To the Moon or Bust: Do Retail Investors Profit From Social Media-Induced Trading?

Liran Eliner*

Botir Kobilov

University of Texas at Dallas

December 2024

Abstract

By combining datasets on retail investors' trading history and stock-specific social media activity, we provide evidence that social media induces retail trading more than other known attention-grabbing factors and is detrimental to investor performance. Specifically, we find retail investors underperform, on average by 2.2%, at both the transaction and portfolio levels from trades placed on days when a stock has abnormally high levels of discussions on social media. We attribute the underperformance to market timing and the disposition effect. These findings are crucial in the context of heightened discourse on the impact of social media on financial markets.

*The paper has benefited from the comments of participants and discussants during presentations made at the Financial Industry Regulatory Authority (2022), Paris Financial Management Conference (2022), HBS Faculty Brownbag (2022), HBS Faculty Symposium Poster Session (2022), Transatlantic Doctoral Conference at LBS (2023), Economics of Financial Technology Conference (2023), the American Accounting Association Annual Meeting 2023 and Helsinki Finance Summit 2023 (poster). We gratefully acknowledge the receipt of the 2023 Call for Papers contest winner award by the Brandes Center at the UC San Diego Rady School of Management. We thank James Kardatzke from Quiver Quant for providing the r/WSB data. All errors and omissions are our own. Correspondence: Botir Kobilov. E-mail: `botir.kobilov@utdallas.edu`

1 Introduction

Stock trading has traditionally been a social endeavor, with decisions often shaped by interpersonal relations, peer influence, and collective sentiment (Shiller and Pound, 1989; Duflo and Saez, 2002; Ouimet and Tate, 2020). Technological advancements and the rise of social media have magnified this interpersonal aspect of stock trading, transforming how information is generated, disseminated, and absorbed (Miller and Skinner, 2015). Retail traders are progressively turning to social media for insights, often sidelining more conventional data sources like financial advisors, newspapers, TV programs, and subscription-based data platforms. Stock picking and day trading videos and streams on YouTube, Twitch, TikTok, Instagram, and text-based messaging boards and groups on Reddit, Discord, Facebook, and Twitter (renamed "X" in July of 2023) have gained significant traction, informing and facilitating the trading activities of many retail traders.[1] With these developments in mind, in this paper we study whether retail investors profit from such social-media induced stock trading.

Despite abundant evidence on the effects of social media on financial markets, a conspicuous gap remains in directly linking social media activity to the trading behavior and performance of individual retail investors. On the one hand, several academic studies have pointed to the positive effects of social media in the investment domain. Blankespoor et al. (2014) has shown that the use of social media by firms can increase the dissemination of information and reduce information asymmetry. Chen et al. (2014) has found that peer-based advice on social media platforms can be useful and value-relevant (Bartov et al., 2018).

On the other hand, however, social media also presents risks and challenges. Discussions on social media platforms can lead to inefficient information processing (Bradley et al., 2021), and they amplify behavioral biases, such as the disposition effect and persuasion bias

[1] FINRA Foundation report found that 60% of retail investors below 35 age use social media as a source of investment information. By contrast, only 35% of investors aged 35 to 54 and just 8% of those 55 and older do the same. See: https://www.finra.org/investors/insights/following-crowd-investing-and-social-media

(Chang et al., 2016a; Heimer, 2016; DeMarzo et al., 2003). Social media can also foster uninformed trading by spreading false information (i.e., "fake news"), rumors, erroneous beliefs, and naive trading strategies, leading to "pump and dump" strategies and trading frenzies (Pedersen, 2022). Growing retail investors' reliance on social media for trading also generates incentives for disseminating false and misleading information for the purpose of price manipulation (Farrell et al., 2022). In the aftermath of the meme-stock frenzy of 2021-2022, regulators have expressed significant concerns about the negative effects that social media may have on market efficiency and on the performance of retail traders. The Securities and Exchange Commission (SEC) advises investors to be cautious and to verify any investment recommendations received through social media. In December 2022, for example, the SEC charged several finance influencers for promoting themselves as successful traders and cultivating millions of followers on such platforms as Twitter and Discord.[2]

Most existing studies of retail investing are based on data sets with relatively small investor base (Barber and Odean, 2000, 2008; Kelley and Tetlock, 2013). This is mostly due to data availability limitations. To address this challenge, Boehmer et al. (2021) created a method to identify marketable retail trades using publicly available U.S. equity transactions data (TAQ). However, Boehmer et al. (2021) method does not allow to observe individual retail traders' performance from those trades, and recently has been found that it identifies only less than one-third of trades generally assumed to be from retail investors (Battalio et. al., 2023).

To investigate the impact of social media on retail trading and the consequences of this relationship on investor performance at both the individual transaction and portfolio levels, we combine a dataset that tracks stock-specific social media activity on one of the most active and popular social media platforms, Reddit's WallStreetBets (r/WSB) with a proprietary dataset on individual retail investor trade logs from a global trading platform *(the Platform)*. We develop a stock-specific social media activity measure that captures abnormal levels of social media discussion on any given day relative to the past 42 trading days. We

[2]See: https://www.sec.gov/news/press-release/2022-221

then split stocks into quintiles based on this measure and classify stock trade as a social media-induced trade if it falls into the top quintile. Our data from the Platform, which serves over 25 million registered users from 140+ countries and covers over 2,500 stocks, allows us to observe individual-level retail trading history, portfolio-level returns, and trader information obtained from Know Your Customer (KYC) answers. Unlike Robinhood, which is more commonly used in previous studies, traders on this Platform can take long and short positions and leverage their trades.

We first conjecture that retail investors are more influenced by social media in making investment decisions, potentially more so than by traditional media sources or other attention-grabbing market events such as momentum or trading volume. Building upon the findings of Barber and Odean (2008), which highlighted the significant impact of attention-getting events on individual investor actions, we analyze the open-close trading imbalance of stocks sorted by trading volume, one-day returns, news coverage, and social media mentions. We find that stocks in the top quintile, sorted by social media activity, are traded almost four times as frequently as those in the bottom quintile. Furthermore, a unit increase in social media activity correlates with a more pronounced shift in retail trading than comparable unit increases in other attention-driven trading factors such as news, momentum, and trading volume. These findings underscore the overpowering influence of social media discussions on retail trading, overshadowing the effects of traditional news, momentum, or trading volumes.

Turning to the main findings of our paper, we find that individual trades placed on days when a stock has an abnormal social media activity have, on average, 1.6-2.8% lower returns than other trades placed on the same day. Next, we find that the share of social media-induced trades in investor portfolios is negatively associated with annualized portfolio returns; specifically, one percentage point increase in the proportion of trades influenced by social media within a trader's portfolio is associated with 1.7-2.8% percentage point decrease in investors' annualized portfolio return.

Consistent with Barber et al. (2023), which reconciled two well-established yet seemingly contradictory empirical findings: retail traders generally underperform, and retail trading imbalance positively predicts short-term returns, we attempt to explain why retail investors engage and underperform from such predictably negative trades. We attribute the underperformance of attention-based trading to poor market timing and the disposition effect, which are exacerbated by social media. Specifically, we find that retail investors predominantly trade on days when stocks exhibit abnormally high social media activity, and their negative performance is concentrated among those trades. Furthermore, we find evidence for the presence of the disposition effect, which is the tendency of retail traders to sell stocks when they are experiencing gains while holding onto stocks when they are incurring losses. This underperformance is compounded by factors such as a short holding period, use of leverage, and frequent trading. Finally, we shed light on heterogeneity among investors, as evidenced by their varying trading behavior and answers to KYC questions. Our findings show that, on average, these investors tend to be younger, male, have less trading experience, have less financial knowledge, exhibit short-term trading strategies, and prefer risk-taking.

Our findings are robust to various measurement choices and regression specifications. We repeat our primary analysis by using the social media activity variable from Refinitiv's MarketPsych Analytics (RMA), which collects data from over 2,000 selected social media sources such as Twitter, StockTwits, Reddit, Investing.com, and other relevant blogs and forums. Our results hold, which underscores that our findings are not driven primarily by data selection. They also remain robust when we employ alternative definitions of our primary variables. Our tight fixed effects specifications also ensure that attention-grabbing factors and retail trading are not confounded by the individual characteristics of investors, specific events on particular days, or inherent characteristics of individual stocks.

The contribution of this paper to the academic literature is twofold. First, our study is closely aligned with research on attention-induced retail trading behavior, which posits that information processing is costly and capacity-constrained investors tend to purchase stocks that have first captured their attention. Such attention-motivated buying results in concentration on stocks that receive investor attention, resulting in poor returns. Barber and Odean (2008) documented three proxies of attention- grabbing events: media coverage, unusual trading volume, and extreme past-day returns. This attention model predicts lower returns, and Barber et al. (2022) recently showed that stocks with the most significant increase in users on the popular Robinhood app tend to have poor returns. Previous studies have also demonstrated stock price reversals following attention-grabbing events, such as Jim Kramer's stock recommendations (Engelberg et al., 2012), the WSJ Dartboard Column (Liang, 1999; Barber and Loeffler, 1993), Google stock searches (Da et al., 2011), and repeat news stories (Tetlock, 2011), which indicate that individual investors overreact to stale information, leading to temporary movements in stock prices. Our study extends existing findings by demonstrating that retail trading, measured by both open-close imbalances and the log number of open trades, increases on days of heightened activity on the social media plat- form. Our study further shows that although social media activity, that is, the level of discussions, positively predicts both aggregate retail trading volume and short-term stock returns, retail traders still lose money when engaging in such trading activity. We attribute this outcome to factors such as market timing and behavioral biases such as the disposition effect.

Second, this paper relates to the literature on herding behavior, which refers to the phenomenon where investors copy each other's actions or make decisions based on the actions of others, influenced by both information and behavior (Shleifer and Summers, 1990; Nofsinger and Sias, 1999; Sias, 2004). Information-driven herding behavior suggests that investors make similar investment choices when they face similar information environments. For instance, Shiller et al. (1984) and De Long et al. (1990) argue that individual investors are susceptible to the influence of fads and fashion. Shleifer and Summers (1990) also contend that retail investors tend to herd when they follow similar signals, such as brokerage house recommendations, popular market influencers, and forecasters. Conversely, behavior-driven herding behavior is linked to psychological biases, such as the representativeness heuristic and disposition effect, and attention-grabbing events (Barber et al., 2009; Merli and Roger, 2014). We contribute to this literature by documenting that discussions on social media platforms can lead to herding behavior by retail investors.

