

Social Media Attention, Stock Returns and Retail Trades ^{*}

Weiting Hu[†] Rusi Yan[‡]

October 2022

Abstract

The rise of social media and retail trading platforms has potentially changed the importance of retail investors in the stock market. Using data from Reddit WallStreetBets, we examine the market consequences of abnormal social media attention and its relationship with retail investor trading activities. We find that high social media attention on Reddit is associated with higher retail trading activities, cumulative returns and price informativeness. When social media attention occurs in conjunction with earnings announcements, social media attention increases the diffusion of information among retail investors, as reflected in discussion tones, directional retail trading, and price informativeness measure. While Reddit helps spread information to the general retail population, investors on Robinhood exhibit opposite trading directions, consistent with previous studies that Robinhood investors are relatively inexperienced. Our findings suggest that high social media attention affects stock returns through diffusion of information to retail investors.

JEL classifications: G11, G12, G14, M41

Keywords: Reddit, WallStreetBets, social media, retail investors, price informativeness, earnings announcement, Robinhood.

^{*}We would like to thank Ana Babus for her invaluable guidance and support throughout the project. We are grateful to Philip Dybvig, Xing Huang, Anthony Cookson, George-Levi Gayle and Yongseok Shin for helpful comments.

[†]Washington University in St. Louis, email: weiting.hu@wustl.edu

[‡]Corresponding author. Washington University in St. Louis, email: rusi.yan@wustl.edu

1 Introduction

Initiated by the Reddit WallStreetBets community, the GameStop event caused significant swing in the company's stock price.¹ The incident has put social media platform and retail investors in the spot light. While social media and retail investors have each played an increasingly important role in the stock market, the previous papers have focused on the impact of investment research provided by social media for retail investors (Chen, De, Hu, and Hwang, 2014; Farrell, Green, Jame, and Markov, 2022). How the volume of social media attention affects retail investors and the financial market is less studied. The intertwined relationship between WallStreetBets and retail investors spurs interest to the question - does social media attention affect retail trading activities and what are the market consequences?

Our paper seeks to address this question by providing evidence that social media attention is the mechanism through which retail investor trades have market consequences. Using submissions and comments from WallStreetBets in 2018-2020, we show that WallStreetBets attention is correlated with retail investor trades, and helps incorporate firm specific fundamental information into price. With high social media attention, the concerned stock sees higher retail trades, cumulative returns and price informativeness. Moreover, in a positive earnings surprise under high social media attention, discussions are more bullish, retail buying trades are higher and price informativeness is higher. Our results show that both the level and content of social media attention are related to retail trading activities and help diffuse firm fundamental information to retail investors. Moreover, we find opposite trading patterns for Robinhood investors, consistent with the literature suggesting heterogeneity in retail investors.

¹Moving from \$17.25 at beginning of the month, GameStop's stock price climbed nearly 30 times to around \$500 at end of the month, and plummeted back to below \$50 at the beginning of February.

The rest of the introduction provides a more detailed summary of our findings.

We focus our attention on the period before the GameStop event to isolate from potential changes in user composition due to the rapid increase of subscribers in the aftermath of the event.² Social media attention is measured by the sum of all submissions and comments collected from Reddit WallStreetBets. Abnormal social media attention is identified when total WallStreetBets mention exceeds a certain threshold multiple of the previous 6-month average attention for at least two consecutive days. Stocks are separated into low and high retail concentration groups.

We start by examining the market consequences of abnormal social media attention. We found that net retail buying volume is significantly higher following the abnormal attention event on Reddit, controlling for the average net retail buying volume of stocks from the same beta group. The increase in trading volume is higher for the group of stocks with high retail investor concentration. Cumulative returns measured at the onset of the abnormal attention event and after the event are both positive after controlling the average return of the stock's beta group, with a larger magnitude of change for the high retail concentration group. Price informativeness is higher from the start of the abnormal social media attention event.

The division of the abnormal social media attention events into continued or dropped attention provides further support on the relationship between social media attention and retail trading activities, and its potential market consequences. Continued attention events maintain abnormal attention for at least five trading days, while dropped attention sees a quick return of mentions to normal level within two days. Trading volume and cumulative returns are closely correlated to abnormal social media attention: retail volume drops when

²WallStreetBets witnessed a nearly 400 growth during early 2021 following the GameStop event as shown in Appendix Figure A1.

abnormal attention drops, and remains high if abnormal attention continues. Cumulative returns and price informativeness continue to grow only when abnormal attention remains.

Is the abnormal social media attention informative for retail trades? Earnings announcement attracting abnormal Reddit attention allows us to assess whether sentiment of abnormal social media attention and retail trades are in the direction of incorporating information related to firm fundamentals. The interaction between earnings surprise and abnormal Reddit attention is the key focus of our analysis. We find that the tone of social media attention is more bullish and retail buying trades are higher for more positive earnings surprise occurred under abnormal social media attention. This result shows that abnormal social media attention disseminates information related to firm fundamentals. Retail investors trade on those earnings announcements that attracted abnormal high social media attention in the direction of the earnings surprise. More positive earnings surprise under abnormal social media attention also correlates with higher price informativeness, which further supports the role of abnormal social media attention in helping diffuse information to retail investors and facilitate price discovery.

While abnormal social media attention is informative for the general retail population, we found opposite result from the Robinhood sample. The interaction of Robinhood users' holding and earnings surprise of the stock is negative, both the day before, on and after the earning announcement. Robinhood users change their holdings on the stock opposite to the direction of the earnings surprise. This finding is consistent with evidence documented by previous literature supporting the argument that Robinhood users are relatively inexperienced (Barber, Huang, Odean, and Schwarz, 2021; Eaton, Green, Roseman, and Wu, 2022).

Our paper is closely related to the literature examining the importance of social media on financial markets. Both using information from the online platform Seeking Alpha, Chen, De, Hu, and Hwang (2014) and Farrell, Green, Jame, and Markov (2022) show that opinions posted on social media platform predict future return and earnings surprise, increasing retail trade informativeness. Jame, Johnston, Markov, and Wolfe (2016) employ the crowdsourced earning forecasts from Estimote and find more benefits from the forecasts with larger crowd size. While the studies mentioned above support the positive role social media plays in the financial market, Cookson, Engelberg, and Mullins (2022) demonstrate the selection bias of users in StockTwits which leads to lower ex-post returns.

There are several contemporaneous studies related to Reddit WallStreetBets, most of which focus on the GameStop short squeeze event (Long, Lucey, and Yarovaya, 2021; Allen, Haas, Nowak, Pirovano, and Tengulov, 2021; Aharon, Kizys, Umar, and Zaremba, 2021; Lyócsa, Baumöhl, and Vÿrost, 2022; Semenova and Winkler, 2021; Gârleanu, Panageas, and Zheng, 2021). Closest to our paper, the work of Hu, Jones, Zhang, and Zhang (2021) and Bradley, Hanousek Jr, Jame, and Xiao (2021) explores the importance of the Reddit WallStreetBets forum for the financial market not limited to the GameStop event. While Hu, Jones, Zhang, and Zhang (2021) study how Reddit activities are connected to retail trading flows and shorting flows, we examine the informativeness of abnormal social media attention for retail investors, and pay a special attention to abnormal social media attention events which occurred together with firm information events. On the other hand, different from the paper by Bradley, Hanousek Jr, Jame, and Xiao (2021) which focuses on investment research reports from WallStreetBets, our paper explores how general social media attention affects retail trading and its market consequences.

Our paper is also closely related to the literature examining the role of retail investors

in the financial market. Evidence from the literature is mixed regarding the impact of retail investor trading activities. Barber and Odean (2008) and Barber, Odean, and Zhu (2009) find that retail investors trades are attention-driven, highly correlated and persistent, which suggests that they are subject to biases. On the other hand, Kaniel, Saar, and Titman (2008) find positive excess return following intense retail buying and negative excess returns following retail selling. Kelley and Tetlock (2013) find retail order flows are able to predict firm news. The mixed evidence from the literature suggests a possibly changing composition of retail investors and the evolving relative importance of different retail investor groups. The rise of social media which facilitates retail investors' information collection and communication could further shape the role of retail investors in the stock market. We add to this literature by studying the interaction between social media attention and retail trading activities.

Finally, our paper also contributes to the literature studying the role of Robinhood users. Ozik, Sadka, and Shen (2021) show that there is a sharp increase in retail trading activities following the COVID-19 pandemic lockdown in early 2020, suggesting the rising impact of retail investors in the stock market. Welch (2022) also documents that Robinhood users increased their holdings in the bear market after the pandemic lockdown and displayed consensus in the portfolio they built. Barber, Huang, Odean, and Schwarz (2021) and Eaton, Green, Roseman, and Wu (2022) provide further evidence on the characteristics of Robinhood investors by employing platform outages as natural experiments. They show that Robinhood investors are relatively inexperienced and engage in more attention-induced trades than other retail investors. Our study employs trading data from Robinhood users, and arrive at consistent findings with the previous literature. The different performance of general retail investor sample and Robinhood sample supports the argument that Robin-

hood users are less sophisticated than other retail investors.

The paper is organized as follows. Section 2 introduces the data sources. Section 3 presents the relationship between social media attention, retail trading and market consequences. Section 4 examines the informativeness of social media attention. Section 5 concludes.

2 Data and Descriptive Statistics

2.1 Reddit WallStreetBets

Reddit is a social media platform organized as a collection of subreddits, where posts of a common theme are gathered together. Our data is collected from WallStreetBets, a subreddit that centered around stock and option trading. Users share their opinions on stocks through post submissions or comments. Created on January 2012, the WallStreetBets community is famous for its aggressive trading strategies and juvenile nature with a large presence of young retail investors. In the GameStop short squeeze event, users on WallStreetBets pushed the valuation of GameStop to almost 30 times compared to beginning of the month, despite the participation of hedge funds.

