods applicable to low-frequency financial data analysis are not suitable for high-frequency financial data. The distributions of returns become increasingly fat-tailed as data frequency increases (smaller interval sizes), making them distinctly unstable (Dacorogna et al. 2001). Additionally, there is evidence of seasonal heteroskedasticity in the form of distinct daily and weekly clusters of volatility. These characteristics necessitate specific approaches for the analysis and modeling of high-frequency financial data to ensure accurate and reliable results.

There are two common models to capture the seasonality: the diurnal dummy variable model and the polynomial trend model (Dacorogna et al. 2001). The diurnal dummy variable model involves using L dummy variables to represent different time intervals over a trading day. Each interval is typically 30 minutes. The log-returns r_{t+k} over shorter time intervals (e.g., 5 minutes) are modeled as:

$$r_{t+k} = \sum_{i=1}^{L} \beta_i D_{i,t+k} + u_{t+k},$$

where $D_{i,t}$ are of dummy variables for each time interval, β_i represents the intercepts for each interval, and u_{t+k} is the error term. Each dummy variable represents a different intercept for the corresponding time interval. To avoid the dummy variable trap, the model does not include an overall intercept. Test for the presence of diurnality is conducted as follows. The null and alternative hypotheses are:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_L \quad \text{(Diurnality is absent)}$$
$$H_a : \beta_i \neq \beta_j \quad \text{for at least one pair } (i, j)$$

Under H_0, the test statistic is:

$$\chi^2(L-1) = \sum_{i=1}^{L} \left(\frac{\hat{\beta}_i - \bar{\beta}}{s_i} \right)^2$$

As mentioned previously, it is expected that intra-daily log returns exhibit heteroscedasticity, therefore, s_i must be a robust standard error.

Diebold 2006 introduced another approach to capture diurnality using a polynomial trend model. This method, while less flexible than a pure diurnal dummy variable model but more parsimonious. Instead of dummy variables, a new variable HRS_{t+k} (or MINS_{t+k}) is introduced, representing the number of hours (minutes) passed from midnight until time $t+k$. Let us consider a quadratic trend model:

$$r_{t+k} = \beta_0 + \beta_1 \mathrm{HRS}_{t+k} + \beta_2 \mathrm{HRS}^2_{t+k} + u_{t+k}$$

or

$$r_{t+k} = \beta_0 + \beta_1 \text{MINS}_{t+k} + \beta_2 \text{MINS}_{t+k}^2 + u_{t+k}$$

The test for the presence of diurnality is conducted as follows. The null and alternative hypotheses are:

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{(Diurnality is absent)}$$
$$H_a : \text{at least one } \beta_i \neq 0,\ i = 1, 2$$

Standard t-tests can be used to test the significance of variables HRS_{t+k} and HRS_{t+k}^2 (or MINS_{t+k} and MINS_{t+k}^2). However, the joint power of such statistical tests will be less than declared. Alternatively, a standard F-test or chi-squared test for goodness of fit for joint significance of the coefficients of variables HRS_{t+k} and HRS_{t+k}^2 (or MINS_{t+k} and MINS_{t+k}^2) can be conducted.

In practice, diurnality in intraday volatility is often a more critical issue than diurnality in intraday returns. GARCH-type models are limited to data with constant unconditional variance, making them unsuitable for raw intraday data. A common solution is to decompose variation in volatility into a deterministic (diurnal) component and a stochastic (GARCH) component (Patton 2015).

Let r_t be a log-return and μ_t be the conditional mean of log-return r_t at time step t. μ_t can be decomposed into an ARMA(p, q) part and a possible seasonal effect denoted as S_t^r. Then, the return model can be represented as follows:

$$\begin{aligned} r_t &= \mu_t + \epsilon_t, \\ \mu_t &= \omega + \sum_{i=1}^{p} \alpha_i r_{t-i} + \sum_{j=1}^{q} \beta_i \epsilon_{t-j} + S_t^r, \\ \epsilon_t &\sim (0, \sigma_t^2). \end{aligned} \tag{1}$$

Let S_t^ϵ be a seasonal component of residual ϵ_t. Let ϵ_t^* be a deseasonalized residual and σ_t^* be a deseasonalized return volatility. Then,

$$\epsilon_t \overset{(1)}{=} \sigma_t \nu_t \overset{(2)}{=} S_t^{\epsilon} \sigma_t^* \nu_t \overset{(3)}{=} S_t^{\epsilon} \epsilon_t^*, \tag{2}$$

where $\nu_t \sim$ i.i.d.$(0, 1)$. Let me briefly explain the Equation 2. The equality 1 illustrates the multiplicative structure of residuals. The equality 2 demonstrates that the conditional volatility σ_t is divided into a seasonal component S_t^{ϵ} and a GARCH component σ_t^*. The equality 3 shows that the GARCH component σ_t^*, along with the independent and identically distributed variable ν_t constitute a deseasonalized residual denoted as ϵ_t^*.

The ultimate objective of these subsequent manipulations is to determine an estimate for the seasonal component S_t^{ϵ} within the log return residual ϵ_t.

The GARCH model is based on a multiplicative framework. However, the aforementioned seasonal adjustment procedure follows an additive structure. As a result, it is necessary to convert the GARCH model into an additive form, for instance, by employing logarithms. Additionally, it is important to highlight that the input to a logarithmic function must always be non-negative. Consequently, prior to applying logarithms, both the right- (ϵ_t) and left-hand ($S_t^{\epsilon} \epsilon_t^*$) sides of the Equation 2 need to be squared.

$$\begin{aligned} \epsilon_t^2 &= S_t^{\epsilon 2} \epsilon_t^{*2}, \\ \ln(\epsilon_t^2) &= \ln(S_t^{\epsilon 2}) + \ln(\epsilon_t^{*2}), \\ \ln(\epsilon_t^2) &= \ln(S_t^{\epsilon 2}) + u_t, \end{aligned} \tag{3}$$

where $u_t = \ln(\epsilon_t^{*2})$.