The findings of this study are significant in the context of heightened regulatory scrutiny of retail trading and ongoing discussions about the impact of social media networks on capital markets. This paper makes a contribution by demonstrating that retail investors underperform in social media-driven trading due to late market entry and a prolonged hold on losing positions as prices predictably decrease over time.

The paper proceeds as follows: Section 2 presents the data sources, sample construction, and measurement of key variables. Section 3 presents the main findings of the study. Section 4 investigates the underlying mechanisms. The paper concludes in Section 5.

2 Data and Key Variables

In this section, we outline the data and methodology. We also present descriptive results on social media activity and discuss investor characteristics in the platform.

2.1 Social Media Activity

2.1.1 Reddit

Reddit is a social media network platform that hosts online communities (known as subreddits), focused on specific topics, such as politics, humorous memes, sports teams, or computer games, among numerous others. One such subforum, r/WSB, was created in 2012. It describes itself as "a community for making money and being amused while doing it." It is primarily used for sharing comments on post-investment advice, stock price expectations, individual trades, and speculative trading strategies. r/WSB has profoundly affected several specific stock trading frenzies: The January 2021 GameStop short squeeze was the most prominent. As of December 2023, r/WSB reached 14.6 million subscribers (See Figure 1)

While r/WSB users communicate on several subforums, our focus is on just two of them: "Tomorrow's Moves" and "Daily Discussion." Each trading day, moderators of r/WSB create a dated "Tomorrow's Moves" thread where users discuss what stocks they are trading the next trading day. "Daily Discussion" also dates threads daily for users to discuss the current day's trading session and comment on specific stocks. (see Figure A1) By including a "cashtag" — a dollar sign ($) followed by a stock's ticker symbol — r/WSB users can specify that their comment refers to a specific stock. We obtained our dataset from a third-party provider that scraped all posts and comments from "Tomorrow's Moves" and "Daily Discussion" for the period from January 1, 2019 to September 30, 2021.[3]

Table 1 displays a selection of typical comments, and Table A1 provides a summary of statistics of stock-specific mentions r/WSB for the period between 2019Q1 and 2021Q3. The r/WSB sample includes 4,006,825 comments for 5,822 assets[4]. On average, comments in r/WSB have bullish sentiments, as evidenced by an average sentiment score of 0.064. User activity is concentrated during trading days and hours, as depicted in Figure A2, which shows the distribution of posted comments by day of the week and hour of the day. Table A2 illustrates the top 20 stocks with the highest cumulative number of mentions in RMA and comments in r/WSB samples during the study period.

2.1.2 Measurement

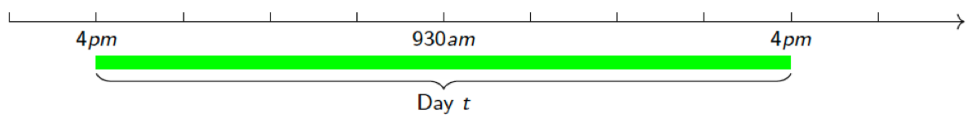
We construct firm-specific social media activity variables by focusing on the abnormal volume of mentions on the day t relative to a benchmark period. Subsequently, we sort stocks into quintiles based on the abnormal daily volume of mentions to accurately identify the stocks with elevated levels of social media discussions that likely attract investor attention.

First, to measure social media activity in r/WSB, we follow Cookson and Niessner (2020) and assign comments on stock s posted between the close of the previous trading day $(t-1)$ the close of the current trading day (t). We also consider comments posted after 4 p.m. on day t, which are assigned to the next day $(t+1)$, as illustrated in the figure below. Thus, the total number of comments for stock s on the day t includes all comments posted between 4

[3]This data is used by the provider for analytics and visualization purposes and is sold on a subscription basis

[4]r/WSB sample covers all assets, including ETFs. In our main analyses throughout the paper, we restrict the r/WSB sample to only stocks, defined as shrcd=10 from CRSP

p.m. on the previous trading day and 4 p.m. on day t. Additionally, we calculate a second version of social media activity by summing up comments posted between 4 PM on day $t-1$ and 9:30 a.m. on day t to capture only overnight social media activity, which is helpful for tests that exploit the timing of comments.



Next, we evaluate the deviation of the daily number of mentions of a stock s in r/WSB from its average number of mentions over the preceding trading days between $t-47$ to $t-6$ (excluding the preceding five trading days). Abnormal social media activity for a stock s on r/WSB sample covers all assets, including exchange-traded funds (ETFs). As noted, throughout the paper, we restrict the r/WSB sample to only stocks, defined as shrcd=10 from CRSP.

$$Abnormal\ Social\ Media\ Activity_{s,t} = \frac{Mentions_{s,t}}{Average\ Mentions_{s,[t-47;t-6]}} \tag{1}$$

Lastly, on the day t, we sort all stocks into quintiles based on their abnormal social media activity. We then define the binary variable $SocialMediaInducedTrade$ as one if the stock's abnormal social media activity falls within the top quintile, and 0 otherwise.

2.2 Retail Trading Activity: *the Platform*

We obtain proprietary data from a multi-asset global online trading platform (the Platform) that allows users to trade individual stocks, exchange-traded funds, stock indices, foreign currencies, commodities, and cryptocurrencies. It has more than 25 million registered users in more than 140 countries, and its users can trade over 2,500 assets, take long or short positions, and use leverage to support their trades. Although the Platform supports trading across 17 exchanges worldwide, we focus exclusively on equity trades conducted at the NYSE and NASDAQ where the Platform recorded 53,726,331 transactions from January 1, 2019 to March 31, 2021.

Our datasets include the Platform's complete trade log, portfolio-level returns, and trader characteristic data with information on each trader's country of residence, gender, and age, as well as answers to KYC questions on prior trading experience, knowledge of trading, preferred trading strategy, primary purpose for trading, attitude toward risk, trading frequency, net annual income, total cash, liquid assets, sources of income, occupation, and more. Table 2 presents a summary of trader characteristics, based on their observed trading history and answers to KYC questions.

We define social-media-induced traders as follows: first, we restrict our sample to only those users who traded at least once on U.S. exchanges during the sample period. Second, we calculate the share of social media-induced trades in the sample period by summing all trades across all asset classes and dividing by the total number of trades. Finally, traders are classified as "Social Media-Induced Traders" if their ratio of social media-induced trades to all trades is in the top 20th percentile among all traders. Table 2 shows that compared to regular retail traders, social media-induced traders are, on average younger, male, have less trading education and experience, and exhibit a preference for short-term returns with high risk-reward ratios.

2.3 Other Datasets

We use several other datasets in this paper, including Refinitiy's *RMA* dataset, which provides us with stock and social media coverage in traditional news outlets, which we use in our robustness tests. We obtain firm-specific fundamentals from *Compustat*, and our asset and index prices are from CRSP. To capture the total retail trading volume in U.S. markets, we use the *Retail Trading Activity Tracker* dataset that tracks individual investor trades according to the classification system used by Boehmer et al. (2021). We define *RetailShare* as a continuous variable that ranges from 0 to 1 and represents the ratio of $USD traded by retail investors in any given ticker divided by the total $USD traded by retail investors across all tickers.

3 Results

This section presents the main results. We first explore the extent to which individuals trade based on social media discussion and how this type of trading compares to other attention-driven trading. We then study the performance of retail traders and explore the potential mechanisms contributing to these results.

3.1 Attention Induced Factors and Retail Trading

Sorting Analysis

We categorize stocks into deciles based on the lagged abnormal trading volume and past-day absolute returns. Then, we determine the number of opened and closed trades for all stocks in each decile for a given day, t. Finally, we calculate the open-close retail trading imbalance for each decile for each date t using the following formula:

$$Open\,Close\,Imbalance = \frac{\sum Open\,Trades - \sum Close\,Trades}{\sum Open\,Trades + \sum Close\,Trades} \tag{2}$$

The results are presented in Table 3. They align with prior literature, demonstrating that individual investors exhibit attention-driven trading behavior on days with high trading volume, extreme one-day returns, and when stocks are in the news. We document the following findings.

i. Columns 1: stocks in the lowest decile (1) by one-day returns receive the highest level of retail investor trading, with an open-close imbalance equal to 6.69. The open-close imbalance decreases up to 2.19 until the ninth decile and increases to 2.96 for stocks with the best past-day returns (decile 10). These are consistent with findings of Barber and Odean (2008).

ii. Column 2: stocks with high abnormal trading volume attract more retail investors trading as the open-close imbalance rises steadily from the lowest decile at 1.34 to the highest decile at 6.74.

iii. Column 3: stocks in the lowest quintile, indicative of the lowest mentions in the news, have an open-close imbalance of 3.58, while stocks in the top quintile have an open-close imbalance of 7.16.

iv. Column 4: stocks in the top quintile have nearly four times more open-close imbalance than stocks in the lowest quintile, 10.9 versus 2.55, respectively.

Regression Analysis

To further examine the relationship between attention-grabbing factors and retail trading, we move beyond the descriptive findings and regress retail trading activity on four attention-inducing factors in both bivariate and multivariate models:

$$RetailTrading_{s,t} = \beta_1 SocialMedia_{s,t} + \beta_2 News_{s,t} + \beta_3 TradingVolume_{s,t} + \beta_4 Return_{s,t} + \epsilon_{s,t} \tag{3}$$

In equation 3, the dependent variable denoted by $RetailTrading_{s,t}$ is represented by either the natural logarithm of the total number of open trades for stock s on the day t or by open-close imbalance value as estimated by equation 2. $SocialMedia_{s,t}$ is an indicator variable equal to one if the abnormal level of discussions on social media for stock s falls in the top 20% of day t, and 0 otherwise. Other independent variables are defined similarly and described in Appendix A. The model includes stock-fixed effects to control for stock-specific and time-invariant unobservables as well as the date-fixed effects to control for unobserved factors that vary across time but are constant across all stocks on a given date.

Results are reported in Table 4. In Panel A, the dependent variable is an open-close imbalance, and in Panel B, the log number of open trades. The first four columns of Panel A of Table 4 regress open-close imbalance on each of the four attention variables separately. Results of columns 1-4 suggest that abnormal levels of social media discussions have a stronger association with changes in the stock open-close imbalance compared to those in news, those with abnormal trading volume, and one-day returns. In Column 5, we regress the open-close imbalance on all four variables of interest. We find that stocks discussed on social media show a more pronounced positive buy-sell imbalance than any other attention-grabbing event. Findings in Panel B of Table 4 reinforce previous observations that stocks with abnormally high volumes of social media discussion tend to have more significant retail trading activity, as measured by the logarithm of the number of open positions.