Number of subscribers to WallStreetBets was around 2 million by December 2020, while a big jump around 400 percent occurred during early 2021. By October 2021, there were above 10 million subscribers to WallStreetBets. The period of rapid increase in subscribers echoed the happening of the GameStop short squeeze event. The number and growth of subscribers to WallStreetBets are shown in the Appendix Figure [A1](#) and Figure [A2](#). Our sample period is from January 2018 to December 2020, starting from 2018 when the subscribers to

WallStreetBets reached a sufficiently large pool of 200,000 members, and before the period of significant increase in subscribers in 2021. To abstract from the possible change in user base and content composition from the sudden inflow of subscribers, we focus our analysis on the period before the GameStop event.

We collect submissions and comments from WallStreetBets during the sample period using the Pushshift API. The Pushshift API keeps a copy of Reddit object. The objects are copied to Pushshift at the point of time they were created. Subsequent edits to the objects will not be updated into Pushshift, therefore we are collecting the original copies of submissions and comments when they were generated. Our search with the Pushshift API relies on two primary endpoints one for submission search, and the other for comment search. Submission search endpoint will collect submissions on Reddit which include the query word in either their title or body during a specified time range. Similarly comment search endpoint will collect comments to Reddit submissions, which include the query words and are generated during a specified time range. The daily count on number of comments and submissions starts from 4pm, ET the day before until 4pm, ET that day, incorporating all the information that could be reflected in trading by market closure that day.

The stock is included in our sample if it satisfies the following two criteria: 1) it has available trading data in NYSE Trade and Quote (TAQ) database; 2) it has share code of 10 or 11 in The Center for Research in Security Prices (CRSP) database. A stock can be mentioned on WallStreetBets by either stock ticker or company name, e.g. ‘Tesla’ or ‘TSLA’. The search for any stock includes both the stock ticker search³ and the company

³Companies with stock ticker being single alphabet cannot be searched using stock ticker as query word, e.g. stock ticker ‘A’. Also, companies with stock ticker being common English word cannot be search using stock ticker as query word, e.g. stock ticker ‘ON’. We manually checked and removed stocks with tickers being common words. In total, there are 247 stocks which cannot be searched using stock ticker as query word.

name search. We conduct the search separately for comments and submissions. The total number of mentions of any stock is the sum of number of mentions in comments and in submissions. Further, we measure the tone of posts by identifying bullish words including “buy”, “long”, “call”, etc; and bearish words including “sell”, “short”, “put”, etc. We count the number of bullish and bearish words in all submissions and comments of a stock, and use bullish words counts minus bearish words counts as a proxy for Reddit tone.

In Table 1, we present the summary statistics of stock mentions in Reddit WallStreetBets on abnormal attention days. Abnormal attention day is defined as the starting day of the event when the number of total mentions of a stock exceeds 150% of its previous 6-month mentions average for at least two days. The stocks are categorized into two groups depending on retail concentration: a group with retail trading share below median, and a group with retail trading share above median.⁴ On the abnormal attention day, stocks in the low retail concentration group are mentioned 5.4 times on average in Reddit, while the stocks of the high retail concentration group are mentioned 28.6 times on average in Reddit. The mention statistics on all other days excluding abnormal attention days are lower.

We also present summary statistics of stock mentions on Reddit WallStreetBets for earnings announcement days which occurred in conjunction with abnormal Reddit attention in Appendix Table A2. The summary statistics is presented for the whole sample period, earnings announcement days, and days with joint occurrence of abnormal Reddit attention and earnings announcement. In days with earnings announcement, Reddit attention is about three times as high as other days. When earnings announcement occurs with abnormal Reddit attention, the average total mention is about ten times that of other days.

⁴See section 2.2 for the BJZZ algorithm in identifying retail trades from the TAQ database.

Table 1: **Summary Statistics (Abnormal Attention Event).**

This table presents summary statistics for the whole sample and abnormal attention events. Wallstreetbets mentions are the sum of mentions in submissions and in comments. Retail trading data is obtained using the BJZZ algorithm from TAQ. Robinhood users holding is obtained from robintrack website. Excess Return is measured using closing price and subtracts risk free rate. Price informativeness is measured using $1 - R^2$ by regressing return on Fama-French 3 factors. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left panel presents statistics for the group of stocks which has below median retail investor concentration during the sample period. Right panel presents statistics for the group which has above median retail investor concentration. The upper panel reports the statistics for the entire sample period. The middle panel reports statistics for the day which the abnormal mention event occurred. The lower panel reports statistics for the entire sample period excluding the abnormal mention day.

	Group: Low Retail Concentration						Group: High Retail Concentration					
	obs	mean	st.dev.	median	Q1	Q3	obs	mean	st.dev.	median	Q1	Q3
Whole sample												
total WSB mentions	478,049	1.0	9.5	0	0	0	429,406	5.9	64.5	0	0	1
net retail buy volume	478,049	-196.9	22346.4	28	-2751	2,843	429,406	7029.9	116500.9	-5781	132	7,605
Robinhood holdings	354,052	1086.3	3584.5	352	141	892	320,150	10622.4	38144.7	1,519	451	5,063
return (percent)	478,049	0.08	2.70	0.08	-1.07	1.20	429,406	0.16	4.32	-0.01	-1.71	1.71
price informativeness	478,049	0.51	0.24	0.50	0.32	0.69	429,406	0.61	0.23	0.63	0.44	0.80
Abnormal mention day												
total WSB mentions	7,069	5.4	16.6	2	1	4	7,240	28.6	157.0	3	1	11
net retail buy volume	7,069	630.7	22924.7	111	-2795	3,404	7,240	15226.3	136113.2	1,051	-6444	14,825
Robinhood holdings	5,174	1910.3	5164.9	598	196	1,729	5,182	19006.0	50758.9	4,132	1,289	11,802
return (percent)	7,069	0.24	4.06	0.18	-1.31	1.61	7,240	2.12	8.16	0.63	-1.36	3.36
price informativeness	7,069	0.48	0.24	0.28	0.47	0.68	7,240	0.57	0.24	0.58	0.40	0.76
Other days												
total WSB mentions	475,401	1.0	9.5	0	0	0	426,217	5.7	63.3	0	0	1
net retail buy volume	475,401	-202.2	22326.2	28	-2751	2,839	426,217	6922.8	116219.1	-5774	125	7,536
Robinhood holdings	338,878	1076.8	3561.1	350	140	886	314,968	10500.0	37915.8	1,498	444	4,988
return (percent)	475,401	0.09	2.69	0.08	-1.07	1.20	426,217	0.14	4.27	-0.01	-1.72	1.70
price informativeness	475,401	0.51	0.24	0.51	0.32	0.70	426,217	0.60	0.23	0.63	0.44	0.79

2.2 Retail Trades

The data on retail trading is constructed from the TAQ database using the algorithm following Boehmer, Jones, Zhang, and Zhang (2021) (BJZZ). Due to Regulation NMS and the resulting institutional arrangements, retail order flows typically receive price improvements in fractions of a penny and are reported to a FINRA trade reporting facility (TRF). The BJZZ algorithm relies on this observation to identify marketable retail orders. Following the algorithm, we keep only trades that are reported to a TRF (with exchange code “D” in TAQ Consolidated Trades), and merge the trade data with CRSP stock data. We include the common stocks with share code 10 or 11 listed on the NYSE, NYSE American (formerly AMEX), and NASDAQ. Trades with fractional penny prices between \$0.006 and \$0.01 are identified as retail buys; trades with fractional penny prices between \$0.00 and \$0.004 are identified as retail sells. Our measure of retail concentration of stocks uses the ratio of the average retail buying volume over the average total buying volume during the sample period.⁵

One potential concern with the BJZZ algorithm is that it identifies only marketable retail orders. The majority of retail trading activities are picked up, though, as limit orders are less informed according to Kelley and Tetlock (2013). We also collect trading data from Robinhood, a retail trading platform which has gained popularity over the past several years. The Robinhood trading data are obtained from the Robintrack website (<https://robintrack.net/>), which scrapes stock data from Robinhood. Robintrack tracks the number of users holding a particular stock between May 2, 2018, and August 13, 2020.⁶

⁵We also use total retail trading volume (including buying and selling) for robustness checks, and the results are similar.

⁶Robinhood discontinued the API Robintrack used to collect data on August 13, 2020.

Robintrack data does not identify individual buy or sell trades. The change in user positions in a particular stock represents the aggregate net increase of users holding the stock in Robinhood.

The summary statistics on retail trading is provided in Table 1 for abnormal attention events and Appendix Table A2 for abnormal attention events occurred together with earnings announcement. Retail trading activities for the general retail population obtained from TAQ and Robinhood users holdings are both presented. We divide the stocks into the low retail concentration group and the high retail concentration group. For the low retail concentration group, on average there are negative 196 shares of net retail buying volume and 1086 Robinhood users holding the stock for a typical day. The average net retail buying volume and Robinhood users holding for the high retail concentration group are much larger, with 7029 shares of net retail buying volume and 10622 users holding the stock. Probing into the statistics on abnormal attention days, retail buying volumes are significantly larger compared to other days. For the high retail concentration group, the average net retail buy volume on abnormal attention days is 15226 number of shares. Moreover, on earnings announcement days, retail trades are twice as many as normal days and Robinhood users holding is close to normal days.

2.3 Stock Data

We collect daily stock returns data from CRSP. The summary statistics on the excess return across stock-day observations is provided in Table 1. The average excess return of the whole sample period is 0.08 percent for the low retail concentration group and 0.16 percent for the high retail concentration group. On the abnormal attention days, the excess return

is significantly higher for both groups. The excess return is 0.24 percent for the low retail concentration group, and 2.12 percent for the high retail concentration group on the abnormal attention days. Based on the daily return data, we use one-month rolling window regressions to obtain daily stock betas to construct beta control group for each stock. The risk-free rate and market return data are obtained from Kenneth French’s website (<https://mba.tuck.dartmouth.edu/pages/faculty/ken.french>).