Next, to produce an estimate $\widehat{\ln(S_t^{\epsilon 2})}$ of the logarithm of the squared seasonal component $\ln(S_t^{\epsilon 2})$, one can employ a pure diurnal dummy variable model as discussed earlier:

$$\begin{aligned} \ln(\epsilon_t^2) &\overset{(1)}{=} \widehat{\ln(S_t^{\epsilon 2})} + \widehat{u}_t, \\ \widehat{\ln(S_t^{\epsilon 2})} &\overset{(2)}{=} \sum_{i=1}^{16} \beta_i D_{i,t}. \end{aligned} \tag{4}$$

Further, exponentiate both sides of the equality 1 from Equation 4 to arrive at the following expression:

$$\begin{aligned}\epsilon_t^2 &= \exp\{\ln(\widehat{S_t^{\epsilon 2}}) + \widehat{u}_t\} \\ &= \exp\{\ln(\widehat{S_t^{\epsilon 2}})\} \exp\{\widehat{u}_t\}.\end{aligned} \tag{5}$$

Note that $\exp\{\ln(\widehat{S_t^{\epsilon 2}})\}$ is a biased estimator of $S_t^{\epsilon 2}$. Therefore, it is usually assumed that $\widehat{S_t^{\epsilon 2}} \coloneqq E[\widehat{\epsilon_t^2 | S_t^\epsilon}]$.

Next, the value $E[\epsilon_t^2 | S_t^\epsilon]$ can be expressed as follows:

$$\begin{aligned}E[\epsilon_t^2 | S_t^\epsilon] &\overset{(1)}{=} E[\exp\{\ln(\widehat{S_t^{\epsilon 2}})\} \exp\{\widehat{u}_t\} | S_t^\epsilon] \\ &\overset{(2)}{=} \exp\{\ln(\widehat{S_t^{\epsilon 2}})\} E[\exp\{\widehat{u}_t\} | S_t^\epsilon] \\ &\overset{(3)}{=} \exp\{\ln(\widehat{S_t^{\epsilon 2}})\} E[\exp\{\widehat{u}_t\}].\end{aligned} \tag{6}$$

Allow me to provide a more detailed explanation of the derivations for Equation 6. The equality 1 arises from substituting Equation 5 into the conditional expectation operator. The equality 2 is a result of factoring out $\exp\ln(\widehat{S_t^{\epsilon 2}})$ from the conditional expectation operator. This factor represents an expression solely dependent on $S_t^{\epsilon 2}$, which the expectation is conditioned upon. The equality 3 holds true when it is assumed that u_t is independent of S_t^ϵ.

Additionally, in order to formulate an estimate for $E[\epsilon_t^2 | S_t^\epsilon]$, denoted as $E[\widehat{\epsilon_t^2 | S_t^\epsilon}]$, in place of the expectation operator $E[\exp\{\widehat{u}_t\}]$ within Equation 6, one can make use of a sample average through the following substitution:

$$\begin{aligned}\widehat{S_t^{\epsilon 2}} &\overset{(1)}{\coloneqq} E[\widehat{\epsilon_t^2 | S_t^\epsilon}] \\ &\overset{(2)}{=} \exp\{\ln(\widehat{S_t^{\epsilon 2}})\} E[\exp\{\widehat{u}_t\}] \\ &\overset{(3)}{\approx} \exp\{\ln(\widehat{S_t^{\epsilon 2}})\} \frac{\sum_{t=1}^{T} \exp\{\widehat{u}_t\}}{T},\end{aligned} \tag{7}$$

where T is a sample size. Equality 1 corresponds to the assumption introduced earlier.

Equality 2 emerges from Equation 6. Approximation 3 arises through the replacement of the expectation operator with a sample average.

In conclusion, the deseasonalized residual estimate $\widehat{\epsilon_t^*}$ is calculated using the following procedure and subsequently employed in the estimation of a deseasonalized GARCH(1, 1) model, as an illustration:

$$\begin{aligned} \widehat{\epsilon_t^*} &= \frac{r_t - \widehat{\mu_t}}{\sqrt{\widehat{S_t^{\epsilon 2}}}}, \\ \epsilon_t^* &\sim \text{i.i.d.}(0, \sigma_t^*), \\ \sigma_t^{*2} &= \omega + \alpha\epsilon_{t-1}^* + \beta\sigma_{t-1}^{*2}. \end{aligned} \tag{8}$$

Note that the distribution of ϵ_t^* is not specified and should be assumed through an *ad hoc* procedure, considering the characteristics of empirical distribution for $\widehat{\epsilon_t^*}$. High-frequency financial time series data tend to exhibit severe kurtosis, hence the common assumption of conditional normality should be loosened by considering alternative fat-tailed distributions such as the Student-t distribution for ϵ_t (Baillie and Bollerslev 1989).

In brief, the main goal of conditional volatility modeling is to obtain a one-step-ahead forecast for the conditional variance. As previously mentioned, the GARCH model provides such a forecast. Therefore, the process for obtaining a one-step-ahead forecast for the conditional variance is as follows:

$$\begin{aligned} V_{t-1}(\epsilon_t) &= \sigma_t^2 \\ &\approx \widehat{S_t^{\epsilon 2}}\sigma_t^{*2} \\ &= \exp\{\widehat{\ln(S_t^{\epsilon 2})}\}\frac{\sum_{t=1}^{T}\exp\{\widehat{u_t}\}}{T}\widehat{\sigma_t^{*2}}. \end{aligned} \tag{9}$$

Short-horizon stock return predictability

Stock return predictability is often discussed in conjunction with market efficiency, a concept introduced by Fama 1970 and further developed by scholars such as Jensen 1978, Black 1986, Malkiel 1992, and Timmermann and Granger 2004. According to Fama 1991, the market efficiency hypothesis posits that "security prices reflect all available information". This hypothesis is frequently linked to the idea that asset prices follow a random walk model, suggesting that the price of an asset in the next period equals its price in the previous period plus an independently and identically distributed mean-zero variable (Malkiel 2003).