To address concerns over the external validity of our findings, which are based on individual user trading records on *the Platform*, we conducted a similar analysis using different dependent variables, which captures the total retail trading volume of stocks in U.S. markets. As presented in Table 5, the results suggest a positive association between retail trading volume and stock placement in the top quintile of social media activity on any given day. This association is, in fact, more substantial than the relationship between abnormal news coverage and retail trading. Regardless of the specification, trading volume variable, social media activity measurement sample, or control variable definitions, our analysis consistently shows that social media information significantly impacts retail trading more than other known attention-grabbing factors. To conclude this section, we have documented that retail investors tend to trade stocks that recently gained their attention and that stock-specific social media discussions serve as an additional attention-grabber.

3.2 Social Media Activity and Investor Returns

Our findings in the previous sections indicate that retail investors trade stocks that have recently captured their attention through abnormal social media activity. The next question is, do retail investors profit from such trading activities? First, we construct investor-date-stock transaction-level data from all opened and closed positions during our sample period. We regress trade-level returns on our four attention-grabbing events and other control variables. Second, we analyze the annualized portfolio returns of retail investors through cross-sectional regressions to associate the social media-induced trades from total trades on investor portfolio returns.

3.2.1 Evidence from Trade-Level Returns

To evaluate the performance of trades induced by social media by retail investors versus those that were not, we estimate the following baseline regression specification:

$$\begin{aligned} Return_{i,s,t} = \beta_1 Social\,Media\,Trade_{s,t} + \beta_2 News_{s,t} + \beta_3 Trading\,Volume_{s,t} + \\ + \beta_4 Return_{s,t} + \lambda_t + \mu_s + \gamma_i \quad (4) \end{aligned}$$

The trade-level return, $Return_{i,s,t}$, is measured for stock s placed on the day t by investor i. The $SocialMediaTrade_{s,t}$ variable takes three forms. First, as an indicator variable that is equal to 1 if the abnormal volume of mentions in r/WSB for stock s on the day t falls in the top quintile among all stocks, and 0 otherwise. Second, as the logarithm of the number of mentions. Third, as an indicator variable that captures comments only between after-market hours on the day $t-1$ and pre-market hours on the day t. The model includes the date, stock, and investor fixed effects represented by λ_t, μ_s, and γ_i, respectively.

Table 6 presents the results, and they demonstrate a statistically significant and economically meaningful association between social media-induced trading and trade-level returns. Specifically, trades executed on days with abnormally high social media activity for a given stock result in lower returns than other equity trades. On average, positions opened on stocks with high abnormal coverage in r/WSB forums underperform other stocks by 1.6% to 2.8%. The short holding period, leveraged, and short positions further exacerbate this negative impact on the performance of social media-induced trades.

3.2.2 Evidence from Portfolio Returns

To evaluate the overall portfolio performance of investors who engage on social media induced trading we estimate the following cross-sectional regression:

$$Return_i = \beta_1 Social\,Media\,Trade\,(\%) + \beta_2 Positions\,(log) + \beta_3 Trading\,Months\,(log) + \\ + \beta_4 \overrightarrow{Asset\,Classes\,(\%)} + \beta_5 Leveraged\,Positions\,(\%) + \beta_6 Short\,Positions\,(\%) + \\ + \beta_7 Gender + \beta_8 Age + \beta_9 Trading\,Knowledge \quad (5)$$

$Return_i$ represents the annualized and market-adjusted monthly portfolio returns of trader i for all trades across all stock classes over the sample period. The main variable of interest, $SocialMediaTrade(\%)$, represents the share of social media-induced trades by a trader from all their equity trades placed in the corresponding exchanges during the sample period. The specification also controls for the total number of trades by an investor, the number of active trading months, leverage, short positions, and trading behavior in other exchanges, such as crypto, commodities, indices, and foreign exchange. The regression also includes investor-level characteristics, such as gender, age, and trading knowledge, which are derived from KYC questions and measured as an indicator variable equal to 1 if the investor has the relevant trading experience, attended trading courses, or holds a relevant degree, and 0 otherwise.

Table 7 presents the results. Column 1 presents results controlling for the number of transactions a trader places. In Column 2, we control for all observable trading behavior by a trader and find that a higher proportion of social media trades in an investor's portfolio is associated with a negative annualized portfolio return. Specifically, an increase in the proportion of trades induced by social media discussions, as a share of an investor's total trades, is associated with a decrease in the annualized returns of the investor's portfolio by 2.8 percentage points, holding all other observable investor characteristics constant. In Column 3, we control for country-specific differences between traders by including a country-fixed effect, which can account for any differences in the financial literacy of investors across countries and find similar results. The results from this table imply that the more an investor relies on social media to make trading decisions, the lower their annualized portfolio returns.

4 Mechanism and Additional Tests

The results so far indicate that social media-induced trades negatively impact retail investor performance. This section examines a few potential driving forces of this phenomenon. Retail investors tend to make performance-reducing trades out of ignorance of their informational disadvantage and overconfidence in their trading abilities (Odean, 1998, 1999; Barber and Odean, 2000). When stocks attract attention, retail investors tend to be net buyers, leading to price increases followed by reversals. With its absence of traditional oversight and its potential for misinformation, social media allows retail investors to disseminate personal interpretations of firm-specific information or disclosures. However, a related question remains: Do all retail investors respond to social media signals by exhibiting similarly sub-optimal trading performances?

4.1 Market Timing

4.1.1 When Do Retail Investors Trade?

The far reach of social media information exacerbates persuasion bias among investors who would be best served by learning to assess repetitive information before they act (DeMarzo et al., 2003). We hypothesize that most retail investors are late entrants to the market and make trades based on stocks they discover on social media. Consistent with **?**, which documented that the losses incurred by day traders are predominantly concentrated in stocks with high retail order imbalance and abnormal trading volume, we expect that most retail trading activity on a specific stock to be concentrated during periods of elevated social media activity. However, the theoretical predictions of Pedersen (2022) posit that short-term rational investors can identify signals about a firm and observe the formation of beliefs on social media, thereby capitalizing on market bubbles for short-term profits. We explore whether social media-induced trading can have episodes of positive returns and whether short-term rational investors can generate positive excess returns before the bubble's peak.

To validate these hypotheses, we carry out an event-study analysis at the individual trade level and run 21 separate regression analyses using the following specifications:

$$Return_{i,s,t} = \beta_1 Social\ Media\ Trade_{t-x} + \beta_n Trade\ Level\ Controls + \lambda_t + \mu_s \tag{6}$$

where, $Return_{i,s,t}$ represents trade-level returns for stock s traded on day t by investor i. The key variable of interest is $SocialMediaTrade_{t-x}$, which is one of 21 time-indicator variables indicating the number of days relative to the day when a stock experiences abnormal social media activity in the top 20%. For instance, the coefficient of $SocialMediaTrade_{t-10}$ reflects the average difference in returns between trades for stocks with abnormally high social media activity and all other equity trades placed ten days before a stock reaches abnormally high social media activity.

Panel A of Figure 2 plots β_1 coefficients and confidence intervals. The figure illustrates that stocks that experience significant levels of discussion on social media present different returns depending on the day a retail investor trades them. In the [-10; +10] window surrounding the peak of social media activity on the day $t = 0$, the average returns generated by social media-induced trades are positive and statistically significant up to five days before day $t = 0$. This suggests that rational short-term retail investors can profit from social media-induced trades if they execute them at least five days before the stock reaches its peak social media activity. However, returns from trades placed starting from day t = 0 are negative when social media activity is at its highest for the stock, supporting the hypothesis and theoretical predictions put forth by Pedersen (2022)'s model that naive retail investors tend to enter the market too late when price bubbles are about to collapse, and returns are reversing. Most importantly, the negative returns documented in Table 6 are mainly concentrated on trades initiated one day before or after the day when the social media activity for the stock exhibits high abnormal volumes.

Similarly, to examine the hypothesis that retail trading volume is highest on days when stocks have abnormally high social media activity, we repeat the specification presented in Equation (6) by replacing the dependent variable with $Trades_{i,s,t}$. This variable represents the natural log of the number of open buy trades for stock s traded by investor i on day t.

We ran 21 separate regressions to analyze this relationship.

$$Trades_{i,s,t} = \beta_1 Social\ Media\ Trade_{t-x} + \beta_n Trade\ Level\ Controls + \lambda_t + \mu_s \tag{7}$$

Panel B of Figure 2 shows the β_1 coefficients and confidence intervals. This figure displays the relationship between the difference in trade volumes and the proximity to a day with abnormally high levels of social media activity ($t = 0$). The coefficients indicate a marked increase in stock trading by retail investors when a stock experiences abnormally high levels of discussion on social media.

4.1.2 Does Social Media Predict Returns?

We next examine the performance of a trading strategy that sorts stocks based on abnormal social media activity on day 0, and tracks returns over a prolonged horizon. The stocks are sorted into two groups — the bottom 20% and top 20% — with abnormal social media activity, as computed by Equation (1). The market-adjusted returns are computed as the difference between daily stock returns and the value-weighted CRSP return. The cumulative market-adjusted returns, weighted by market capitalization, are then calculated for each quintile of stocks. Figure 3 illustrates the results, displaying the returns for 150 days, beginning when stocks with the highest (red line) and the lowest (blue line) abnormal social media activity are sorted. The figure indicates that stocks with the highest abnormal social media activity exhibit inferior performance over time compared to stocks with the lowest abnormal social media activity.

Next, to validate our findings and to further investigate the trajectory of market-adjusted future returns for social media-induced trades, we estimate a panel regression of the following form:

$$\begin{aligned} Return_{s,t+x} = {} & \beta_1 Social\ Media\ Trade + \beta_2 News + \beta_3 Past\ Returns + \\ & + \beta_4 Trading\ Volume + \lambda_t + \mu_s + \gamma_t \end{aligned} \tag{8}$$

The dependent variable is the return of a stock s on the day $t + x$. Table 8 presents the results and shows that controlling for the past news, trading volume, and past returns, the social media activity predicts statistically significant negative market-adjusted returns for at least up to a month if measured by a broader set of social media networks for a period of at least six months. Although social media-induced trades exhibit positive returns on the day of a trade, they reverse entirely the next day, and decrease over time. Tests in this section illustrate that stocks that have high levels of discussion in social media perform predictably worse compared to other equity trades. Taken together with earlier evidence from actual retail trades, we show that retail investors are worse off by trading such predictably performance-worsening stocks.