To construct measures of price informativeness, we use Fama and French (1992) three factors (equity premium, return difference between small and large stocks, and return difference between high and low book-to-market stocks) data from Kenneth French’s website. Our measure of price informativeness identifies price nonsynchronicity or the firm-specific variation in prices. Following Morck, Yeung, and Yu (2000), we measure the amount of information incorporated into prices using $1 - R^2$, where R^2 is obtained by regressing daily stock excess return on Fama and French (1992) three factors. A higher value of $1 - R^2$ indicates that more firm specific information is incorporated into prices, or higher price informativeness. The summary statistics is presented in Table 1. For the low retail concentration group, the mean price informativeness measure is 0.51 during the whole sample period, while that for the high retail concentration group is 0.61.

Firm controls used in regressions include market capitalization (number of shares outstanding times price) and book-to-market ratio (book equity divided by market equity) calculated using data obtained from CRSP and COMPUSTAT. We also include industry control at the 4-digit SIC level.

Firm earnings announcement dates are obtained from COMPUSTAT (data item RDQE). Using data from I/B/E/S and COMPUSTAT, we construct measures on earnings surprises

at the earnings announcements. We use the analyst forecast errors (AFE), defined as actual quarterly earnings per share (EPS) less the median analyst forecast of EPS. In Appendix Table A1, we provide a summary table for all the variable sources and definitions.

3 Social Media Attention, Retail Trades and Market Consequences

In this section, we document the market consequences and retail trading patterns when social media attention is high on Reddit. We first present the return drifts and increase in price informativeness around heated social media discussions. Then we document the retail trading activities when social media attention is high, which could be a driving force behind the large excess returns.

3.1 Market Consequences Around Social Media Attention

We first examine the excess returns patterns around heated discussions on Reddit using event study methodology. For stocks with high proportions in retail trading, the impact of social media attention could be potentially larger compared to stocks with low proportions in retail trading. Therefore, we separate the stocks into a group with below median retail trading proportions and a group with above median retail trading proportions. The event study is conducted for the two groups separately.

In the event study, we define a day as “abnormal attention event” if it marks the beginning of an event when the total number of mentions of a stock on Reddit exceeds 150%

of its previous 6-month mentions average for at least two days. The onset of the abnormal attention is event day 1 in our analysis. Event day 0 is the last day when social media attention is below the abnormal threshold. Note that for each stock, it is possible that it has multiple abnormal mention events. For stocks with high proportions in retail trading, the attention from Reddit is higher compared to stocks with low proportions in retail trading as shown in Appendix Figure A3. For each abnormal mention event, we calculate the excess cumulative return for the stock involved over a $[-5, 5]$ window around the event to examine the return patterns around high social media attention. The cumulative return in the $[t_0, t_1]$ window for an event-stock i is

$$CR_i(t_0, t_1) = \prod_{t_0}^{t_1} R_{i,t}, \quad (1)$$

where $R_{i,t}$ is the gross return of stock i on date t . We also benchmark all the returns against a control group of stocks. Specifically, we sort all stocks in the sample into percentiles based on their stock betas computed using one-month rolling windows. We use a stock's beta group as its control group to control for the systematic risk of stocks. The excess cumulative returns for stocks are then

$$ECR_i(t_0, t_1) = CR_i(t_0, t_1) - \sum_{Control} \omega_{Control} CR_{Control}(t_0, t_1), \quad (2)$$

we take the weights $\omega_{Control}$ as equal for all stocks in the control group. The mean ECR for each group is the equal-weighted average of the individual stock CRs in the group benchmarked against the mean CRs for all control stocks.

In Figure 1, we plot the ECR over a $[-5, 5]$ window around the abnormal social media event of low retail concentration group in Panel A; and the ECR over the same window of high retail concentration group are plotted in Panel B. In both panels, we exclude the

top and bottom 1% of the returns to address the concern of the sensitivity of our results to outliers.

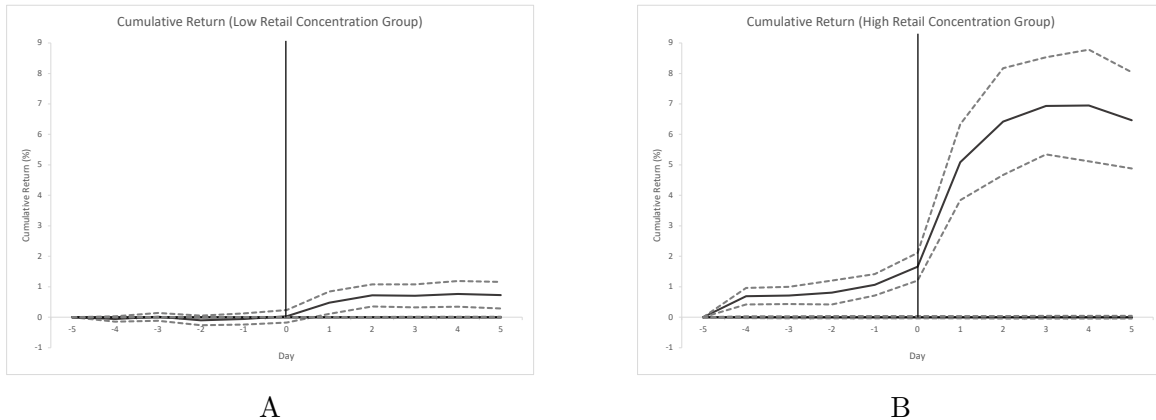


Figure 1: **Average Excess Cumulative Return (Abnormal Attention Event).**

This figure plots the average of excess cumulative return of stocks on a 11-day window. The excess return is measured controlling the average of the stock’s respective beta group. The cumulative return is calculated relative to stock price five days prior to the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval for the average return. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average cumulative return for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

As seen in the charts in Figure 1, the excess returns in both groups display upward drifts following the abnormal attention event. For the low retail concentration group, the excess cumulative returns are around zero before the event, but the returns climb to 1 percent after the event and do not go down for at least 5 days. For the high retail concentration group, the excess cumulative returns are already rising to above 1 percent before the abnormal attention event. Starting from event-day 1 after the abnormal retail mentions, the excess cumulative returns show a strong upward drift and continue to increase until 3 days after. The peak of cumulative excess returns is around 7 percent at event-day 3, while it decreases slightly to

6 percent at event-day 5 following the high social media discussions. As evidenced by the pointwise 95% confidence interval for the mean return (the dashed grey lines), the excess cumulative returns around the abnormal attention event for the high retail concentration stocks are significantly different from zero.

To put the economic magnitude of the returns drift in perspective, we also calculate the average excess cumulative returns on all other 11-day windows in the sample period excluding windows centered around the abnormal mention days. The results are plotted in the solid grey lines in Figure 1. The black dashed lines represents the pointwise 95% confidence bands.⁷ As seen in both panels, the average excess cumulative returns are essentially zero on other days.

To more formally assess the magnitudes of excess stock returns around heated social media discussions, we conduct the simple dummy variable regression model over a series of return windows:

$$ECR_t = \beta_0 + \beta_1 \mathbf{1}_t(\text{Reddit_abnormal_mention}) + \beta_c \text{Controls}_t + FE_s + \epsilon_t, \quad (3)$$

where ECR_t is the average excess cumulative return for all stocks in the group on event-day t . The explanatory variable is a dummy variable that is equal to one on specified windows around Reddit abnormal mention days and zero otherwise. For firm controls, we include log book-to-market ratio and log market capitalization. We also include industry fixed effect control at the 4-digit SIC level, as well as time fixed effect control at the monthly level. In the regression, β_1 is the mean excess return differential on Reddit abnormal attention windows versus other days. The constant β_0 measures the unconditional mean excess returns earned

⁷We use Newey-West standard errors when computing confidence intervals to account for the overlapping of days in the windows of all other days.

on all time periods outside of the Reddit abnormal attention windows.

Table 2 and 3 report coefficient estimates for β_1 over different return windows. Table 2 reports three windows for cumulative returns until the event day: market close on one day before the event ($t - 1$) to market close on the event day, three days before the event ($t - 3$) to market close on the event day, and five days before the event ($t - 5$) to market close on the event day. Table 3 reports three windows for cumulative returns until several days after the event: market close on one day before the event ($t - 1$) to market close on day one after the event ($t + 1$), three days before the event ($t - 3$) to market close on three days after the event ($t + 3$), and five days before the event ($t - 5$) to market close on five days after the event ($t + 5$).

In both tables, the left panel reports estimates for the low retail concentration group, and the right panel reports estimates for the high retail concentration group. For the high retail concentration group, it is notable that β_1 is significant for all six windows. In Table 2, the excess cumulative return from five days before the abnormal attention event is on average 1.431 percent higher than other periods outside of the abnormal attention event. The coefficients for the regressions for three-day window before the event and one-day window before the event are smaller, though both are still significant. The coefficients are larger when the windows expand to several days after the event in Table 3. For the low retail concentration group, the excess returns are smaller compared to the high retail concentration group and β_1 is significant only for three-day and one-day window before the event.

After showing the stock returns pattern around the general abnormal social media attention event, we proceed to examine price informativeness under abnormal social media attention. Price informativeness measures how much firm specific information is incorpo-

Table 2: **Cumulative Return Regression Results (Onset of Abnormal Attention).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock’s respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal mention day dummy for each stock. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close
Dummy	0.092*** (0.038)	0.138*** (0.067)	0.111 (0.078)	0.615*** (0.066)	1.028*** (0.116)	1.431*** (0.018)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

rated into price and is calculated from one-month rolling windows including the current trading day. Detailed description of the measure is presented in Section 2.3. We benchmark the measure against a control group of stocks sorted by stock betas. We conduct the event study regressions as described in equation (4). The onset of the abnormal attention is event day 1 in our analysis. Regression results are presented in Table 4. Price informativeness is significantly higher from the start of abnormal social media attention and continues to grow in the five-day window after the start of the abnormal attention. The magnitude of increase is close for the low and high retail concentration group, leveling around 0.03 in five days since the start of the abnormal attention event. This increase in magnitude amounts to around 5% of average price informativeness in normal days without abnormal social media attention.