Malkiel, a prominent advocate of the efficient market hypothesis, has been a strong proponent of passive investing. However, Malkiel 2003 acknowledged a shift in the intellectual landscape, noting that "by the start of the twenty-first century, the intellectual dominance of the efficient market hypothesis had become far less universal". He highlighted the significance of "psychological and behavioral elements" as well as certain "fundamental valuation metrics" in effectively predicting future stock prices.

When the analysis shifts to high-frequency financial data, market microstructure effects become significantly important. For instance, the well-known Roll 1984 model incorporates transaction costs through the bid-ask spread, impacting asset price variation. A key finding of Roll's model is that the covariance between successive tick-by-tick price changes is negative due to the bid-ask bounce effect, rather than the arrival of new information. Furthermore, the magnitude of this covariance is directly related to the size of the effective bid-ask spread.

Huang and Stoll 1994 examined numerous microstructure theories to understand short-term stock return behavior, focusing on 5-minute stock log-returns. Their research demonstrated that using data observed at five-minute intervals effectively captures past return behavior without requiring a complex lag structure. The study modeled several microstructure effects: the adverse information effect, inventory effect, and order processing cost.

The adverse information effect, initially described by Copeland and Galai 1983 and Glosten and Milgrom 1985b, represents the value of private information conveyed by public sellers or buyers. Since dealers take proprietary positions in traded stocks, they bear inventory risk. The inventory holding cost, as first by Stoll 1978 and Ho and Stoll 1981, compensates dealers for holding unwanted inventory. Order processing costs cover expenses related to trade processing, such as communication, clearing, and record keeping. Both adverse information and inventory effects tend to lower bid and ask prices.

An additional effect formulated by Huang and Stoll 1994 is the induced order arrival effect, which corrects the assumption in many microstructure models that the probabilities of public purchase and sale are equal. Key variables identified by the researchers include the logarithm of the average bid and ask prices quoted just before a trade and the logarithm of the transaction price. They noted that the transaction price typically deviates from the average bid and ask prices by about half the effective bid-ask spread, as transactions often occur within the quoted spread.

The study concluded that the lagged difference between the logarithm of the transaction price and the logarithm of the average bid and ask prices just before a trade is crucial in multivariate prediction models of 5-minute stock log-returns. Other variables used in their models include lagged changes in the S&P 500 index futures prices, lagged quote returns, lagged cumulative signed volume, lagged indicators for large public sales and buys, and the lagged difference between quoted volumes at ask and bid prices.

Huang and Stoll 1994 demonstrated that the coefficient of the lagged difference between the logarithm of the transaction price and the average bid-ask prices is statistically significant in predicting price log-returns. This was shown for 20 actively traded stocks in the Major Market Index over all trading days in 1988, showing a negative coefficient, which indicates a reduced bid-ask bounce in price returns. The findings align with a mix of order-processing, adverse-information, and inventory-holding theories.

The authors argue that predicting stock returns based on microstructure variables does not necessarily contradict market efficiency. Microstructure theory suggests that prices adjust to past prices and trades to incorporate private information, manage inventory, and cover operating costs.

Retail investor activity

Retail investors tend to approach buying and selling stocks differently. The decision to buy involves higher search costs compared to selling, which is typically based on the range of stocks they already own. Generally, retail investors do not engage in short selling (Barber and Odean 2007). Their investment decisions are significantly impacted by bounded rationality, limited attention resources, processing capabilities, and overconfidence, which are intrinsic human traits (Barber and Odean 2007; Odean 1998; Seasholes and Wu 2007). Recognizing these factors necessitates moving away from the notion of full investor rationality. Often, the problem of finding stocks to purchase is reduced to choosing among attention-getting stocks (Barber and Odean 2007; Odean 1998; Seasholes and Wu 2007). According to Barber and Odean 2007, attention-grabbing stocks are usually actively discussed in the news, exhibit extreme returns, or show unusual trading volumes. Significant company-related news can lead to divergent investor views, increasing trading activity. For example, a dramatic price surge of GME common stock was triggered by news of Ryan Cohen and his colleagues joining GameStop's board, highlighted by online investment communities. Empirical evidence shows that individual investors are net buyers of attention-grabbing stocks and tend to purchase stocks they haven't owned previously following such events (Seasholes and Wu 2007). In contrast, professional investors do not exhibit this attention-induced buying behavior (Barber and Odean 2007). Individual non-professional investors often trade suboptimally, resulting in lower returns due to excessive trading (Odean 1999; Barber and Odean 2000; Barber and Odean 2001; Barber and Odean 2002). Sophisticated investors can profitably exploit the predictable behavior of non-professional investors towards attention-grabbing stocks (Seasholes and Wu 2007).

Recent studies suggest that retail investor activity can predict stock returns. It has been documented that aggregate retail buy–sell imbalances are contrarian and positively predict stock returns over a three-week horizon (Barrot and Sauvagnat 2016). Additionally, individual investors' trades can be uninformative at intraday and long horizons but informative at short horizons (5-20 days) (Fong et al. 2014). Boehmer, Charles Jones, et al. 2021 also find that marketable retail order flow is useful in predicting stock returns.

However, retail order flow is hard to obtain from public resources, with research typically relying on proprietary data sets. Boehmer, Charles Jones, et al. 2021 suggest that retail order flow was previously proxied by trade size, but this approach became obsolete as market structures became more algorithm-driven. They propose a new method to isolate retail order flow from aggregated order flow.

Due to regulatory restrictions in the United States, retail order flow can receive price improvements, unlike institutional order flow (U.S. Securities and Exchange Commission 2019). This price improvement means executing orders at better prices than current quotes, benefiting retail investors. The SEC initiated a retail price improvement program in 2012, which became permanent in 2019. The minimum size of retail price improvement is $0.001. Marketable orders are treated as retail purchases if the transaction price is slightly below a round penny and as retail sells if slightly above (Boehmer, Charles Jones, et al. 2021).

Marketable retail equity orders in the United States typically share another common characteristic: they are usually executed off-exchange. This can happen either through internalization by brokers or by selling to wholesalers (Boehmer, Charles Jones, et al. 2021). The wholesalers' market is competitive, encouraging them to offer price improvements to attract order flow from retail brokerages. Researchers suggest that internalizers are similarly motivated to provide price improvements to comply with regulatory requirements and ensure optimal order execution.