4.2 Disposition Effect

Another driving force behind the persistent losses experienced by retail investors from social media-influenced trades may be attributed to the disposition effect, which is one of the most well-established findings in the study of individual trading behavior (Kahneman and Tversky, 1979; Barberis and Xiong, 2012; Ingersoll and Jin, 2013). Retail traders tend to sell stocks when they are experiencing gains and hold onto stocks when they are incurring losses. We hypothesize that social media networks have amplified this behavioral bias, given that retail investors are susceptible to forming the belief that they are part of a more significant movement, or narrative, and tend to follow allegedly profitable trading strategies promoted on social media, such as the Reddit-based expresssion “diamond hands.” This term, often depicted in emoji form, refers to an investor with a high-risk tolerance for high-volatility stocks who holds onto their investment even under pressure to sell. The acronym “HODL,” which stands for “hold on for dear life” is also frequently seen in investment strategy discussions on social media networks.

To examine the presence of the disposition effect among retail investors on the Platform, we employ the following specification to determine if they have a higher tendency to sell stocks that are experiencing gains compared to those that are incurring losses:

$$Sale_{i,s,t} = \beta_1 Gain + \beta_2 Social\ Media\ Trade_{i,s,t} + \beta_3 Gain \times Social\ Media\ Trade_{i,s,t} \tag{9}$$

Our tests were conducted at the account (i), stock (s), and date (t) level and were restricted to a sample of long equity trades on U.S. exchanges. The variable $Sale_{i,s,t}$ is a dummy indicator equal to 1 if investor i sold stock s on date t, and 0 otherwise. The $Gain$ is another indicator variable equal to 1 if stock s was in a state of gain at the close of date t, and 0 otherwise. The coefficient of $Gain$ measures the increase in the probability of selling a position if it was at a gain rather than at a loss. The interaction term between $SocialMediaTrade_{i,s,t}$ and $Gain$ measures the difference in the disposition effect for stocks with high social media coverage and other stocks. To eliminate the possibility that retail investor inattention solely drives the results to their accounts rather than deliberate choices to sell, we followed the method in Chang et al. (2016b) and restricted our sample to only those periods (months) in which a retail investor conducted a sale of any security in their account. This ensures that the retail investor was attentive to their portfolio during that period. Table 9 presents the results. The first column shows the results for the full sample, but to eliminate the impact of day trading, which accounts for roughly one-third of all trades on *the Platform*, we exclude these trades from our sample and present the results in column 2. The coefficient of the interaction term is positive and statistically significant. It can be interpreted to indicate that when a retail investor sells stock, they are 0.9% more likely to sell a stock if it is at a gain. Columns 3 and 4 present the results when the sample is restricted to only non-leveraged trades and stocks, but they show similar results, confirming the presence of the disposition effect.

5 Conclusion

The distinct characteristics of social media — user-generated content, the lack of peer review, and its speed and reach through network effects — suggest that it can uniquely influence capital markets. With the rise of meme stock investment, this impact has been amplified in recent years, but its influence in the capital markets is not a brand new phenomenon. The SEC raised concerns as early as 2014 about the growing reliance of U.S. investors on social media as a source of information for stock research, investment strategy guidance, current news, and market discussions.[5]

It's also important to understand that social media-driven trading is not limited to Reddit, Discord, or Twitter ("X"), and that its audience is broader than might be expected. For instance, Facebook hosts numerous private and public groups with hundreds of thousands of members who share, discuss, and exchange information related to stock picking. TikTok has videos labeled as "stock picks" and "Robinhood investors" that collectively receive 4 billion views. "Traditional" social media platforms such as YouTube and Twitter feature influencers who discuss stocks, live stream their trading sessions, and post their stock recommendations to massive audiences.

The SEC is currently exploring using gamification and behavioral prompts to determine potential actions that can increase investor protection. This examination is crucial, as it impacts retail investor behavior, market prices, and gains and losses. Our research also included an investigation of how interactions between social media and other key players in financial markets — traditional news outlets, analysts, institutional investors, short sellers — impact the level of social media discussion about particular stocks. This paper offers important insights into the relationship between social media, retail trading, and investor performance. Our results indicate that stock discussions on social media platforms significantly impact individual trading decisions more than other attention-grabbing factors. They also show that retail investors underperform from trades placed when an asset has high abnormal discussions and mentions across all social media platforms. Lastly, our results reveal

[5] https://www.sec.gov/oiea/investor-alerts-bulletins/ia_socialmediafraud.html. Concerns regarding the impact of social media on capital markets have been echoed globally by securities market regulators. The European Securities and Markets Authority (ESMA) has expressed concern that spreading misleading trading advice via social media is putting investors at risk. Additionally, the International Organization of Securities Commissions (IOSCO) has established the Retail Market Conduct Task Force, identifying the impact of social media on investor behavior as a top concern.

that social media investors perform worse not only in their equity trades but also in trades involving foreign exchange currencies, cryptocurrencies, and commodities.

References

Barber, B. M., X. Huang, T. Odean, and C. Schwarz (2022). Attention-Induced Trading and Returns: Evidence from Robinhood Users. *The Journal of Finance 77*(6), 3141–3190. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13183.

Barber, B. M., S. Lin, and T. Odean (2023, May). Resolving a Paradox: Retail Trades Positively Predict Returns but Are Not Profitable. *Journal of Financial and Quantitative Analysis*, 1–35.

Barber, B. M. and D. Loeffler (1993). The "Dartboard" Column: Second-Hand Information and Price Pressure. *The Journal of Financial and Quantitative Analysis 28*(2), 273–284. Publisher: Cambridge University Press.

Barber, B. M. and T. Odean (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *The Journal of Finance 55*(2), 773–806. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00226.

Barber, B. M. and T. Odean (2008, April). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies 21*(2), 785–818.

Barber, B. M., T. Odean, and N. Zhu (2009, January). Do Retail Trades Move Markets? *Review of Financial Studies 22*(1), 151–186.

Barberis, N. and W. Xiong (2012, May). Realization utility. *Journal of Financial Economics 104*(2), 251–271.

Bartov, E., L. Faurel, and P. S. Mohanram (2018, May). Can Twitter Help Predict Firm-Level Earnings and Stock Returns? *The Accounting Review 93*(3), 25–57.

Blankespoor, E., G. S. Miller, and H. D. White (2014, January). The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter™. *The Accounting Review 89*(1), 79–112.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2021, October). Tracking Retail Investor Activity. *The Journal of Finance 76*(5), 2249–2305.

Bradley, D., J. Hanousek Jr., R. Jame, and Z. Xiao (2021). Place Your Bets? The Market Consequences of Investment Advice on Reddit's Wallstreetbets. *SSRN Electronic Journal*.

Chang, T. Y., D. H. Solomon, and M. M. Westerfield (2016a, February). Looking for Someone to Blame: Delegation, Cognitive Dissonance, and the Disposition Effect: Looking for Someone to Blame. *The Journal of Finance 71*(1), 267–302.

Chang, T. Y., D. H. Solomon, and M. M. Westerfield (2016b, February). Looking for Someone to Blame: Delegation, Cognitive Dissonance, and the Disposition Effect: Looking for Someone to Blame. *The Journal of Finance 71*(1), 267–302.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang (2014, May). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Review of Financial Studies 27*(5), 1367–1403.

Cookson, J. A. and M. Niessner (2020). Why Don't We Agree? Evidence from a Social Network of Investors. *The Journal of Finance 75*(1), 173–228. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12852.

Da, Z., J. Engelberg, and P. Gao (2011). In Search of Attention. *The Journal of Finance 66*(5), 1461–1499. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2011.01679.x.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy 98*(4), 703–738. Publisher: University of Chicago Press.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003, August). Persuasion Bias, Social Influence, and Unidimensional Opinions*. *The Quarterly Journal of Economics 118*(3), 909–968.

Duflo, E. and E. Saez (2002, July). Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics 85*(1), 121–148.

Engelberg, J., C. Sasseville, and J. Williams (2012). Market Madness? The Case of "Mad Money". *Management Science 58*(2), 351–364. Publisher: INFORMS.

Farrell, M., T. C. Green, R. Jame, and S. Markov (2022, August). The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics 145*(2), 616–641.

Heimer, R. Z. (2016). Peer Pressure: Social Interaction and the Disposition Effect. *The Review of Financial Studies*, 3177–3209.

Ingersoll, J. E. and L. J. Jin (2013, March). Realization Utility with Reference-Dependent Preferences. *The Review of Financial Studies 26*(3), 723–767.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–291. Publisher: [Wiley, Econometric Society].

Kelley, E. K. and P. C. Tetlock (2013, June). How Wise Are Crowds? Insights from Retail Orders and Stock Returns: Retail Orders and Stock Returns. *The Journal of Finance 68*(3), 1229–1265.

Liang, B. (1999, January). Price Pressure: Evidence from the "Dartboard" Column. *The Journal of Business 72*(1), 119–134.

Merli, M. and T. Roger (2014, July). What drives the herding behavior of individual investors?:. *Finance Vol. 34*(3), 67–104.

Miller, G. S. and D. J. Skinner (2015). The Evolving Disclosure Landscape: How Changes in Technology, the Media, and Capital Markets Are Affecting Disclosure. *Journal of Accounting Research 53*(2), 221–239. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-679X.12075.

Nofsinger, J. R. and R. W. Sias (1999). Herding and Feedback Trading by Institutional and Individual Investors. *The Journal of Finance 54*(6), 2263–2295. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00188.

Odean, T. (1998). Are Investors Reluctant to Realize Their Losses? *The Journal of Finance 53*(5), 1775–1798. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00072.

Odean, T. (1999, December). Do Investors Trade Too Much? *American Economic Review 89*(5), 1279–1298.

Ouimet, P. and G. Tate (2020). Learning from Coworkers: Peer Effects on Individual Investment Decisions. *The Journal of Finance 75*(1), 133–172. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12830.

Pedersen, L. H. (2022, June). Game on: Social networks and markets. *Journal of Financial Economics*, S0304405X22000964.

Shiller, R. J., S. Fischer, and B. M. Friedman (1984). Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity 1984*(2), 457.

Shiller, R. J. and J. Pound (1989, August). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization 12*(1), 47–66.

Shleifer, A. and L. H. Summers (1990, May). The Noise Trader Approach to Finance. *Journal of Economic Perspectives 4*(2), 19–33.