Table 3: **Cumulative Return Regression Results (After Abnormal Attention).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock’s respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal day dummy. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)
Dummy	0.094* (0.053)	0.254*** (0.098)	0.235* (0.129)	0.877*** (0.093)	1.585*** (0.175)	2.158*** (0.242)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

A potential concern is that the excess return patterns might be mixed with the return drifts around important announcement events. Ball and Brown (1968), Kishore, Brandt, Santa-Clara, and Venkatachalam (2008) and others have documented post firm earnings announcements drifts. Hence, as a robustness check, we disentangle the return pattern around heated social media discussions from the firms earnings announcement day effects. Specifically, we remove any stock if its abnormal mention day on Reddit is within 5-day windows around earnings announcement days. The event study figure excluding firms earnings announcements and the corresponding regression results are presented in the Internet Appendix. See Figure IA1 for the cumulative excess return in a $[-5, 5]$ window around the abnormal mention days. The regressions for six windows around the abnormal mention days are shown in Table IA1 and IA2. Still, the significantly large excess return around heated

Table 4: **Price Informativeness Regression Results (Abnormal Attention).**

This table presents regression results for event study regression results of price informativeness around the abnormal mention event. Price informativeness is measured at the excess level by controlling the average value of the stock's beta group sorted in decile. The left column presents result for stocks with below median retail trade concentration. The right column presents results for for stocks with above median retail trade concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail	High Retail
	Price Informativeness	Price Informativeness
day-5	0.00130 (0.00375)	-0.00483 (0.00346)
day-4	0.00197 (0.00375)	-0.00350 (0.00346)
day-3	-0.000768 (0.00375)	-0.00123 (0.00346)
day-2	-0.00235 (0.00375)	-0.00162 (0.00346)
day-1	-0.00184 (0.00375)	-0.000988 (0.00346)
day0	-0.000939 (0.00375)	-0.000379 (0.00346)
day1	0.00777** (0.00375)	0.0117*** (0.00346)
day2	0.0140*** (0.00375)	0.0182*** (0.00346)
day3	0.0222*** (0.00375)	0.0261*** (0.00346)
day4	0.0255*** (0.00375)	0.0281*** (0.00346)
day5	0.0274*** (0.00375)	0.0315*** (0.00346)
Firm Controls	Y	Y
Industry FE	Y	Y
Time FE	Y	Y
Obs.	1,030,136	951,468

social mediadiscussions survives after excluding stocks with abnormal Reddit mention days close to earnings announcements.

3.2 Retail Trading Around Social Media Attention

To examine the driving force behind the return drifts when retail attention is high, we turn our focus to the trading activities of retail investors in this section. Retail tradings are identified by the BJZZ algorithm discussed in Section 2.2 from the TAQ database. We investigate retail net buying volume around abnormal mention event on Reddit as defined in Section 3.1.

Denote the net retail buying volume for stock i at time t as $NB_{i,t}$. We have

$$NB_{i,t} = BV_{i,t} - SV_{i,t}, \quad (4)$$

where $BV_{i,t}$ is the buying volume of retail trades for stock i at date t and $SV_{i,t}$ is the selling volume. As in Section 3.1, we benchmark the trading measure against a stock's beta group to control for the systematic risk of stocks. The excess retail net buying volume for stock i at date t is

$$ENB_{i,t} = NB_{i,t} - \sum_{Control} \omega_{Control} NB_{Control,t}. \quad (5)$$

And again, we take the weights $\omega_{Control}$ as equal for all stocks in the control group. The mean ENB for each group is the equal-weighted average of the individual stock-event NBs in the group benchmarked against the mean NBs for all control stocks.

In Figure 2, we plot the mean ENB for the low retail concentration group in Panel A; and the mean ENB for the high retail concentration group is plotted in Panel B. The solid

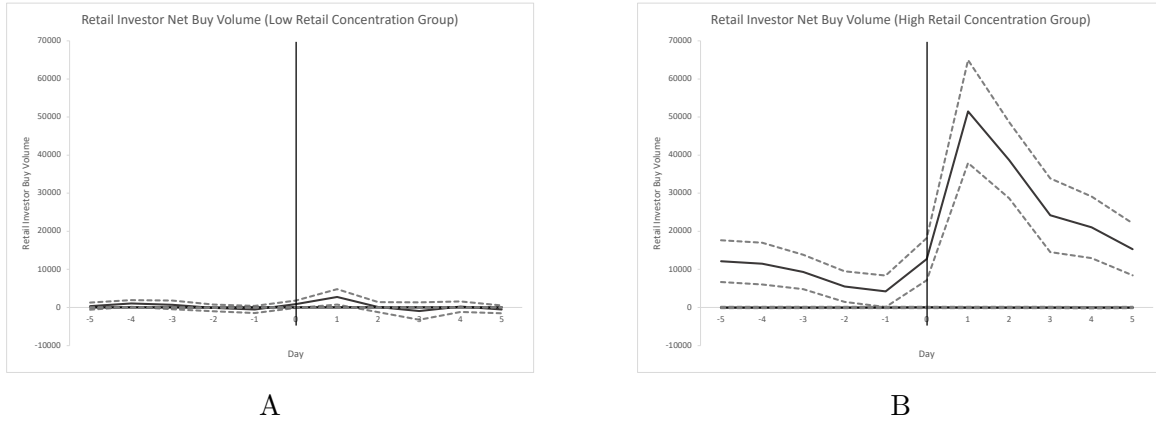


Figure 2: **Average Excess Retail Net Buy Volume (Abnormal Attention Event).** This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess volume is measured controlling the average of the stock’s respective beta group. The vertical line marks the day before the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from TAQ following the BJZZ algorithm. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

grey lines in both panels show the mean ENB for all other windows excluding the abnormal attention event, which are essentially zero for both stock groups. As shown in Panel A, the low retail concentration group doesn’t see much difference in the excess retail net buying volume around the abnormal attention events on Reddit compared to other windows. For the high retail concentration group, the increase in the excess retail net buying volume is very significant. The retail net buying volume is around 10,000 shares higher than normal times before the abnormal attention event, and it rises sharply to a peak of around 50,000 shares higher on event-day 1, which is the starting day of the abnormal attention. The excess net buying volume decreases steadily after event-day 1, though it is still significantly higher than normal times for at least 5 days after the event.

We also use Robinhood users holdings data to proxy for the aggregate buying of the stock within Robinhood. Note that users holding is different from net buying volume as we don't have data on how many shares each user holds. We conduct the same event study on the Robinhood users holding data, and the excess holding patterns around the abnormal attention events on Reddit are shown in the Figure 3. Although excess Robinhood users holdings are higher around the abnormal attention events on Reddit compared to normal times, there is no apparent drift in trading on or following the event day. Excess holdings for the high retail concentration group is larger in scale than the low retail concentration group. These suggest that while in general Robinhood users prefer stocks that are popular on social media platform, not many Robinhood users adjust their holdings of stocks following discussions on Reddit.

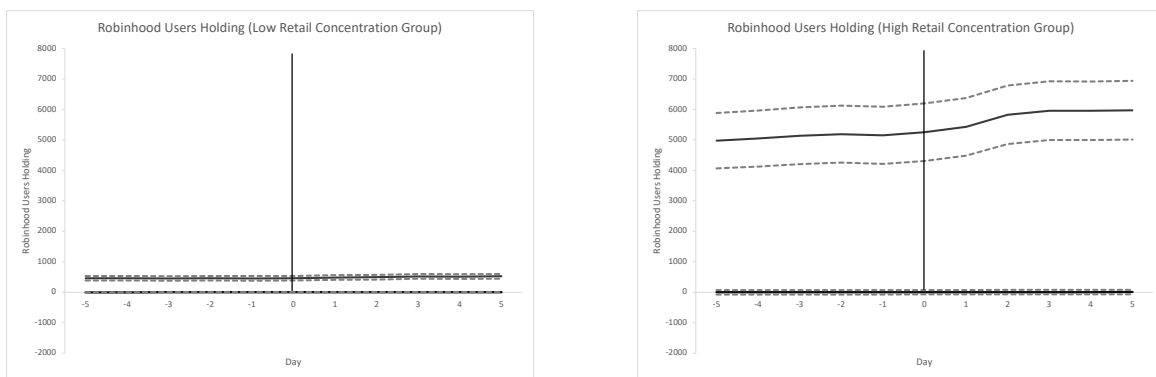


Figure 3: **Average Excess Robinhood Users Holding (Abnormal Attention Event).** This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess holding is measured controlling the average of the stock's respective beta group. The vertical line marks the day before the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from Robintrack. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

To quantitatively examine the trading activities around the abnormal attention events on Reddit, we conduct a series of event study regressions for the general retail population and the Robinhood investors respectively:

$$Y_{i,t} = \beta_0 + \sum_{j=-5}^5 \beta_j \cdot \mathbf{1}(j=t) + \beta_c \text{Controls}_{i,t} + FE_s + \epsilon_{i,t}, \quad (6)$$

where $Y_{i,t}$ is excess net buying volume for the general TAQ retail trading data; for the Robinhood investors data, $Y_{i,t}$ is excess Robinhood users holding. We include a full set of event-time dummies $\mathbf{1}(j=t)$, $j = \{-5, \dots, 5\}$. Note that β_0 is the unconditional mean for the dependent variable on all windows excluding the abnormal mention events, and β_j measures the difference on event-day j during the abnormal mention events. Firm controls and fixed effects controls are included. The results are shown in Table 5.