The majority of marketable retail equity orders are nondirected, meaning that the broker is not instructed to execute an order at a specific exchange venue (Boehmer, Charles Jones, et al. 2021). Institutional orders, on the contrary, are rarely internalized or sold to wholesalers; they are sent directly to exchanges (Boehmer, Charles Jones, et al. 2021). The minimum pricing increments for stocks priced greater than $1 per share is $0.01, as required by the National Market System (U.S. Securities and Exchange Commission 2024). Therefore, institutional orders are usually priced in round pennies, except for occasions when the transaction is executed at a midpoint price between the best bid and best offer and the quoted spread is 1 cent (Boehmer, Charles Jones, et al. 2021).

Boehmer, Charles Jones, et al. 2021 identify the following rule for the identification of retail purchases and sales. Let P_{it} be the transaction price of stock i at time t and $Z_{it} \equiv 100 * \text{mod}(P_{it}, 0.01)$. If $Z_{it} \in (0, 0.4)$, then the trade is identified as a retail sell transaction. If $Z_{it} \in (0.6, 1)$, then the trade is identified as a retail buy transaction.

Hu et al. 2021 find that the positive predictive power of retail order flow on next-day stock returns remains robust even when social media variables are included in the regression analysis. This indicates the significant influence of retail investor activity on short-term stock returns.

Short selling activity

There are at least two distinct motivations for shorting: one is based on the fundamental analysis of the stock, leading to the belief that the stock is overvalued; the other is hedging, such as when market-makers trading options engage in shorting to manage risk (Boehmer, Charles M. Jones, et al. 2008).

The events surrounding GameStop in late 2020 and early 2021 revealed unique dynamics in the stock market, particularly related to short selling. During this period, GME short interest exceeded 100%, making it a rare case where the number of shares sold short was greater than the shares outstanding (U.S. Securities and Exchange Commission 2021). Notably, a broad retreat of short sellers began approximately two months before the GameStop events, indicating an early shift in the market environment that prompted significant losses for short positions across various highly shorted stocks (Gârleanu et al. 2021).

Gârleanu et al. 2021 suggest that these shorting strategies experienced historically poor returns from November 2020 to January 2021, leading to a significant decline in short interest during these months. This decline did not revert to previous levels in the following six months, implying a shift to a new equilibrium with less shorting activity. This retreat was likely driven by fear of an evolving market environment rather than changes in stock fundamentals or lending market structure.

In their model, Gârleanu et al. 2021 incorporate both rational and overly optimistic (irrational) investors, illustrating how shorting can be fickle and prone to sudden reversals. The market clearing condition for share lending is defined as the equivalence between aggregate short interest (demanded short shares) and a fraction of the lent-out aggregate long position (supply of lendable shares). The lending fee, which is determined through negotiations between brokers and security lenders, is a key component of this dynamic. This fee depends positively on the fraction of lent-out stock, meaning that as brokers lend out a higher fraction of their stock, the lending fee increases.

Gârleanu et al. 2021 illustrate that this relationship creates a feedback loop. As more stock is lent out, the effective rate of return for overly optimistic investors holding long positions increases primarily because they can earn additional income through lending fees. When the demand for borrowing shares rises, the competition among short sellers to obtain shares intensifies, driving up the lending fees. Stockholders who lend their shares earn these higher lending fees, which become a significant component of their total return.

Consequently, the overall return for these long investors is augmented by the lending income. This makes holding the stock more attractive, especially for overly optimistic investors, who continue to benefit from the heightened lending fees. This additional income incentivizes them to hold onto their shares rather than selling them, reinforcing the cycle of increased lending and higher returns.

As this cycle continues, the demand for shares drives up their price, and the stock's Sharpe ratio consequently lowers. A lower Sharpe ratio indicates a decrease in the risk-adjusted return, making the stock appear less attractive to rational investors. However, the growing lending income keeps the overly optimistic investors engaged. As the Sharpe ratio declines, the appetite for short-selling grows because short sellers anticipate that the stock is overvalued, which further increases the demand for lent shares and perpetuates the cycle.

The model presented by Gârleanu et al. 2021 suggests that the market can reach multiple equilibria due to this self-reinforcing effect. For example, if rational short sellers are motivated to exit their positions—regardless of whether the trigger is related to stock fundamentals or changes in the stock lending market—the lending income will decrease. With reduced lending income, the incentive for stockholders to lend their shares diminishes. This results in a lower supply of lendable shares and further decreases the short interest. The reduction in shorting activity alleviates downward pressure on the stock price, which can lead to an increase in the stock's Sharpe ratio. This shift results in a new equilibrium characterized by a higher Sharpe ratio and reduced shorting activity. As a result, there appears to be a negative relationship between the cfraction of lent-out stock and the stock's Sharpe ratio. Thus, the market can transition to a new state where shorting is less prevalent, demonstrating the potential for sudden and significant changes in market dynamics.

This theoretical framework aligns with empirical observations, showing that short interest and lending fees negatively correlate with expected excess returns. Understanding these mechanisms helps explain the retreat of short sellers observed in the GameStop events and underscores the broader implications for market equilibrium and investor behavior.

Boehmer, Charles M. Jones, et al. 2008 highlight that short sellers are generally sophisticated and well-informed market participants, engaging in shorting based on fundamental analysis or hedging strategies. They find that larger short sale orders are more informative about future price movements, suggesting that trade size correlates with the quality of information possessed by informed traders. Charles Jones et al. 1994 support this by documenting a positive relationship between trading volume and volatility, indicating that informed traders tend to trade larger amounts at given prices.

To measure shorting flows, Boehmer, Charles M. Jones, et al. 2008 use three approaches:

1. The number of executed short sale orders,
2. The total number of shares sold short, and
3. The cfraction of the total number of shares sold short relative to the total trading volume in the stock.

Boehmer, Charles M. Jones, et al. 2008 partition short sale orders by their size and find that orders of at least 5,000 shares are the most informative about future price moves.