Sias, R. W. (2004, January). Institutional Herding. *Review of Financial Studies 17*(1), 165–206.

Tetlock, P. C. (2011, May). All the News That's Fit to Reprint: Do Investors React to Stale Information? *Review of Financial Studies 24*(5), 1481–1512.

6 Figures

Figure 1. The Popularity of r/WSB

This figure illustrates the number of registered users of r/WSB over time. The Y-axis is in millions. Source: Reddit.

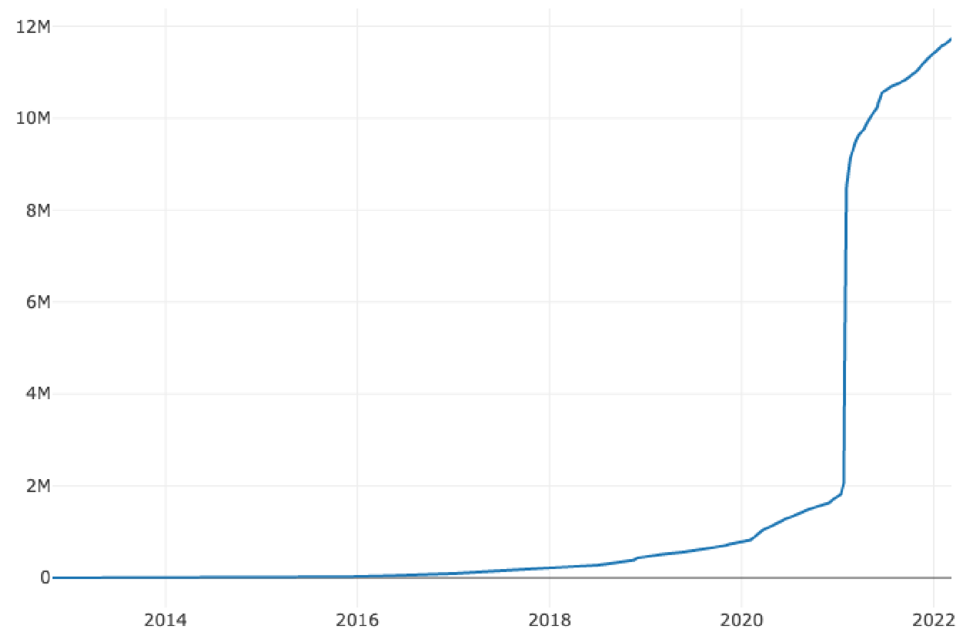
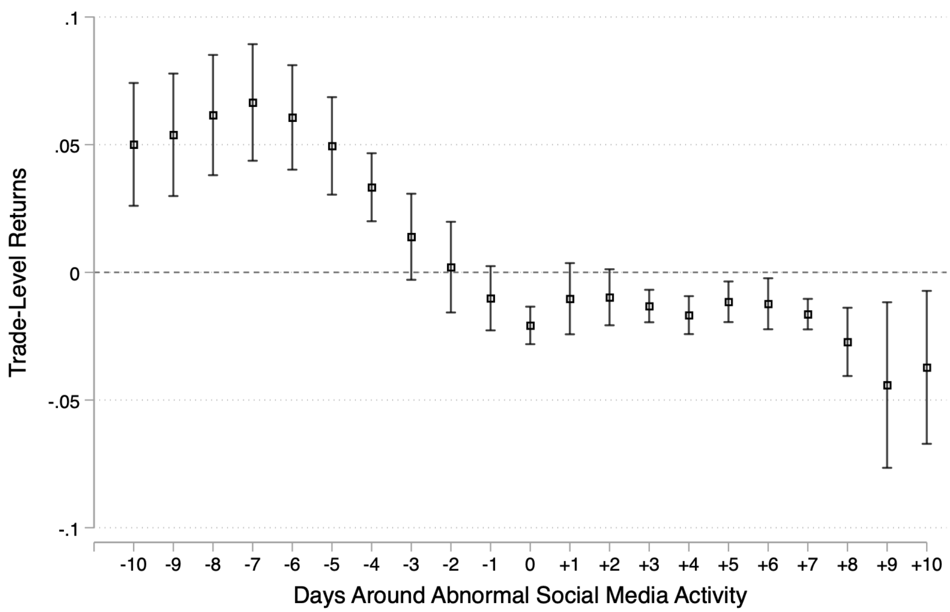


Figure 2. Trade-level Returns of Retail Investors and Social Media: RMA

The figure shows investor trade-level performance and trading behavior around high abnormal social media activity days. Panel A plots the coefficient of the variable of interest from 21 regressions in equation (6), where the dependent variable is trade-level returns. Panel B plots the coefficients of log number of opened positions estimated from equation (7). Each regression has one of the 21-time indicator variables, from $t-10$ to $t+10$, indicating the number of days relative to the day t when the stock was classified as a social media trade, zero otherwise. Figures plot the coefficients from these time indicator variables. For example, the coefficient on the $t-10$ shows the average difference in trading between stocks with abnormally high social media activity in 10 days and all other stocks.

Panel A: Trade-Level Returns



Panel B: Timing of the Trading Behavior of Users in the Platform

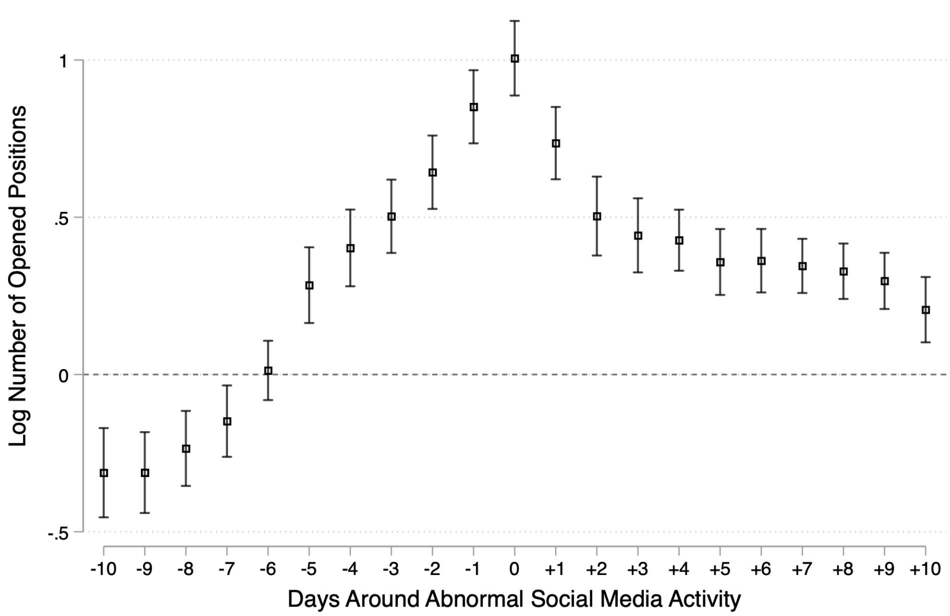
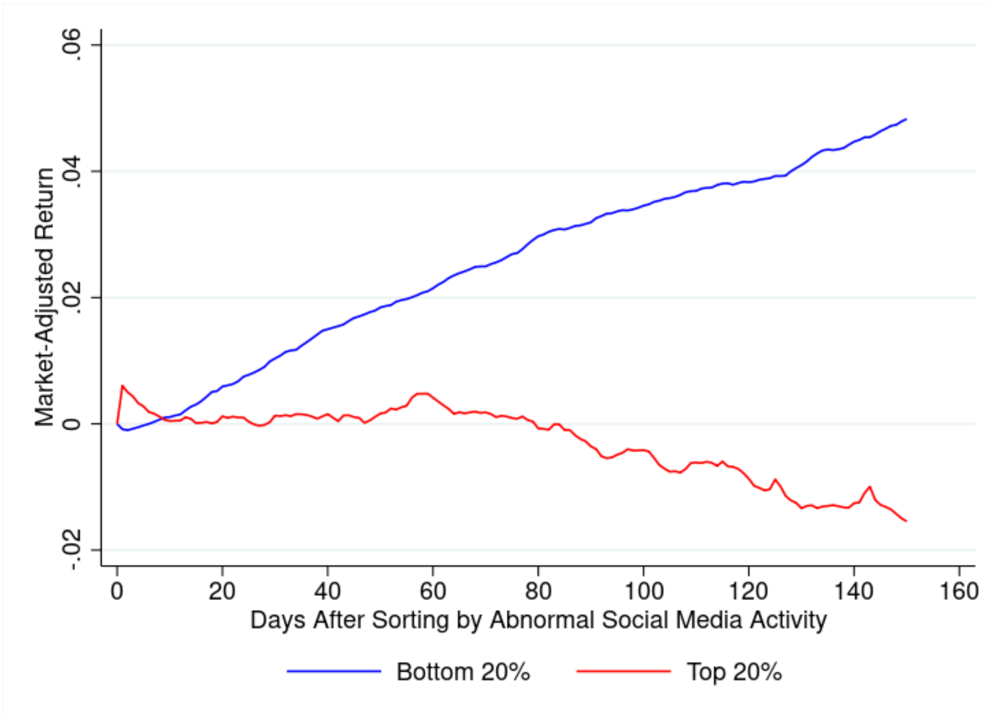


Figure 3. Market-Adjusted Returns of Stocks in Top and Bottom Quintiles by Social Media Activity

The figure depicts the cumulative market-adjusted returns over 150 days after market close on day 0 for stocks in the top and bottom quintiles of abnormal social media activity.



7 Tables

Table 1. r/WSB sample comments

This table illustrates comments posted in the "Tomorrow's Moves" and "Daily Discussions" sub-forums of r/WSB from 12:53 p.m. to 12:56 p.m. on March 26, 2021. Column 1 shows the ticker extracted from a comment; column 2 shows the timestamp (EST), and column 3 shows the sentiment score for each comment, calculated by the VADER method. Panel B shows five comments from r/WSB's "Tomorrow's Moves" and "Daily Discussions." Source: Reddit.