In Table 5, the left panel shows the results for the low retail concentration group, and the right panel shows the results for the high retail concentration group. For retail net buying volume, the high concentration group displays significant abnormal volume from the start of abnormal attention to five days after the abnormal attention, though the scale of excess volume is steadily decreasing from event-day 1. The excess retail net buying volume on event-day 1 is 45,451 shares for the high retail concentration group, while on event-day 5 it decreases to 10,873 shares. The low retail concentration group does not see any stable trading patterns around the abnormal attention events. For Robinhood investors, the aggregate holdings are significantly higher around the abnormal attention events for both group, but there is no upward drift in holdings following the event in the window that we examine. Also, the excess holdings for the high retail concentration group is much larger in scale compared to the low retail concentration group.

Table 5: **Retail Trading Activity Regression Results (Abnormal Attention Event).**

This table presents regression results for event study regression results of retail trading activity around the abnormal mention event. All dependent variable are measured at the excess level by controlling the average value of the stock's beta group sorted in decile. The first column uses net retail buying volume as dependent variable. The second column uses robinhood users holding as the dependent variable. The left panel reports results for stocks with below median retail trade concentration. The right panel reports regression results for stocks with above median retail trade concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration		High Retail Concentration	
	Retail Net Buy Volume	Robinhood holding	Retail Net Buy Volume	Robinhood holding
day-5	186.9 (327.7)	380.4*** (38.37)	8507.9*** (1798.8)	3360.3*** (565.1)
day-4	853.2*** (327.8)	388.9*** (38.40)	6942.7*** (1799.5)	3566.0*** (564.8)
day-3	1101.5*** (327.5)	408.5*** (38.36)	5960.8*** (1799.8)	3751.0*** (566.7)
day-2	48.33 (327.7)	420.8*** (38.37)	2542.9 (1797.9)	3678.1*** (566.2)
day-1	-55.53 (327.3)	428.8*** (38.44)	638.0 (1796.1)	3666.2*** (566.2)
day0	903.9*** (327.4)	436.0*** (38.51)	9053.0*** (1797.9)	3850.3*** (567.2)
day1	1706.4*** (328.0)	469.2*** (38.51)	45451.3*** (1798.8)	4224.6*** (567.2)
day2	522.1 (327.7)	498.4*** (38.61)	37066.4*** (1799.5)	4819.2*** (568.0)
day3	-1047.9*** (327.5)	522.0*** (38.66)	22126.6*** (1798.8)	4910.3*** (568.5)
day4	394.0 (327.5)	521.5*** (38.59)	19136.7*** (1797.8)	4909.4*** (568.5)
day5	-111.0 (327.5)	521.5*** (38.66)	10873.4*** (1798.4)	4941.3*** (568.4)
Firm Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Obs.	1,030,141	707,606	951,508	610,061

To rule out the concern that results in this section are driven by retail trading activities around earnings announcements which attracted abnormal social media attention, we conduct robustness checks of the retail trading results removing stocks with abnormal attention days within 5 days around the earnings announcement days. The results are presented in the Internet Appendix. Figure IA2 shows the event study graph for the excess retail buying volume, and Figure IA3 shows the event study graph for the excess Robinhood users holdings. The results of the event study regressions for retail trading activities are shown in Table IA4. Still, while the high retail concentration group displays a sharp increase in excess retail buying volume following the abnormal attention day, while the Robinhood users holdings stay at a roughly constant level which is significant higher than normal times around the abnormal attention day. The regression results confirms these observations. The robustness checks shows that the findings in this section are not driven by firms earnings announcements.

3.3 Continued versus Dropped Attention

In this section, we provide further evidence on the relationship between social media attention and retail trading activities, and its market consequences. We spilt the abnormal attention events based on the persistence of high mentioning and examine the pattern differences. Specifically, we investigate two groups: a group with abnormal attentions that dropped in 2 days, and another group with abnormal attentions that remained high for at least 5 days.

For the excess cumulative returns around the abnormal attention events, we conduct the dummy variable regression in Equation (3) over the same set of windows as in Section 3.1. The results for the group with mentions dropped in 2 days are shown in Table 6, and the results for the group with mentions remained at least 5 days are shown in Table 7. The

coefficients measured after the abnormal attention event for the continued mention events are almost twice as large as measured at the beginning of the abnormal attention for the high retail concentration group. This suggests that for stocks that has more persistent mentions on Reddit, the cumulative return increase is larger. The excess cumulative returns regression results for the time window measured at the onset of the abnormal attention are included in Appendix Table A3 and Table A4. Comparing results from the two event groups, the abnormal excess cumulative returns measured at the onset or after the abnormal attention are close in magnitude for the events with dropped attention.

Table 6: **Cumulative Return Regression Results - After Abnormal Attention (Attention Dropped in 2 days).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal day dummy for which mention dropped in 2 days. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)
Dummy	0.157*** (0.053)	0.136 (0.094)	0.322** (0.125)	0.660*** (0.106)	0.650*** (0.196)	0.955*** (0.274)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

Regression results for the two types of abnormal attention events using price informativeness as dependent variable are presented in Appendix Table A5. Comparing the coefficients

Table 7: **Cumulative Return Regression Results - After Abnormal Attention (Attention Remained at least 5 days).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal day dummy for which mention remained high for at least 5 days. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)
Dummy	0.123 (0.108)	0.220 (0.197)	0.437* (0.259)	0.720*** (0.161)	1.746*** (0.302)	2.550*** (0.418)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

for the two types of events we can observe that while price informativeness remains relatively constant in a dropped attention event, it continues to grow in an abnormal attention event with persistent high attention.

We next look at the retail trading activities around the abnormal attention events. We conduct the event study regressions in Equation (6), where $Y_{i,t}$ is the excess net buying volume here. The regression analysis are presented for the two groups of events with dropped mention in 2 days and the group with continued mention in at least 5 days. The results are shown in Table 8. We focus on the high retail concentration stocks as the impact of Reddit attention is cleaner and more significant there. Note that while the dropped attention group only sees significant excess net retail buying volume on event-day 0, 1 and 2, the remained

Table 8: **Retail Trading Activity Regression Results (Attention Dropped or Remained High).**

This table presents regression results for event study regression results of retail trading activity around the abnormal mention event. Dependent variable is measured at the excess level by controlling the average value of the stock's beta group sorted in decile. Column 1 and 2 present results for stocks with below median retail trade concentration. Column 3 and 4 present results for for stocks with above median retail trade concentration. For each type of event, the left column reports results for abnormal mention event in which attention dropped in 2 days. The right column reports regression results abnormal mention event in which attention remained high for at least 5 days. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration		High Retail Concentration	
	Attention dropped	Attention Remain	Attention dropped	Attention Remain
day-5	138.6 (319.5)	-324.5 (651.1)	3049.8 (2044.3)	21845.3*** (3082.4)
day-4	440.2 (319.7)	1883.3*** (651.7)	-1224.6 (2042.8)	20538.4*** (3080.7)
day-3	-206.3 (319.3)	1686.9*** (651.1)	2522.8 (2040.9)	12441.8*** (3089.2)
day-2	-914.7*** (319.6)	1268.5* (651.7)	-782.4 (2040.9)	6879.4** (3080.7)
day-1	434.0 (319.8)	1194.4* (651.7)	947.4 (2040.9)	7291.1** (3077.3)
day0	688.9** (319.7)	1958.8*** (651.7)	4343.1** (2042.6)	22278.0*** (3085.8)
day1	2274.3*** (319.8)	2117.4*** (652.2)	25015.1*** (2045.9)	78677.4*** (3079.0)
day2	263.0 (319.6)	-0.161 (651.7)	3689.4* (2038.5)	62466.6*** (3087.5)
day3	-649.1** (319.3)	2207.7*** (651.1)	-1433.7 (2039.9)	52395.8*** (3084.1)
day4	-382.2 (319.4)	2986.4*** (651.1)	-3007.4 (2047.3)	54528.4*** (3085.8)
day5	82.17 (319.3)	-265.1 (651.7)	-1497.0 (2041.9)	40675.6*** (3085.7)
Firm Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Obs.	1,030,141	1,030,141	951,508	951,508

attention group has significant coefficients on all event days. The excess net retail buying volume peaks at event-day 1, with 78,677 shares above the volume on normal days. And the excess volume remains significant for at least 5 days after the event, averaging at 57,742 shares, which is about 8.1 times as high as mean net retail buying volume in other days without abnormal social media attention.

We plot the average excess retail net buying volume for an eleven-day window around the event day. Appendix Figure A4 shows the event study plot for the group with mention dropped in 2 days, and Appendix Figure A5 shows the event study plot for the group with mention remained high for at least 5 days. As shown in the figures, there is significant difference in the trading patterns between the two groups. Events with continued mentions sees more persistent retail trading drift, and the drift is larger compared to the dropped mention group. The analysis presented in this section lends further support to the role played by social media attention in affecting retail trading activities and shaping market consequences.

4 Information Diffusion Among Retail Investors

In this section, we examine the impact of Reddit attention on information diffusion among retail investors. News and earnings positively affects the diffusion of information among investors, as documented in previous literature (e.g. Peress (2014)). We further test whether social media attention on Reddit helps spread the information from earnings announcements to retail investors. In section 4.1, we study the information diffusion among the general retail investors; and we focus on Robinhood investors in Section 4.2.

4.1 The General Retail Investors

To utilize the information from firm news release, we examine firm earnings announcements which occur in conjunction with abnormal attention events on Reddit, and investigate whether price informativeness and other measures on how information is perceived are higher relative to earnings announcements that are not accompanied by abnormal social media attention.

As described in Section 2.3, we use earnings surprises measured by analyst forecast errors (AFE) to proxy the direction of information released from earnings announcements. If AFE of a firm is more positive, earnings announcement reveals good news about the firm, hence we should expect an investor who learns about the information to talk more positively about the firm on social media and buy more shares. Otherwise an investor should talk more negatively about the firm on social media and sell more shares at the mean time. Specifically, we use price nonsynchronicity to identify price informativeness as described in Section 2.3. Retail buying volume is identified from the TAQ database as in Section 2.2. And Reddit

tone is the net bullish tendency in Reddit posts including submissions and comments as described in Section 2.1. Observations are winsorized at top and bottom 1%.