Hu et al. 2021 add another dimension by showing that social media activity can deter short sellers, potentially due to the risk of a short squeeze, as seen in the GameStop case. Their research indicates that social media activity negatively predicts shorting flows, and this effect remains robust even when controlling for shorting flow variables in stock return predictions. Social media traffic exacerbates the negative impact of shorting flows on stock returns, highlighting the influence of online platforms on market dynamics.

Variables

This document provides descriptions of the variables used in the analysis. The variables are grouped into relevant categories for clarity. Each variable is considered for asset i at time period t.

No.	Variable	Description	Data source
1	log_return_{it}	Last trade price 5-minute log-return	tickdatamarket.com
2	$trade\_vlm_{it}$	Total trading volume over a 5-minute interval	Bloomberg database
3	$trade\_num_{it}$	Number of trades over a 5-minute interval	Bloomberg database
Reddit Sentiment and Activity			
4	$mentions_{it}$	GME: number of mentions of "gme" or "gamestop" in comments on Reddit Wallstreetbets over a 5-minute interval during regular trading hours over 09:30 – 16:00 EST.	Baumgartner et al. 2020
5	$ovrnght\_mentions_{it}$	GME: number of mentions of "gme" or "gamestop" in comments on Reddit Wallstreetbets over non-regular trading hours and nontrading days.	Baumgartner et al. 2020
6	$neg\_word\_cnt_{it}$	Number of negative words in comments posted over a 5-minute interval	Baumgartner et al. 2020
7	$pos\_word\_cnt_{it}$	Number of positive words in comments posted over a 5-minute interval	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
8	$total\_word\_cnt_{it}$	Total number of words in comments posted over a 5-minute interval	Baumgartner et al. 2020
9	$neg\_word\_fr_{it}$	Negative word frequency in comments posted over a 5-minute interval, $neg\_word\_fr_{it} = \frac{neg\_word\_cnt_{it}}{total\_word\_cnt_{it}}$	Baumgartner et al. 2020
10	$pos\_word\_fr_{it}$	Positive word frequency in comments posted over a 5-minute interval, $pos\_word\_fr_{it} = \frac{pos\_word\_cnt_{it}}{total\_word\_cnt_{it}}$	Baumgartner et al. 2020
11	$tone_{it}$	Tone of comments posted over a 5-minute interval, $tone_{it} = neg\_word\_fr_{it} - pos\_word\_fr_{it}$	Baumgartner et al. 2020
12	neg_dummy_{it}	Indicator of whether comments posted over a 5-minute interval had greater negative word frequency than an average negative word frequency of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $neg\_dummy_{it} = \begin{cases} 1, & \text{if } neg\_word\_fr_{it} > \frac{1}{T}\sum_{t=1}^{T} neg\_word\_fr_{it} \\ 0, & \text{otherwise,} \end{cases}$ where T is a number of 5-minute intervals	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
13	$non\_neg\_dummy_{it}$	Indicator of whether comments posted over a 5-minute interval had lower negative word frequency than an average negative word frequency of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $non\_neg\_dummy_{it} = 1 - neg\_dummy_{it}$	Baumgartner et al. 2020
14	$tone\_pos\_dummy_{it}$	Indicator of whether comments posted over a 5-minute interval had greater value of tone than an average tone of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $tone\_pos\_dummy_{it} = \begin{cases} 1, & \text{if } tone_{it} > \frac{1}{T}\sum_{t=1}^{T} tone_{it} \\ 0, & \text{otherwise,} \end{cases}$ where T is a number of 5-minute intervals	Baumgartner et al. 2020
15	$tone\_non\_pos\_dummy_{it}$	Indicator of whether comments posted over a 5- minute interval had lower value of tone than an average tone of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $tone\_non\_pos\_dummy_{it} = 1 - tone\_pos\_dummy_{it}$	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
16	$score_{it}$	Average score of comments posted over a 5-minute interval	Baumgartner et al. 2020
17	$score\_dummy_{it}$	Indicator of whether an average score of comments posted over a 5-minute interval was higher than an average score of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $score\_dummy_{it} = \begin{cases} 1, & \text{if } score_{it} > \frac{1}{T}\sum_{t=1}^{T} score_{it} \\ 0, & \text{otherwise,} \end{cases}$ where T is a number of 5-minute intervals	Baumgartner et al. 2020
18	$popular\_neg\_dummy_{it}$	Indicator of whether an average score of comments and negative word frequency of comments posted over a 5-minute interval was higher than an average score and average negative word frequency of all comments posted during trading hours over January 11, 2021 – February 4, 2021, respectively, $popular\_neg\_dummy_{it} = neg\_dummy_{it} * score\_dummy_{it}$	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
19	$popular\_non\_neg\_dummy_{it}$	Indicator of whether an average score of comments posted over a 5-minute interval was higher than an average score and negative word frequency of comments posted over a 5-minute interval was lower than average negative word frequency of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $popular\_non\_neg\_dummy_{it} = non\_neg\_dummy_{it} * score\_dummy_{it}$	Baumgartner et al. 2020
20	$popular\_tone\_pos\_dummy_{it}$	Indicator of whether an average score of comments and tone of comments posted over a 5-minute interval was higher than an average score and average tone of all comments posted during trading hours over January 11, 2021 – February 4, 2021, respectively, $popular\_neg\_dummy_{it} = tone\_pos\_dummy_{it} * score\_dummy_{it}$	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
21	$popular\_tone\_non\_pos\_dummy_{it}$	Indicator of whether an average score of comments posted over a 5-minute interval was higher than an average score and tone of comments over a 5-minute interval was lower than average tone of all comments posted during trading hours over January 11, 2021 – February 4, 2021, $popular\_tone\_non\_pos\_dummy_{it} = tone\_non\_pos\_dummy_{it} * score\_dummy_{it}$	Baumgartner et al. 2020
22	$ovrnght\_total\_fr\_neg_{it}$	cfraction of comments that have a higher score than average score of all comments posted overnight and higher negative word frequency than the average negative word frequency of all comments posted overnight out of total number of comments that were posted overnight	Baumgartner et al. 2020
23	$ovrnght\_total\_fr\_non\_neg_{it}$	cfraction of comments that have a higher score than average score of all comments posted overnight and lower negative word frequency than the average negative word frequency of all comments posted overnight out of total number of comments that were posted overnight	Baumgartner et al. 2020