Comment	Company	Sentiment score
TSLA earnings call is a joke. So much uncertainty and indirect answers. No reason for the stock price to moon.	Tesla	-0.5506
I bought FB 252.5 puts for like 30$, cant wait to see my gains tomorrow	Facebook	0.5994
Lmaoooo, $MTCH missed across the board and has to pay half their revenue because of a lawsuit and it's still green Lmaoo	Match	-0.5423
$PYPL missed by 2 cents and is -30. Lmaoooooooooo	Paypal	-0.296
Time to buy GOOG calls were yesterday but I'll do so today because it will increase a few hundred points after the split	Google	0.4497

Table 2. Trader Characteristics

	All Traders	**Social-Media Induced Traders**	**Regular Traders**	**χ^2 or T-Stat**
Number of Traders	1,083,319	208,041	875,278	
Number of Trades	53,721,873	7,247,041	46,448,947	
Short Positions, %	6.49%	6.70%	6.43%	-
Leveraged Trades, %	26.92%	21.65%	27.76%	
Average Holding Days	12.99	9.03	13.62	
Female, %	12.85	12.61	12.91	t=3.7 (p<0.01)
Age Range, %				
18-24	16.31	21.38	15.11	
25-34	41.47	45.79	40.45	
35-44	25.69	21.38	26.72	χ^2=1.1e+04
45-54	11.31	7.95	12.11	(p<0.01)
55-64	4.13	2.78	4.46	
>65	0.90	0.60	0.98	
Trading Education, %				
No Financial Knowledge	41.86	43.67	41.43	
Trading Courses	34.74	34.02	34.91	χ^2=422.7
Degree or experience	12.52	12.28	12.57	(p<0.01)
Professional	10.87	10.02	11.07	
Trading Experience - Equities (past year), %				
Never traded	38.69	41.21	38.06	
0-10 times	31.95	33.34	31.61	χ^2=2.1e+03
10-20 times	11.99	11.45	12.12	(p<0.01)
Above 20 times	17.38	13.99	18.21	
Net Annual Income (USD), %				
Up to $10K	19.55	21.04	19.20	
$10K-$50K	47.55	47.61	47.54	
$50K-$200K	24.79	24.41	24.88	χ^2=736.4
$200K-$500K	2.92	2.50	3.02	(p<0.01)
$500K-$1M	1.65	1.41	1.70	
>$1M	3.53	3.03	3.65	
Primary Purpose of Trading, %				
Short Term Returns	19.65	22.58	18.96	
Additional Revenues	51.95	50.36	52.33	χ^2=1.5e+03
Future Planning	20.68	19.19	21.03	(p<0.01)
Saving For Home	7.72	7.88	7.68	
Preferred Risk-Reward Scenario, %				
Gain 5% / Lose -3%	3.35	3.48	3.32	
Gain 10% / Lose -6%	9.28	9.23	9.29	χ^2=47.1
Gain 20% / Lose -12%	26.99	26.49	27.11	(p<0.01)
Gain 40% / Lose -24%	34.46	34.61	34.43	
Gain 80% / Lose -48%	25.91	26.19	25.85	
Trading Strategy (Duration), %				
Short (Up to 24 Hours)	16.70	15.11	17.07	
Medium (Few Weeks to Months)	54.20	57.14	53.51	χ^2=769.2
Long (More Than Several Months)	29.10	27.75	29.42	(p<0.01)

Table 3. Attention Induced Trading and Retail Trading in the Platform: Sorting

This table reports average open-close imbalances (estimating equation (2) for stocks in relation to four attention-grabbing variables. In the first two columns, stocks are sorted into ten deciles by past-day returns and abnormal trading volume. In column 3, stocks are sorted into quintiles based on abnormal news article volume. In column 4, stocks with at least five mentions on social media are sorted into quintiles.

Source:	**CRSP**	**CRSP**		**RMA**	**r/WSB**
	1	**2**		**3**	**4**
Decile	$Return_{t-1}$	$Abnormal$ $TradingVolume_{t-1}$	*Quintile*	$Abnormal$ $News_{t-1}$	$Abnormal$ $SocialMedia_t$
1	0.0669	0.0134	Zero/Few	-	0.0255
2	0.0641	0.0250	1	0.0358	0.0224
3	0.0600	0.0331	2	0.0221	0.0378
4	0.0557	0.0442	3	0.0322	0.0414
5	0.0517	0.0513	4	0.0475	0.0694
6	0.0500	0.0524	5	0.0716	0.1090
7	0.0445	0.0560			
8	0.0354	0.0629			
9	0.0219	0.0592			
10	0.0296	0.0674			

Table 4. Attention Inducing Factors and Retail Trading in the Platform: Regression

This table reports the coefficient of standard OLS regressions from estimating equation (3). All specifications include stock and date effects. Standard errors are clustered at the stock and date levels and are in parentheses. *p<0.10, ** p<0.05, *** p<0.01.

Panel A: Open - Close Imbalance

Dependent Variable:	Open-Close Imbalance				
	(1)	(2)	(3)	(4)	(5)
Social Media Trade	0.070***				0.063***
	(0.005)				(0.005)
News Trade		0.004**			0.001
		(0.002)			(0.002)
Volume Trade			0.013***		0.007***
			(0.002)		(0.002)
Return Trade				0.012***	0.010***
				(0.002)	(0.003)
Date FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	556,229	480,125	545,259	545,931	471,750
R^2	0.046	0.039	0.045	0.045	0.039

Panel B: Log Number of Open Trades

Dependent Variable:	Open-Close Imbalance				
	(1)	(2)	(3)	(4)	(5)
Social Media Trade	1.218***				1.003***
	(0.039)				(0.035)
News Trade		0.247***			0.135***
		(0.008)			(0.006)
Volume Trade			0.477***		0.381***
			(0.010)		(0.009)
Return Trade				0.445***	0.323***
				(0.011)	(0.011)
Date FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	516,591	445,960	506,554	507,207	438,292
R^2	0.748	0.745	0.753	0.750	0.765

Table 5. Attention Inducing Factors and Retail Trading in the Market

This table reports OLS regression results. The dependent variable represents the ratio of total USD traded by retail investors in a given stock divided by the total USD traded by retail investors across all stocks in U.S. exchanges. All equations include the date and stock fixed effects. Standard errors are clustered at the stock and date levels and are in parentheses. *p<0.10, ** p<0.05, *** p<0.01.

Dependent Variable:	Retail Share		
	(1)	(2)	(3)
Social Media Trade	0.0024***	0.0023***	0.0022***
	(0.0002)	(0.0002)	(0.0002)
Return Trade		0.0001***	
		(0.0000)	
Volume Trade		0.0001***	
		(0.0000)	
News Trade		0.0000***	
		(0.0000)	
Log News			0.0000***
			(0.0000)
Return [t-2;t-1]			-0.0000
			(0.0000)
Return [t-5;t-2]			0.0000
			(0.0000)
Return [t-21;t-6]			0.0000*
			(0.0000)
Log Trading Volume			0.0002***
			(0.0000)
Date FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Observations	1,917,344	1,744,492	1,742,066
R^2	0.598	0.601	0.605

Table 6. Social Media-Induced Trading and Trade Returns

This table reports OLS regression results from estimating equation (4). The dependent variable is trade-level returns for stock s, traded on the day t, by an investor i. $SocialMediaTrade$ takes two forms - an indicator variable defined as in the equation (1) and a log number of mentions. $Trade\ Size$ - share of total equity balance of investor i invested to stock position s on the day t, $Short\ Position$ - indicator variable equal to one for short positions, $Leveraged\ Position$ - indicator variable equal to one for leveraged positions, $Holding\ Days$ - natural logarithm of the number of calendar days between open and close dates of the transaction. All equations include investor, stock, and date fixed effects. Standard errors are clustered at the stock and date levels and are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Dependent Variable	Trade-level Returns				
Social Media Trade Variable Definition:	Abnormal Indicator	Abnormal Indicator	Raw Log	Raw Log	Abnormal Indicator (Premarket)
	(1)	(2)	(3)	(4)	(5)
Social Media Trade	-0.0224***	-0.0161***	-0.0178***	-0.0172***	-0.0277***
	(0.0054)	(0.0059)	(0.0014)	(0.0014)	(0.0084)
News Trade		-0.0062*			-0.0046
		(0.0036)			(0.0036)
Volume Trade		-0.0085***			-0.0034
		(0.0028)			(0.0026)
Return Trade		-0.0082***			-0.0064**
		(0.0025)			(0.0025)
Trade Size, % of Account Balance			0.0040***	0.0031**	0.0018
			(0.0011)	(0.0013)	(0.0012)
Short Position			-0.0308***	-0.0309***	-0.0315***
			(0.0056)	(0.0057)	(0.0056)
Leveraged Position			-0.0068**	-0.0099***	-0.0049
			(0.0033)	(0.0032)	(0.0033)
Holding Days (Log)			0.0348***	0.0337***	0.0356***
			(0.0023)	(0.0023)	(0.0023)
Return [t-1]				-0.0811***	
				(0.0199)	
Return [t-5;t-2]				-0.0507***	
				(0.0065)	
Return [t-21;t-6]				-0.0147***	
				(0.0030)	
Trading Volume (Log)				0.0120***	
				(0.0034)	
News Volume (Log)			-0.0084***	-0.0054***	
			(0.0020)	(0.0016)	
Date FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Investor FE	Yes	Yes	Yes	Yes	Yes
Observations	46,298,502	42,951,475	42,960,549	41,845,930	42,951,475
R^2	0.120	0.125	0.144	0.147	0.140

Table 7. Attention Inducing Factors and Retail Trading in the Market

This table reports OLS regression results. The dependent variable represents the ratio of total $USD traded by retail investors in a given stock divided by the total $USD traded by retail investors across all stocks in U.S. exchanges. All equations include the date and stock fixed effects. Standard errors are clustered at the date level. Standard errors are clustered at the stock and date levels and are in parentheses. *p<0.10, ** p<0.05, *** p<0.01.