To test whether discussions on Reddit indeed helps diffuse the information from earnings announcement, we estimate the following regression:

$$Y_{i,t} = \alpha_0 + \alpha_1 AFE_i + \alpha_2 \mathbf{1}_{i,t}(\text{Reddit mention}) + \alpha_3 AFE_i \cdot \mathbf{1}_{i,t}(\text{Reddit mention}) + \alpha_c \text{Controls}_{i,t} + FEs + \epsilon_{i,t}, \quad (7)$$

where $Y_{i,t}$ is taken as Reddit tone, number of retail buying trades and price informativeness measure respectively. For firm controls, we include log book-to-market ratio and log market capitalization. We also include industry fixed effects controlled at the 4-digit SIC level, as well as time fixed effects controlled at the monthly level. Our focus is on the incremental effect on information diffusion when earnings announcement is accompanied by Reddit abnormal attention, as represented by the coefficient α_3 of the multiplier term. Note that at normal earnings announcement dates, the change in the dependent variable when AFE changes percent is $\alpha_1 \Delta AFE$. At earning announcement dates accompanied by abnormal Reddit mentions, the change in the dependent variable when AFE changes is $\alpha_1 \Delta AFE + \alpha_3 \Delta AFE$. Hence α_3 measures the incremental effect of social media attention on information diffusion on earnings announcement events. The estimation results are shown in Table 9. For one standard deviation increase in AFE of the high retail concentration group under abnormal social media attention, price informativeness would increase 0.023, which amounts to 3.8% of average price informativeness in the high retail concentration group on normal days.

Table 9: **Earnings Announcement with Abnormal Attention (Day t).**

This table presents regression results exploring the interaction term between abnormal attention on Reddit and the earning surprise on earning announcement events. All dependent variable are measured at the excess level by controlling the average value of the stock's beta group sorted in decile. Dependent variable is measured at the same day as the independent variable. Column 1-3 present results for stocks with below median retail trade concentration. Column 4-6 present results for stocks with above median retail trade concentration. For each group, dependent variable of first column is reddit bullish tone. Second column is retail buy trade. Third column is price informativeness measured using $1 - R^2$ by regressing return on Fama-French 3 factors. Earnings surprise is measured using difference of actual EPS and median forecast from I/B/E/S. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11 and earning forecast from I/B/E/S. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	Reddit Tone	Retail Buy Trade	Price Informativeness	Reddit Tone	Retail Buy Trade	Price Informativeness
Abnormal	0.400*** (0.008)	102.4*** (1.391)	0.010*** (0.001)	2.967*** (0.060)	382.1*** (5.807)	0.019*** (0.001)
Forecast Error	0.015 (0.050)	14.58* (8.743)	0.0001 (0.005)	-0.007 (0.014)	-1.118 (1.335)	0.0002 (0.0002)
Abnormal*Forecast Error	0.207*** (0.079)	65.44*** (13.87)	0.027*** (0.009)	0.862** (0.372)	132.0*** (36.23)	0.017*** (0.005)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	434,523	434,523	434,523	358,029	358,029	358,029

As shown in Table 9, for both the low retail concentration group and the high retail concentration group, the coefficient α_3 is positive and significant for Reddit tone, number of retail buy trades and price informativeness. And the coefficients for Reddit tone and number of retail buy trades for the high retail concentration group are larger than that of the low retail concentration group. This suggests that on earnings announcements events, high Reddit attention and content of discussion help retail investors perceive the information from earnings surprises. Also, note that for the excess number of retail buy trades in the high concentration group, while the coefficient on AFE is insignificantly negative on normal earnings announcement dates, the coefficient on AFE on the earnings announcements dates in conjunction with abnormal attention events is positive and large. For one standard deviation increase in AFE of the high retail concentration group under abnormal social media attention, number of retail buy trades would increase by 494, which amounts to 55% of average retail buy trades in the high retail concentration group on normal days. Retail investors take correct trading directions when earnings announcements gained high attention on social media. The retail investors' discussion tone tilts more towards the right direction when earnings announcements occur with abnormal Reddit mention events, and the general informativeness of pricing also increases on these events.

We conduct the same test in Equation (7) on date $t + 1$ after the earnings announcement dates. The dependent variable is $Y_{i,t+1}$ in this case. The regression results are shown in Table 10. As seen in Table 10, the incremental effect when earnings announcements are accompanied by Reddit attention is still positive and significant for Reddit tone and price informativeness. However, the effect on the retail buying volume is positive but not significant.

Table 10: **Earnings Announcement with Abnormal Attention (Day $t + 1$).**

This table presents regression results exploring the interaction term between abnormal attention on Reddit and the earning surprise on earning announcement events. All dependent variable are measured at the excess level by controlling the average value of the stock's beta group sorted in decile. Dependent variable is measured at $t + 1$ as the independent variable. Column 1-3 present results for stocks with below median retail trade concentration. Column 4-6 present results for stocks with above median retail trade concentration. For each group, dependent variable of first column is reddit bullish tone. Second column is number of retail buy trades obtained from TAQ using the BJZZ algorithm. Third column is price informativeness measured using $1 - R^2$ by regressing return on Fama-French 3 factors. Earning surprise is measured using difference of actual EPS and median forecast from I/B/E/S. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11 and earning forecast from I/B/E/S. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	Reddit Tone	Retail Buy Trade	Price Informativeness	Reddit Tone	Retail Buy Trade	Price Informativeness
Abnormal	0.206*** (0.008)	32.40*** (0.651)	0.013*** (0.001)	2.316*** (0.060)	310.3*** (5.812)	0.023*** (0.001)
Forecast Error	0.059 (0.050)	13.78*** (4.056)	0.014*** (0.006)	-0.006 (0.014)	-1.899 (1.341)	0.0002 (0.0002)
Abnormal*Forecast Error	0.189*** (0.078)	6.529 (6.763)	0.025*** (0.009)	1.819*** (0.373)	45.36 (36.39)	0.018*** (0.006)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	434,523	434,523	434,523	358,029	358,029	358,029

4.2 Robinhood investors

We examine the impact of social media attention on information diffusion among Robinhood investors in this section. From Robintrack data, we obtain the aggregate users holdings of stocks, which can be a noisy proxy for the Robinhood buying activity.

To investigate Robinhood investors' trading pattern at earnings announcements dates with high retail interest on Reddit, we estimate a regression model within the framework of Equation (7), where we take Y_i as aggregate Robinhood users holdings at date $t - 1$, t and $t + 1$ around the events. The estimation results are shown in Table 11.

As seen in Table 11, the coefficient α_3 is surprisingly negative and significant on all three dates $t - 1$, t and $t + 1$, and for both the low concentration and high concentration groups. For earnings announcements that occurred in conjunction with abnormal attention events on Reddit, Robinhood investors change holdings opposite to the direction of earnings surprise both before, on and after the event. The evidence of Robinhood investors decreasing their holdings on stocks with more positive earnings forecast errors suggests that Robinhood investors are inexperienced and tends to make uninformed trades. Combined with the evidence in Section 3.1 that the general retail investors make informed trades at earnings announcement dates when social media attention is high, we show that retail investors are heterogeneous, with Robinhood investors being inexperienced in particular. This finding is consistent with previous studies on Robinhood investors such as Barber, Huang, Odean, and Schwarz (2021) and Eaton, Green, Roseman, and Wu (2022).

Table 11: **Earnings Announcement with Abnormal Attention (Robinhood Users Holding).**

This table presents regression results exploring the interaction term between abnormal attention on Reddit and the earning surprise on earning announcement events. All dependent variable are measured at the excess level by controlling the average value of the stock's beta group sorted in decile. Dependent variable is robinhood users holding, and is measured at $t - 1$, t and $t + 1$ as the independent variable. Robinhood users holding is obtained from robintrack website. Column 1-3 present results for stocks with below median retail trade concentration. Column 4-6 present results for stocks with above median retail trade concentration. Earning surprise is measured using difference of actual EPS and median forecast from I/B/E/S. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11 and earning forecast from I/B/E/S. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	Robinhood Users Holding (t-1)	Robinhood Users Holding (t)	Robinhood Users Holding (t+1)	Robinhood Users Holding (t-1)	Robinhood Users Holding (t)	Robinhood Users Holding (t+1)
Abnormal	195.3*** (13.32)	214.8*** (13.33)	226.3*** (13.34)	773.3*** (209.5)	1006.1*** (209.7)	1216.0*** (209.8)
Forecast Error	-88.32 (86.58)	-90.13 (85.54)	-73.97 (85.42)	-10.88 (257.8)	-12.94 (257.8)	-59.51 (257.8)
Abnormal*Forecast Error	-410.6*** (132.0)	-400.4*** (131.1)	-421.5*** (131.1)	-7766.6*** (1829.7)	-8320.3*** (1892.7)	-8380.1*** (1892.5)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	306,380	306,380	306,380	247,968	247,968	247,968

5 Conclusion

With the increasing popularity of social media platform for sharing investment opinions and more significant participation of retail investors in the stock market, an important question rises: what is the role of social media in retail investors' decision making process and how does it affect market consequences? To answer this question, we first document the correlation between abnormal social media attention and retail trading activities. When abnormal attention occurs in Reddit WallStreetBets, retail trading activities significantly increases especially for the group of stocks with high retail investor concentration. The abnormal attention in Reddit goes in accordance with retail trading activities - lasting abnormal attention is accompanied by lasting high retail trading activities, and abnormal attention that drops quickly also witness a drop in retail trades. Moreover, abnormal attention on Reddit is associated with high cumulative returns following the event, suggesting potentially high impact of retail investors on price.