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
24	$ovrnght\_total\_fr\_tone\_pos_{it}$	cfraction of comments that have a higher score than average score of all comments posted overnight and greater value of tone than the average tone of all comments posted overnight	Baumgartner et al. 2020
25	$ovrnght\_tone\_non\_pos_{it}$	cfraction of comments that have a higher score than average score of all comments posted overnight and smaller value of tone than the average tone of all comments posted overnight out of total number of comments that were posted overnight	Baumgartner et al. 2020
Option markets activity-related variables			
26	$call\_itm\_opn\_int_{it}$	In the money call option open interest at a 5-minute frequency	cboe.com
27	$call\_atm\_opn\_int_{it}$	At the money call option open interest at a 5-minute frequency	cboe.com
28	$call\_otm\_opn\_int_{it}$	Out of the money call option open interest at a 5-minute frequency	cboe.com
29	$put\_itm\_opn\_int_{it}$	In the money put option open interest at a 5-minute frequency	cboe.com
30	$put\_atm\_opn\_int_{it}$	At the money put option open interest at a 5-minute frequency	cboe.com

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
31	$put\_otm\_opn\_int_{it}$	Out of the money put option open interest at a 5-minute frequency	cboe.com
		Short selling activity-related variables	
32	$short\_vlm_{it}$	Total short volume over a 5-minute interval	cboe.com
33	ss_{it}	Total shares sold short in reported by CBOE as a cfraction of total trading volume reported by Bloomberg database over a 5-minute interval, $ss_{it} = \frac{short\_vlm_t}{trade\_vlm_t}$	cboe.com
34	$short\_trd\_num_{it}$	Number of short trades over a 5-minute interval	cboe.com
		Retail Trading Activity	
35	$mrbuytrd_{it}$	Total retail buy trades count over a t^{th} 5-minute interval	tickdatamarket.com
36	$mrselltrd_{it}$	Total sell buy trades count over a 5-minute interval	tickdatamarket.com
37	$mrbuyvol_{it}$	Total retail buy volume over a 5-minute interval	tickdatamarket.com
38	$mrsellvol_{it}$	Total sell buy volume over a 5-minute interval	tickdatamarket.com
39	$mrtrd_{it}$	Total retail trades count over a 5-minute interval	tickdatamarket.com

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
40	$mroibvol_{it}$	Marketable volume retail order imbalance over a 5-minute interval, $mroibvol_{it} = \frac{mrbuyvol_{it} - mrsellvol_{it}}{mrbuyvol_{it} + mrsellvol_{it}}$	tickdatamarket.com
41	$mroibtrd_{it}$	Marketable trade retail order imbalance over a 5-minute interval, $mroibtrd_{it} = \frac{mrbuytrd_{it} - mrselltrd_{it}}{mrbuytrd_{it} + mrselltrd_{it}}$	tickdatamarket.com
		Conditional Volatility	
42	$garch\_vol_{it}$	True volatility estimated through an ARMA-GARCH model using LTP log-returns	tickdatamarket.com
43	$deseason\_garch\_vol_{it}$	Deseasonalized volatility estimated through an ARMA-GARCH model using LTP log-returns	tickdatamarket.com
		Trade Restrictions and Influences	
44	$trd\_restriction\_dummy_{it}$	Indicates whether there has been a trading restriction over a 5-minute interval, $trd\_restriction\_dummy_{it} = \begin{cases} 1, & \text{if the period is over January 27 - February 4} \\ 0, & \text{otherwise,} \end{cases}$	tickdatamarket.com

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
45	$musk\_tweet_{it}$	Time-decaying indicator illustrating a decaying influence of Elon Musk's tweet mentioning GME on financial markets, $elon\_musk\_tweet_{it} = \begin{cases} \frac{847}{t}, & t \geq 847 \\ 0, & \text{otherwise}, \end{cases}$	tickdatamarket.com
Control Variables			
46	mm_{it}	Lagged difference between a logarithm of the transaction price and a logarithm of the average of bid and ask prices quoted just prior to a trade, $mm_{it+1} = \ln(P_{it}) - \ln\left(\frac{Ask_{it} + Bid_{it}}{2}\right)$	tickdatamarket.com
47	$price\_dsp_{it}$	The difference between the highest and the lowest price over a 5-minute interval, $price\_dsp_{it} = \frac{High_{it} - Low_{it}}{\frac{High_{it} + Low_{it}}{2}}$	Bloomberg database
48	$ovrnght\_return_{it}$	Overnight LTP log return based on the last trade price at 16:00 EST on the previous trading day and last trade price at 09:35 EST on the current trading day	tickdatamarket.com

Continued on next page

Table 3 – continued from previous page

No.	Variable	Description	Source
49	$lag\_daily\_log\_return_{it}$	Lagged daily GME or BB log-return	https://uk.finance.yahoo.com/
50	$lag\_daily\_trd\_vlm_{it}$	Lagged daily GME or BB trading volume	https://uk.finance.yahoo.com/
51	$lag\_daily\_vix_{it}$	Lagged daily VIX	https://uk.finance.yahoo.com/
Other variables			
52	$last\_trd\_price_{it}$	Last trade price at a 5-minute frequency	tickdatamarket.com
53	$close\_price_{it}$	Close price at a 5-minute frequency	Bloomberg database

Bibliography

Allen, Franklin et al. (2021). "Squeezing Shorts Through Social News Platforms". In: *SSRN Electronic Journal*.

Authers, John (2021). *GameStop Is Rage Against the Financial Machine*. URL: https://www.bloomberg.com/opinion/articles/2021-01-27/gamestop-short-squeeze-is-rage-against-the-financial-machine.

Baillie, Richard T. and Bollerslev, Tim (1989). "The Message in Daily Exchange Rates: A Conditional-Variance Tale". In: *Journal of Business Economic Statistics* 7.3, pp. 297–305. URL: http://www.jstor.org/stable/1391527.