Dependent Variable:	Annualized Portfolio Returns		
	(1)	(2)	(3)
Social Media Trade %	-0.0172***	-0.0277***	-0.0280***
	(0.0009)	(0.0008)	(0.0008)
Log # of Positions	-0.0207***	-0.0084***	-0.0081***
	(0.0001)	(0.0001)	(0.0001)
Log # of Trading Months		0.0146***	0.0139***
		(0.0003)	(0.0003)
Crypto Trade %		0.0607***	0.0568***
		(0.0006)	(0.0007)
FX Trade %		-0.1394***	-0.1292***
		(0.0028)	(0.0028)
Commodity Trade %		-0.1417***	-0.1415***
		(0.0017)	(0.0017)
Index Trade %		-0.1071***	-0.1141***
		(0.0022)	(0.0022)
Leveraged Trade %		-0.0547***	-0.0525***
		(0.0008)	(0.0008)
Short Positions %		-0.1068***	-0.1057***
		(0.0015)	(0.0015)
Male			-0.0079***
			(0.0005)
Trading Knowledge			0.0008**
			(0.0003)
Age 18-24			-0.0168***
			(0.0015)
Age 25-34			-0.0107***
			(0.0015)
Age 35-44			-0.0083***
			(0.0015)
Age 45-54			-0.0074***
			(0.0015)
Age 55-64			-0.0045***
			(0.0016)
Observations	1,082,330	1,082,330	1,082,007
R^2	0.030	0.125	0.130

Table 8. Social Media Activity and Stock Returns

This table reports OLS regression results from estimating the equation (8). All equations include the date and stock fixed effects. Standard errors are clustered at the date and stock level. T-statistics are reported in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Dependent Variable:	**Market-Adjusted Return**							
Period:	**[t]**	**[t;t+1]**	**[t;t+2]**	**[t;t+5]**	**[t;t+10]**	**[t;t+21]**	**[t;t+63]**	**[t;t+126]**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Media Trade	0.040***	-0.013***	-0.017***	-0.020**	-0.034***	-0.053***	-0.115***	-0.158***
	(2.897)	(-5.453)	(-5.515)	(-2.252)	(-2.981)	(-3.636)	(-4.381)	(-3.196)
Log (News Volume$_{t-1}$)	-0.002***	-0.002***	-0.002***	-0.002**	-0.002	-0.004	-0.008**	-0.015***
	(-5.978)	(-6.827)	(-5.210)	(-2.462)	(-1.066)	(-1.381)	(-2.362)	(-2.999)
Returns$_{t-1}$	-0.001	0.004	0.002	-0.004	-0.011***	-0.022***	-0.046***	-0.081***
	(-0.054)	(1.054)	(0.489)	(-1.039)	(-3.226)	(-6.201)	(-7.220)	(-4.631)
Returns$_{t-5;t-2}$	-0.011	0.000	-0.002	-0.007***	-0.015***	-0.026***	-0.052***	-0.090***
	(-0.883)	(0.153)	(-1.074)	(-5.118)	(-10.627)	(-9.189)	(-7.447)	(-4.405)
Returns$_{t-21;t-6}$	-0.002***	-0.002***	-0.004***	-0.011***	-0.020***	-0.036***	-0.072***	-0.126***
	(-5.502)	(-3.964)	(-7.951)	(-17.766)	(-12.350)	(-7.911)	(-7.098)	(-4.221)
Log (Trading Volume$_{t-1}$)	0.019***	0.014***	0.017***	0.024***	0.032***	0.042***	0.053**	0.065*
	(9.147)	(7.622)	(6.411)	(4.205)	(3.412)	(3.224)	(2.487)	(1.927)
Date FE	Y	Y	Y	Y	Y	Y	Y	Y
Stock FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,332,056	2,332,011	2,331,987	2,331,964	2,331,934	2,331,880	2,331,562	2,327,287
R-squared	0.015	0.012	0.011	0.013	0.021	0.038	0.099	0.154

Table 9. Disposition Effect

This table reports OLS regression results from estimating the equation (9). T-statistics are reported in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Sample:	All	Non-Day Trades	Non-Leveraged Trades	Only Stocks
	(1)	(2)	(3)	(4)
Gain	0.021***	0.021***	0.022***	0.021***
	(11.261)	(10.423)	(13.602)	(10.208)
Social Media Trade	0.002	0.003*	0.004**	0.001
	(1.283)	(1.744)	(2.531)	(0.516)
Gain × Social Media Trade	0.006***	0.009***	0.007***	0.008***
	(3.097)	(4.506)	(3.981)	(4.092)
Observations	382,895,203	341,652,699	290,998,460	321,259,970
R-squared	0.002	0.002	0.003	0.002

Appendix

Other Attention Grabbing Factors

Barber and Odean (2008) found that retail investors display attention-driven trading behavior. Retail investors concentrate on buying stocks on high-trading volume days, on days with both extremely negative and extremely positive one-day returns, and on days when stocks are in the news. We study how social media-induced attention trading relates to other attention-inducing factors: abnormal trading volume, past-day returns, and coverage of stock in the news.

News Coverage

Our data on traditional news coverage comes from Refinitiv's MarketPsych Analytics (RMA), which monitors all major news outlets (top 4,000 international business sources, top regional news sources, and leading industry sources), both in print and online. Following a methodology similar to abnormal social media activity, we define abnormal news coverage for a stock s on the day t to be:

$$Abnormal\ News\ Coverage_{s,t} = \frac{News\ Mentions_{s,t}}{Average\ News\ Mentions_{s,[t-47;t-6]}}$$

Then, on the day t, we sort the abnormal news coverage variable for all stocks with at least two news articles to quintiles and define indicator variable $News\ Induced\ Trade$ equal to one if the stock's one day lagged abnormal news coverage variable falls in the top 20% and 0 otherwise.

Trading Volume

We calculate abnormal trading volume similar to our social media activity measure. On day t, we calculate the abnormal trading volume for stock s as a ratio of trading volume on the day t, as reported in CRSP, to an average trading volume through day $t47$ to $t6$. We use the lagged abnormal trading volume in our analysis and define the $VolumeInducedTrade$ indicator variable as equal to 1 if the stock's one-day lagged abnormal trading volume falls in the top 20% percentile on the day t.

Past Day Returns

When stocks have extreme daily returns, it is likely to draw the attention of retail investors, and we expect those retail investors to trade in response to both negative and positive price changes. To test this, we sort stocks into deciles based on the previous trading day's returns to account for the fact that many investors learn of, or react to, prices after the market closes. On day t, we calculate the stock's return from day $t-2$ to $t-1$ and define the indicator variable $Return\ Induced\ Trade$ equal to 1 if a stock's absolute past day return is in the top 20% percentile.

Figure A1. Message rooms of r/WSB.

This figure shows a screenshot of r/WSB's thematic threads: "Tomorrow's Moves" and "Daily Discussions." Source: Reddit.

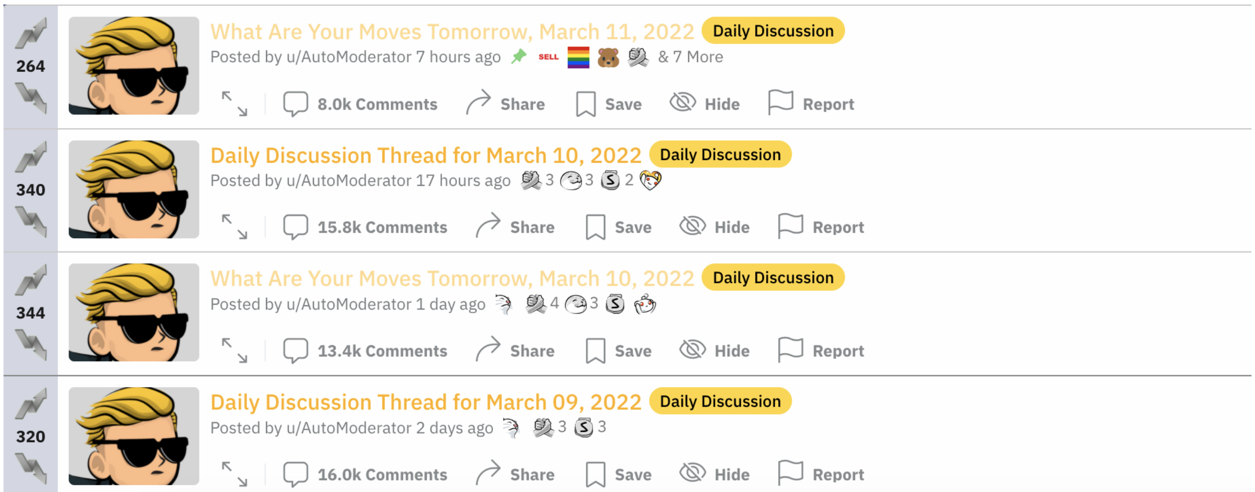


Figure A2. Distribution of comments in r/WSB by days and hours activity

This figure shows the distribution of comments in r/WSB by weekdays and by hours of the day. Source: Reddit