While the above evidence shows the correlation between social media attention and retail investor trading activities, what remains unclear is whether the attention is informative or not. Does the attention facilitate the spread of information among retail investors? To address this question, we focus our analysis on a particular type of event which we have accounting information related to firm fundamentals. We assess the sentiment of social media attention on earnings announcement event when social media attention is abnormally high. The tone of abnormal social media attention is more bullish when positive earnings surprise occurs in conjunction with abnormal social media attention. In accordance with the bullish sentiment, retail buying trades are also higher, and price informativeness is higher. This provides evidence supporting the informativeness of social media attention in helping

spread information to retail investors in the process of price discovery. We also found opposite trading directions for Robinhood users, consistent with the previous literature suggesting the presence of heterogeneity in retail investors.

Our paper shows the relationship between social media attention and retail trading activities, and its potential impact for the financial market. We illustrate the role of abnormal social media attention in spreading firm specific information to retail investors. Given the close relationship between social media attention and retail investor trading activities, and the presence of heterogeneity in retail investors, an interesting direction of future work would be to explore the heterogeneity in social media participants and how they coordinate with each other. The relative role each group plays could potentially affect the aggregate impact of social media attention and retail investors for the financial market.

References

- David Y Aharon, Renatas Kizys, Zaghum Umar, and Adam Zaremba. Did david win a battle or the war against goliath? dynamic return and volatility connectedness between the gamestop stock and the high short interest indices. Working paper, 2021.
- Franklin Allen, Marlene Haas, Eric Nowak, Matteo Pirovano, and Angel Tengulov. Squeezing shorts through social media platforms. Working paper, 2021.
- Ray Ball and Philip Brown. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, pages 159–178, 1968.
- Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2):785–818, 2008.
- Brad M Barber, Terrance Odean, and Ning Zhu. Systematic noise. *Journal of Financial Markets*, 12(4):547–569, 2009.
- Brad M Barber, Xing Huang, Terrance Odean, and Christopher Schwarz. Attention induced trading and returns: Evidence from robinhood users. *Journal of Finance*, forthcoming, 2021.
- Ekkehart Boehmer, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang. Tracking retail investor activity. *The Journal of Finance*, 76(5):2249–2305, 2021.
- Daniel Bradley, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao. Place your bets? the market consequences of investment research on reddit’s wallstreetbets. Working paper, 2021.
- Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5):1367–1403, 2014.
- J Anthony Cookson, Joseph Engelberg, and William Mullins. Echo chambers. Working paper, 2022.

- Gregory W Eaton, T Clifton Green, Brian S Roseman, and Yanbin Wu. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, 146(2):502–528, 2022.
- Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.
- Michael Farrell, T Clifton Green, Russell Jame, and Stanimir Markov. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, 145(2):616–641, 2022.
- Nicolae B Gârleanu, Stavros Panageas, and Geoffery X Zheng. A long and a short leg make for a wobbly equilibrium. Working paper, National Bureau of Economic Research, 2021.
- Danqi Hu, Charles M Jones, Valerie Zhang, and Xiaoyan Zhang. The rise of reddit: How social media affects retail investors and short-sellers’ roles in price discovery. Working paper, 2021.
- Russell Jame, Rick Johnston, Stanimir Markov, and Michael C Wolfe. The value of crowd-sourced earnings forecasts. *Journal of Accounting Research*, 54(4):1077–1110, 2016.
- Ron Kaniel, Gideon Saar, and Sheridan Titman. Individual investor trading and stock returns. *The Journal of Finance*, 63(1):273–310, 2008.
- Eric K Kelley and Paul C Tetlock. How wise are crowds? insights from retail orders and stock returns. *The Journal of Finance*, 68(3):1229–1265, 2013.
- Runeet Kishore, Michael W Brandt, Pedro Santa-Clara, and Mohan Venkatachalam. Earnings announcements are full of surprises. Working paper, 2008.
- Cheng Long, Brian M Lucey, and Larisa Yarovaya. ”i just like the stock”: The role of reddit sentiment in the gamestop share rally. Working paper, 2021.
- Štefan Lyócsa, Eduard Baumöhl, and Tomáš Vřrost. Yolo trading: Riding with the herd during the gamestop episode. *Finance Research Letters*, 46:102359, 2022.

- Randall Morck, Bernard Yeung, and Wayne Yu. The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of financial economics*, 58(1-2):215–260, 2000.
- Gideon Ozik, Ronnie Sadka, and Siyi Shen. Flattening the illiquidity curve: Retail trading during the covid-19 lockdown. *Journal of Financial and Quantitative Analysis*, 56(7): 2356–2388, 2021.
- Joel Peress. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance*, 69(5):2007–2043, 2014.
- Valentina Semanova and Julian Winkler. Social contagion and asset prices: Reddit’s self-organised bull runs. Working paper, 2021.
- Ivo Welch. The wisdom of the robinhood crowd. *The Journal of Finance*, 77(3):1489–1527, 2022.

Appendix

Table A1: Variable Descriptions

Variable	Source	Definition
<i>Comments</i>	Wallstreetbets	Number of comments posted in Reddit Wallstreetbets about a stock collected using Pushshift API. Comments made after the end of the previous trading day and before the end of the trading day are categorized into comments made in the respective trading day.
<i>Submissions</i>	Wallstreetbets	Number of submissions posted in Reddit Wallstreetbets about a stock collected using Pushshift API. Submissions made after the end of the previous trading day and before the end of the trading day are categorized into submissions made in the respective trading day.
<i>Attention Tone</i>	Wallstreetbets	Categorization of submissions and comments to bullish tendency or bearish tendency by identifying words in the respective direction.
<i>Abnormal Attention Event</i>	Wallstreetbets	Occurrence of the total mention of a stock exceeding 150% of the previous 6-month mention average for at least two days.
<i>Number of retail trades</i>	TAQ	Following Boehmer, Jones, Zhang and Zhang (BJZZ,2021) algorithm, information on daily retail buy or sell trades obtained from TAQ.
<i>Retail trading volume</i>	TAQ	Following BJZZ algorithm, information on daily retail buy or sell trading volume obtained from TAQ.
<i>Robinhood User holdings</i>	Robintrack	Number of Robinhood users holding a stock on a trading day obtained from information available on Robintrack.

<i>Number of Shares Traded</i>	CRSP	Number of shares traded in the market of a stock on a trading day.
<i>Volume of Shares Traded</i>	CRSP	Volume of shares traded in the market of a stock on a trading day.
<i>Adjusted stock price</i>	CRSP	Adjusted stock price using the cumulative price adjustment factor obtained from CRSP.
<i>Excess Return</i>	CRSP	Stock return calculated using stock price information from CRSP controlling for beta group.
<i>Beta</i>	CRSP	Calculated with one-year rolling window using daily return information.
<i>Price informativeness</i>	CRSP	Calculated using $1 - R^2$ from one-month rolling window regression by regressing excess return on Fama-French 3 factor. A higher value indicates more firm specific information incorporated into price.
<i>Earnings announcement day</i>	Compustat	Marks whether the trading day is an earning announcement day for the company. Takes the value of one if yes and value of zero if no.
<i>Earnings surprise</i>	I/B/E/S	Compare the announced EPS with the analyst forecast to obtain whether the earning announcement is of positive or negative surprise to the market.
<i>Analyst forecast error</i>	I/B/E/S	Difference between actual EPS and median analyst forecast from I/B/E/S.
<i>Market Capitalization</i>	CRSP	Market capitalization of a stock calculated using outstanding share of the stock times stock price.
<i>Book Value</i>	Compustat	Firm book value calculated using total assets minus total liabilities
<i>Industry Classification</i>	Compustat	Firm industry information obtained at 4-digit SIC classification.

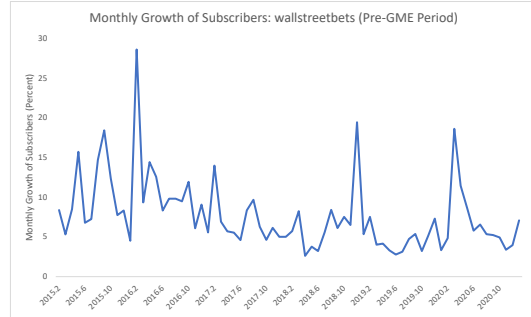
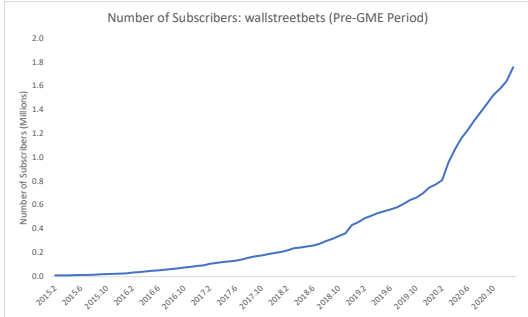


Figure A1: **Subscribers of WallStreetBets (Pre-GME Period)**. This figure shows the number of subscribers and subscriber growth of the subreddit WallStreetBets. The figure only covers the period before the GameStop event. The time period is from Feb 2015 to Dec 2020. Statistics is obtained from the following website: <https://subredditstats.com/r/WallStreetBets>.