Barber, Brad M. and Odean, Terrance (2000). "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors". In: *Journal of Finance* 55, pp. 773–806.

— (2001). "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment". In: *Quarterly Journal of Economics* 116, pp. 261–292.

— (2002). "Online Investors: Do the Slow Die First?" In: *The Review of Financial Studies* 15.2, pp. 455–487.

— (Dec. 2007). "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors". In: *The Review of Financial Studies* 21.2, pp. 785–818. URL: https://doi.org/10.1093/rfs/hhm079.

Barrot, Jean-Noël and Sauvagnat, Julien (May 2016). " Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks *". In: *The Quarterly Journal of Economics* 131.3, pp. 1543–1592. URL: https://doi.org/10.1093/qje/qjw018.

Baumgartner, Jason et al. (2020). "The Pushshift Reddit Dataset". In: URL: https://arxiv.org/pdf/2001.08435.pdf.

Beber, Alessandro and Pagano, Marco (2013). "Short-Selling Bans Around the World: Evidence from the 2007-09 Crisis". In: *Journal of Finance* 68.1. URL: https://www.eief.it/files/2013/01/pagano_beber_joff_2013.pdf.

Black, Fischer (1986). “Noise”. In: *The Journal of Finance* 41.3, pp. 528–543. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04513.x.

Boehmer, Ekkehart, Jones, Charles, Zhang, Xiaoyan, and Zhang, Xinran (2021). “Tracking Retail Investor Activity”. In: *Journal of Finance* 76.5, pp. 2249–2305. URL: https://EconPapers.repec.org/RePEc:bla:jfinan:v:76:y:2021:i:5:p:2249-2305.

Boehmer, Ekkehart, Jones, Charles M., and Zhang, Xiaoyan (2008). “Which Shorts Are Informed?” In: *The Journal of Finance* 63.2, pp. 491–527. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2008.01324.x.

Chen, Hailiang, De, Prabuddha, Hu, Yu (Jeffrey), and Hwang, Byoung-Hyoun (2014). “Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media”. In: *Review of Financial Studies* 27.5, pp. 1367–1403. URL: https://EconPapers.repec.org/RePEc:oup:rfinst:v:27:y:2014:i:5:p:1367-1403..

CNBC (2021). *GameStop jumps as Elon Musk tweets out Reddit board that’s hyping stock*. URL: https://www.cnbc.com/2021/01/26/gamestop-jumps-as-elon-musk-tweets-out-reddit-board-thats-hyping-stock.html.

Cohen, Ryan (Nov. 2020). *Letter to The Board of Directors of GameStop Corporation*. https://www.sec.gov/Archives/edgar/data/1326380/000101359420000821/rc13da3-111620.pdf.

Copeland, Thomas E. and Galai, Dan (1983). “Information Effects on the Bid-Ask Spread”. In: *The Journal of Finance* 38.5, pp. 1457–1469. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1983.tb03834.x.

Dacorogna, Michel M. et al. (2001). “5 - BASIC STYLIZED FACTS”. In: *An Introduction to High-Frequency Finance*. Academic Press, pp. 121–173. URL: https://www.sciencedirect.com/science/article/pii/B9780122796715500083.

Diebold, F.X. (2006). *Elements of Forecasting*. South-Western College Publishing.

Fama, Eugene F. (May 1970). “Efficient Capital Markets: A Review of Theory and Empirical Work”. In: *Journal of Finance* 25.2, pp. 383–417. URL: https://ideas.repec.org/a/bla/jfinan/v25y1970i2p383-417.html.

— (Dec. 1991). “Efficient Capital Markets: II”. In: *Journal of Finance* 46.5, pp. 1575–1617.

Fong, Kingsley Y. L., Gallagher, David R., and Lee, Adrian D. (2014). “Individual Investors and Broker Types”. In: *The Journal of Financial and Quantitative Analysis* 49.2, pp. 431–451. URL: http://www.jstor.org/stable/43303952.

Gaffney, Devin and Matias, J. Nathan (2018). “Caveat Emptor, Computational Social Science: Large-Scale Missing Data in a Widely-Published Reddit Corpus”. In: *PLOS ONE* 13.7. DOI: 10.1371/journal.pone.0200162. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200162.

GameStop Corp. (Jan. 2021a). *GameStop Announces Additional Board Refreshment to Accelerate Transformation*. GameStop News Release. URL: https://news.gamestop.com/news-releases/news-release-details/gamestop-announces-additional-board-refreshment-accelerate.

— (Mar. 2021b). *GameStop Corp., 2020 Annual Report*. https://news.gamestop.com/static-files/470c5a4c-bb4f-48d4-abec-befd467d3210.

Gârleanu, Nicolae, Panageas, Stavros, and Zheng, Geoffery (2021). “A Long and a Short Leg Make For a Wobbly Equilibrium”. In: *DecisionSciRN: Stock Market Decision-Making (Sub-Topic)*. URL: https://api.semanticscholar.org/CorpusID:236344656.

Glosten, Lawrence R. and Milgrom, Paul R. (1985a). “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders”. In: *Journal of Financial Economics* 14.1, pp. 71–100. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(85)90044-3. URL: https://www.sciencedirect.com/science/article/pii/0304405X85900443.

— (1985b). “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders”. In: *Journal of Financial Economics* 14.1, pp. 71–100. URL: https://www.sciencedirect.com/science/article/pii/0304405X85900443.

Ho, Thomas and Stoll, Hans R. (1981). “Optimal dealer pricing under transactions and return uncertainty”. In: *Journal of Financial Economics* 9.1, pp. 47–73. URL: https://www.sciencedirect.com/science/article/pii/0304405X81900209.

Hoffstein, Corey (2021). *Need to Understand the Influence of Social Media on Trades*. URL: https://www.bloomberg.com/news/videos/2021-01-22/need-to-understand-the-influence-of-social-media-on-trades-corey-hoffstein-video.