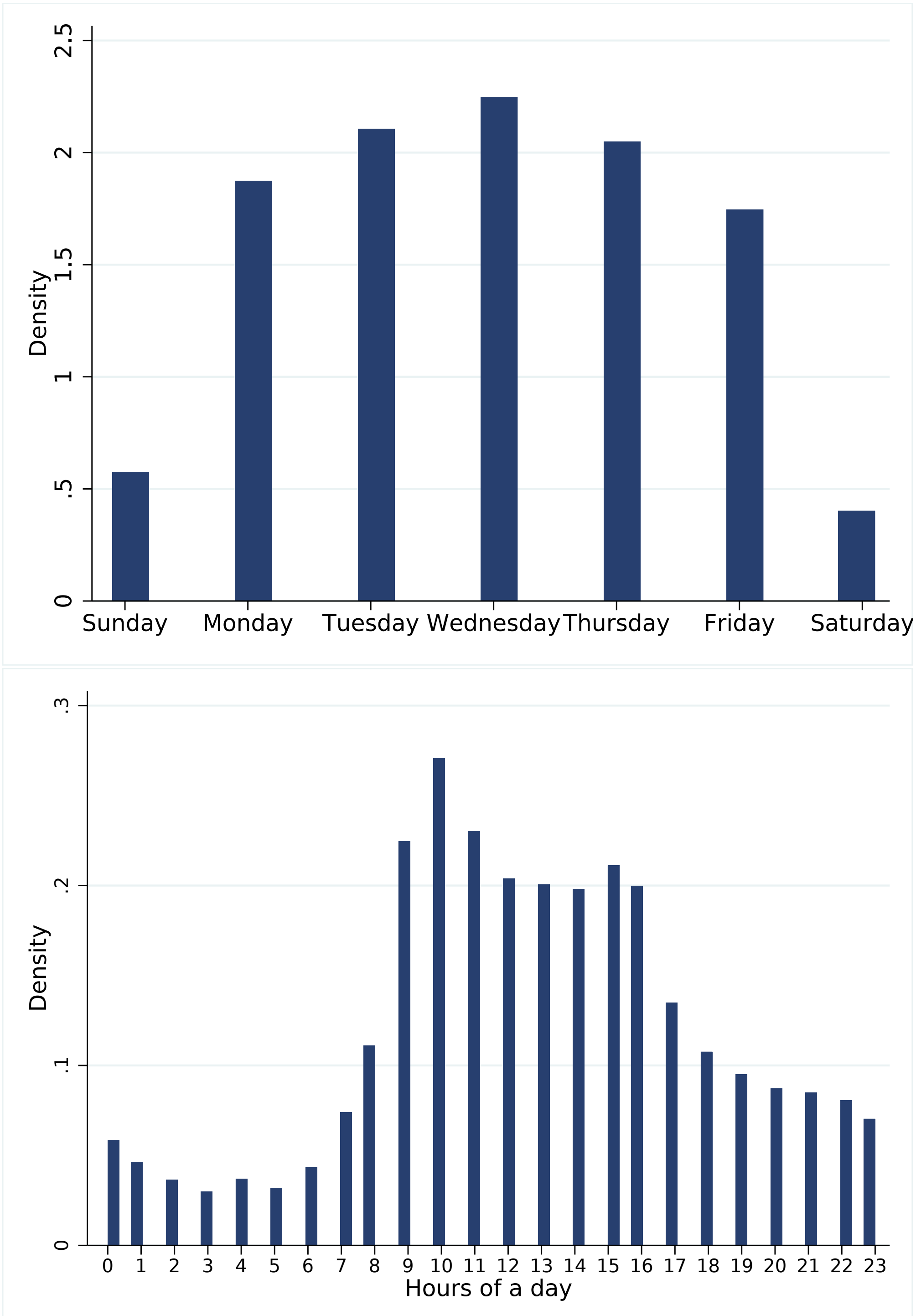
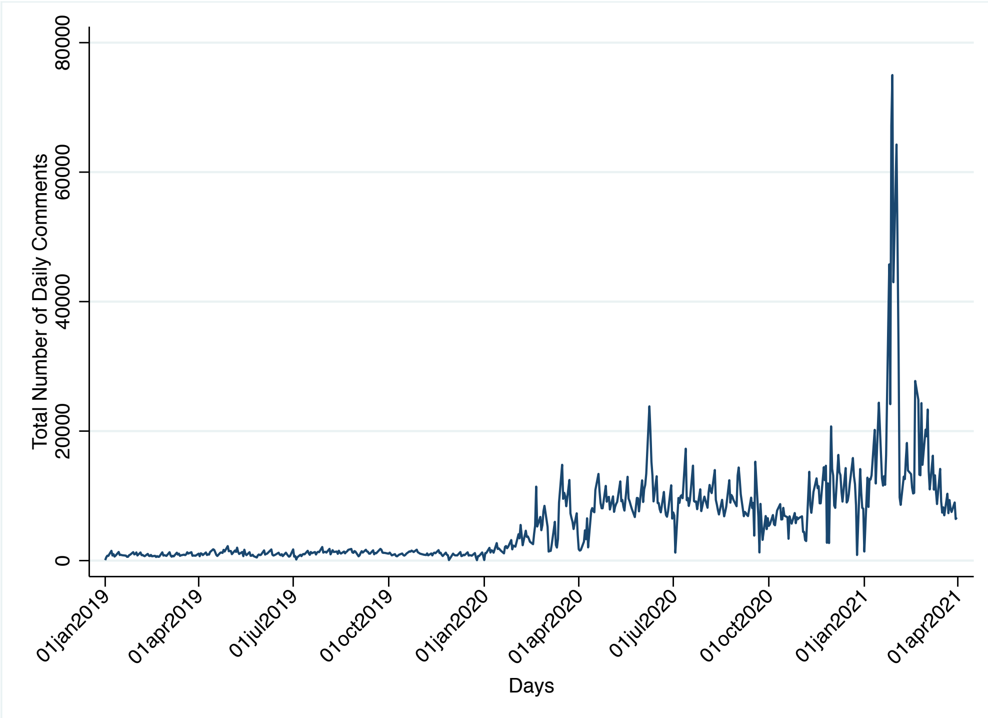


Figure A3. Daily activity in r/WSB.

This figure shows the daily total number of comments (upper panel) and the daily total number of mentioned stocks (lower panel) in r/WSB during the sample period. Source: Reddit



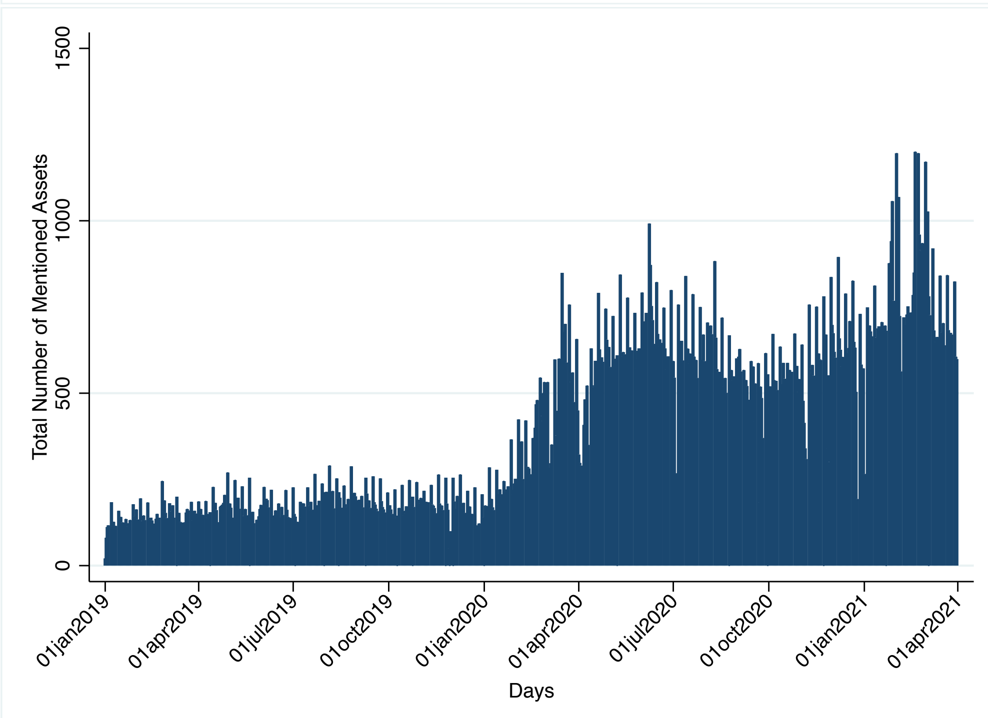


Table A1. Summary Statistics of RMA and r/WSB Samples

Panel B: r/WSB sample

	Comments			Stocks	Sentiment Score		Users
	All	Bullish	Bearish	N	Mean	SD	N
Full Sample	4,006,825	1,549,196	1,072,787	5,822	0.064	0.429	9,696,811
By quarters							
2019 Q1	50,267	19,949	15,000	1,198	0.050	0.442	520,542
2019 Q2	66,430	26,717	19,435	1,236	0.058	0.438	583,400
2019 Q3	74,608	28,074	22,453	1,402	0.037	0.435	663,339
2019 Q4	58,936	22,166	17,845	1,372	0.037	0.437	773,269
2020 Q1	245,924	88,424	75,278	2,180	0.024	0.433	1,070,055
2020 Q2	514,205	190,639	149,169	2,990	0.040	0.430	1,308,914
2020 Q3	523,658	194,864	147,475	2,913	0.046	0.425	1,523,676
2020 Q4	530,543	198,494	141,695	2,981	0.054	0.419	1,762,121
2021 Q1	1,010,455	402,972	258,135	3,845	0.078	0.436	9,709,213
2021 Q2	596,720	247,727	139,914	3,626	0.103	0.425	10,620,004
2021 Q3	332,145	128,098	85,519	3,332	0.070	0.420	10,893,421

Panel A: RMA Sample

	Mentions	Stocks	Sentiment Score	
	All		Mean	SD
Full Sample		4,918	0.052	0.427
By quarters				
2019 Q1	7,648,300	4,104	-0.026	0.503
2019 Q2	7,867,931	4,157	0.024	0.470
2019 Q3	6,626,285	4,149	0.042	0.438
2019 Q4	7,354,159	4,160	0.049	0.432
2020 Q1	8,391,080	4,155	0.013	0.418
2020 Q2	9,338,791	4,145	0.017	0.403
2020 Q3	11,816,515	4,153	0.036	0.399
2020 Q4	11,461,362	4,200	0.085	0.385
2021 Q1	11,800,767	4,316	0.109	0.388
2021 Q2	11,577,061	4,455	0.124	0.404
2021 Q3	9,592,437	4,534	0.131	0.401

Table A2. Top-20 Assets With The Highest Number Mentions in RMA and Comments In r/WSB

RMA			r/WSB		
Ticker	Name	Number of Mentions	Ticker	Name	Number of Comments
TSLA	Tesla Inc	6,719,764	GME	Gamestop Corp New	324,592
AMC	AMC Entertainment Holdings Inc	5,010,710	SPY	SPDR S&P 500 ETF Trust	297,243
AAPL	Apple Inc	3,517,540	TSLA	Tesla Inc	203,905
AMZN	Amazon Com Inc	2,736,730	AMC	AMC Entertainment Holdings Inc	183,392
GME	Gamestop Corp New	2,088,310	PLTR	Palantir Technologies Inc	137,763
NIO	NIO Inc	1,270,427	BB	Blackberry Ltd	137,723
BA	Boeing Co	1,206,104	AAPL	Apple Inc	97,026
MSFT	Microsoft Corp	1,098,436	AMD	Advanced Micro Devices Inc	93,730
AMD	Advanced Micro Devices Inc	1,086,395	NIO	NIO Inc	69,043
NFLX	Netflix Inc	894,194	MSFT	Microsoft Corp	60,786
SRNE	Sorrento Therapeutics Inc	822,574	RKT	Rocket Companies Inc	50,035
DIS	Disney Walt Co	805,231	NOK	Nokia Corp	49,045
PFE	Pfizer Inc	784,416	SPCE	Virgin Galactic Holdings Inc	48,711
WKHS	Workhorse Group Inc	762,677	BA	Boeing Co	47,412
IPOA	Social Cap Hedosophia Hldgs Corp	735,728	AMZN	Amazon Com Inc	47,348
BABA	Alibaba Group Holding Ltd	722,820	BABA	Alibaba Group Holding Ltd	42,357
ROKU	Roku Inc	691,966	CLOV	Clover Health Investments Corp	37,953
NVAX	Novavax Inc	660,337	FB	Facebook Inc	34,477
INO	Inovio Pharmaceuticals Inc	655,517	QQQ	Invesco QQQ Trust	33,028
FCEL	Fuelcell Energy Inc	642,607	DIS	Disney Walt Co	31,691