Hu, Danqi, Jones, Charles M., Zhang, Valerie, and Zhang, Xiaoyan (2021). “The Rise of Reddit: How Social Media Affects Retail Investors and Short-sellers’ Roles in Price Discovery”. In: *Behavioral & Experimental Finance eJournal*. URL: https://api.semanticscholar.org/CorpusID:233652370.

Huang, Roger D. and Stoll, Hans R. (Spring 1994). “Market Microstructure and Stock Return Predictions”. In: *The Review of Financial Studies* 7.1, pp. 179–213. URL: http://links.jstor.org/sici?sici=0893-9454%28199421%297%3A1%3C179%3AMMASRP%3E2.0.CO%3B2-Q.

Jensen, Michael C. (1978). “Some anomalous evidence regarding market efficiency”. In: *Journal of Financial Economics* 6.2, pp. 95–101. URL: https://www.sciencedirect.com/science/article/pii/0304405X78900259.

Jensen, Michael C. and Ruback, Richard S. (1983). “The market for corporate control: The scientific evidence”. In: *Journal of Financial Economics* 11.1, pp. 5–50. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(83)90004-1. URL: https://www.sciencedirect.com/science/article/pii/0304405X83900041.

Jones, Charles, Kaul, Gautam, and Lipson, Marc (1994). “Transactions, volume and volatility”. In: *Review of Financial Studies* 7.4, pp. 631–652.

Levine, Matt (Jan. 2021). *Will WallStreetBets Face SEC Scrutiny After GameStop Rally?* Bloomberg Opinion. URL: https://www.bloomberg.com/opinion/articles/2021-01-26/will-wallstreetbets-face-sec-scrutiny-after-gamestop-rally.

Malkiel, Burton G. (1992). "Efficient Market Hypothesis". In: *New Palgrave Dictionary of Money and Finance*. Ed. by Peter Newman, Murray Milgate, and John Eatwell. London: Macmillan.

— (Winter 2003). "The Efficient Market Hypothesis and Its Critics". In: *Journal of Economic Perspectives* 17.1, pp. 59–82.

Musk, Elon (2021). *Gamestonk!!* URL: https://x.com/elonmusk/status/1354174279894642703.

Odean, Terrance (1998). "Are investors reluctant to realize their losses?" In: *The Journal of Finance* 53.5, pp. 1775–1798.

— (Dec. 1999). "Do Investors Trade Too Much?" In: *American Economic Review* 89.5, pp. 1279–1298. URL: https://www.aeaweb.org/articles?id=10.1257/aer.89.5.1279.

Patel, Sahil (2020). *Reddit Claims 52 Million Daily Users, Revealing a Key Figure for Social-Media Platforms*. URL: https://www.wsj.com/articles/reddit-claims-52-million-daily-users-revealing-a-key-figure-for-social-media-platforms-11606822200.

Patton, A. (2015). *Quantitative Finance*. This guide was prepared for the University of London International Programmes by A. Patton, Department of Economics, Duke University. University of London.

Pedersen, Lasse Heje (2022). "Game on: Social networks and markets". In: *Journal of Financial Economics* 146.3, pp. 1097–1119. DOI: https://doi.org/10.1016/j.jfineco.2022.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0304405X22000964.

Pew Research Center (Apr. 2021). *Social Media Use in 2021*. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

Reddit Help (2022). *Reddiquette*. URL: https://support.reddithelp.com/hc/en-us/articles/205926439-Reddiquette.

Reddit Inc. (2021). *Reddit Recap 2021*. URL: https://www.redditinc.com/blog/reddit-recap-2021#:~:text=In%202021%2C%20redditors%20created%20366,%2C%20a%201%25%20increase%20YoY.

Roll, Richard (1984). "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market". In: *The Journal of Finance* 39.4, pp. 1127–1139. DOI: https://doi.org/10.1111/j.1540-6261.1984.tb03897.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1984.tb03897.x.

Salihefendic, Amir (Dec. 2015). *How Reddit Ranking Algorithms Work*. URL: https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9.

Seasholes, Mark S. and Wu, Guojun (Dec. 2007). "Predictable behavior, profits, and attention". In: *Journal of Empirical Finance* 14.5, pp. 590–610.

Securities and Exchange Comission (2020). *Form SC 13D/A - RCI Hospitality Holdings, Inc.* https://www.sec.gov/Archives/edgar/data/1326380/000101359420000670/rc13d-082820.htm. Filing Date: 2020-08-28.

Securities Exchange Act of 1934 (1934). Public Law 73-291, 48 Stat. 881. URL: https://www.govinfo.gov/content/pkg/COMPS-1885/pdf/COMPS-1885.pdf.

SimilarWeb (2020). *Reddit.com Analytics*. URL: https://www.similarweb.com/website/reddit.com/.

Stoll, Hans R. (1978). "The Supply of Dealer Services in Securities Markets". In: *The Journal of Finance* 33.4, pp. 1133–1151. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2326945.

Timmermann, Allan and Granger, Clive (2004). "Efficient market hypothesis and forecasting". In: *International Journal of Forecasting* 20.1, pp. 15–27. URL: https://EconPapers.repec.org/RePEc:eee:intfor:v:20:y:2004:i:1:p:15-27.

Tsay, Ruey S. (2002). *Analysis of financial time series. financial econometrics*. Wiley.

U.S. Securities and Exchange Commission (2012). *National Examination Risk Alert: Investment Adviser Use of Social Media*. URL: https://www.sec.gov/about/offices/ocie/riskalert-socialmedia.pdf.

— (2019). *Self-Regulatory Organizations; New York Stock Exchange LLC; Order Granting Approval of a Proposed Rule Change Amending the Listed Company Manual to Modify the Provisions Relating to Direct Listings*. Release No. 34-85160. URL: https://www.sec.gov/files/rules/sro/nyse/2019/34-85160.pdf.

— (Oct. 2021). *Staff Report on Equity and Options Market Structure Conditions in Early 2021*. https://www.sec.gov/files/staff-report-equity-options-market-struction-conditions-early-2021.pdf.

— (2024). *Sub-Penny Rule (Rule 612)*. URL: https://www.sec.gov/divisions/marketreg/subpenny612faq.htm.

s