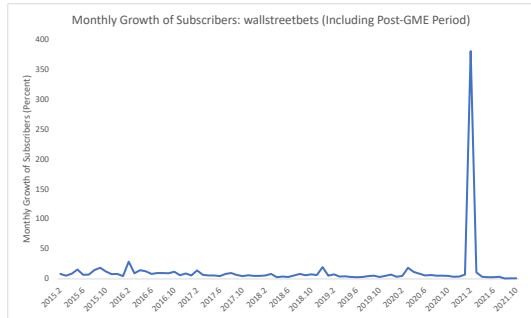
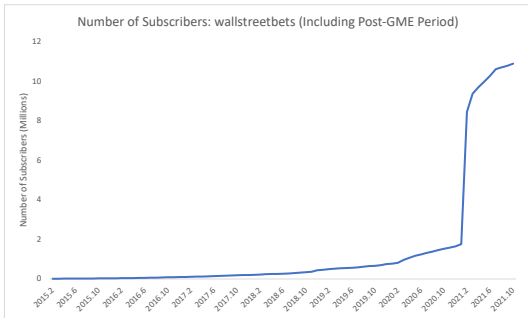
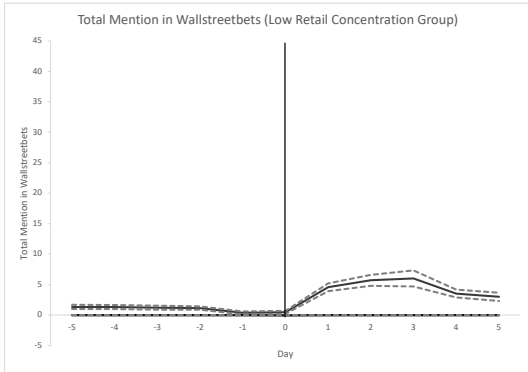


Figure A2: **Subscribers of WallStreetBets (Including Post-GME Period)**. This figure shows the number of subscribers and subscriber growth of the subreddit WallStreetBets. The figure covers the GameStop event. The time period is from Feb 2015 to Oct 2021. Statistics is obtained from the following website: <https://subredditstats.com/r/WallStreetBets>.

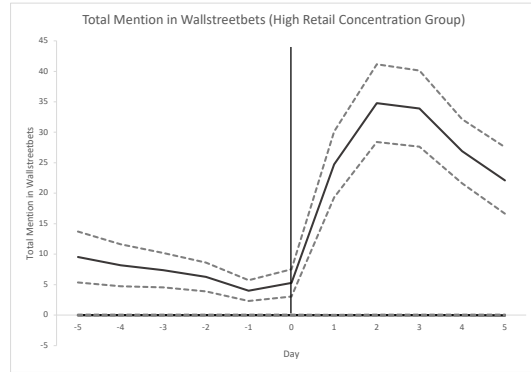
Table A2: **Summary Statistics (Earnings Announcement with Abnormal Attention).**

This table presents summary statistics for the whole sample and abnormal attention events occurred with earnings announcement. Wallstreetbets mentions are the sum of mentions in submissions and in comments. Retail trading data is obtained using the BJZZ algorithm from TAQ. Robinhood users holding is obtained from robintrack website. Price informativeness is measured using $1 - R^2$ by regressing return on Fama-French 3 factors. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, with CRSP share code of 10 or 11, and have analyst coverage from I/B/E/S. Left panel presents statistics for the group of stocks which has below median retail investor concentration during the sample period. Right panel presents statistics for the group which has above median retail investor concentration. The upper panel reports the statistics for the entire sample period. The middle panel reports statistics for the days with earnings announcement. The lower panel reports statistics for the days which the abnormal mention event occurred in conjunction with earnings announcement.

	Group: Low Retail Concentration						Group: High Retail Concentration					
	obs	mean	st.dev.	median	Q1	Q3	obs	mean	st.dev.	median	Q1	Q3
Whole Sample (Only Include Stocks with Earnings Forecast Analyst Coverage)												
total WSB mentions	454,152	1.0	9.7	0	0	0	394,907	6.2	66.90	0	0	1
number of buy trade	454,152	173.4	322.9	87	34	208	394,907	893.8	3579.8	159	35	601
Robinhood holdings	318,594	1337.6	4190.6	430	168	1,126	272,591	12937.6	42662.5	2,133	654	6,593
price informativeness	454,152	0.51	0.24	0.50	0.32	0.69	394,907	0.60	0.23	0.62	0.43	0.79
Earnings Announcement												
total WSB mentions	6,224	3.0	16.9	0	0	1	5,509	18.7	91.70	0	0	4
number of buy trade	6,224	384.6	724.9	172	68	403	5,509	1550.1	3960.3	321	72	1,195
Robinhood holdings	5,130	1437.6	4325.1	470	188	1,224	4,485	13920.8	46096.6	2,178	681	7,067
forecast error	5,667	0.07	0.58	0.04	-0.02	0.12	5,509	-0.31	12.40	0.11	-0.03	0.03
price informativeness	6,224	0.53	0.23	0.52	0.34	0.71	5,509	0.59	0.22	0.61	0.42	0.79
Earnings Announcement with Abnormal Mention												
total WSB mentions	1,102	10.7	31.9	3	1	8	1,383	56.9	161.10	9	3	39
number of buy trade	1,102	798.2	1227.1	406	184	870	1,383	3435.5	5702.3	1365	545	3,774
Robinhood holdings	970	3271.8	7737.6	1,237	462	3,072	1,198	28487.6	60269.6	7,307	2,844	22,108
forecast error	1,053	0.08	0.58	0.04	0.00	0.14	1,273	0.03	0.85	0.05	-0.01	0.13
price informativeness	1,102	0.52	0.23	0.52	0.35	0.71	1,383	0.54	0.24	0.54	0.36	0.72



A



B

Figure A3: **Average Excess Wallstreetbets Mention (Abnormal Attention Event).** This figure plots the average of excess mention of stocks on Reddit Wallstreetbets in a 11-day window. The excess mention is measured controlling the average of the stock's respective beta group. The vertical line marks the day before the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

Table A3: **Cumulative Return Regression Results - Onset of Abnormal Attention (Attention Dropped in 2 days).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal day dummy for which mention dropped in 2 days. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close
Dummy	0.089*** (0.037)	0.241*** (0.072)	0.283*** (0.086)	0.348*** (0.074)	0.639*** (0.132)	0.934*** (0.176)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

Table A4: **Cumulative Return Regression Results - Onset of Abnormal Attention (Attention Remained at least 5 days).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal day dummy for which mention remained high for at least 5 days. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close
Dummy	0.099 (0.075)	0.119 (0.135)	0.096 (0.177)	0.542*** (0.066)	0.793*** (0.116)	0.987*** (0.128)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

Table A5: **Price Informativeness Regression Results (Attention Dropped or Remained High).**

This table presents regression results for event study regression results of price informativeness around the abnormal mention event. Price informativeness is measured at the excess level by controlling the average value of the stock's beta group sorted in decile. Column 1 and 2 present results for stocks with below median retail trade concentration. Column 3 and 4 present results for for stocks with above median retail trade concentration. For each type of event, the left column reports results for abnormal mention event in which attention dropped in 2 days. The right column reports regression results abnormal mention event in which attention remained high for at least 5 days. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration		High Retail Concentration	
	Attention dropped	Attention Remain	Attention dropped	Attention Remain
day-5	-0.000124 (0.00366)	0.00576 (0.00747)	0.00907** (0.00392)	-0.0148** (0.00594)
day-4	-0.000509 (0.00366)	0.00947 (0.00747)	0.00718* (0.00392)	-0.0117** (0.00594)
day-3	0.000306 (0.00366)	0.00929 (0.00747)	0.00289 (0.00392)	-0.00848 (0.00594)
day-2	0.00212 (0.00366)	0.00547 (0.00747)	0.00449 (0.00392)	-0.00980* (0.00594)
day-1	0.00325 (0.00366)	0.00480 (0.00747)	0.00491 (0.00392)	-0.00617 (0.00594)
day0	0.00227 (0.00366)	0.00639 (0.00747)	0.00344 (0.00392)	-0.00318 (0.00594)
day1	0.0112*** (0.00366)	0.0108 (0.00747)	0.00871** (0.00392)	0.0118** (0.00594)
day2	0.0158*** (0.00366)	0.0172** (0.00747)	0.0141*** (0.00392)	0.0151** (0.00594)
day3	0.0146*** (0.00366)	0.0290*** (0.00747)	0.0162*** (0.00392)	0.0294*** (0.00594)
day4	0.0141*** (0.00366)	0.0425*** (0.00747)	0.0171*** (0.00392)	0.0361*** (0.00594)
day5	0.0145*** (0.00366)	0.0469*** (0.00747)	0.0166*** (0.00392)	0.0456*** (0.00594)
Firm Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Obs.	1,030,136	951,468	1,030,136	951,468

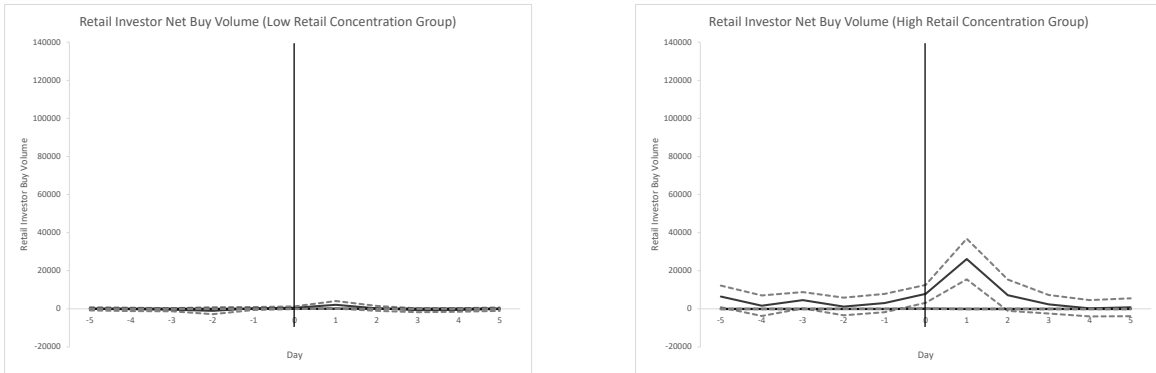


Figure A4: Average Excess Retail Net Buy Volume (Attention Dropped in 2 Days).

This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess volume is measured controlling the average of the stock’s respective beta group. The vertical line marks the day before the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from TAQ following the BJZZ algorithm. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

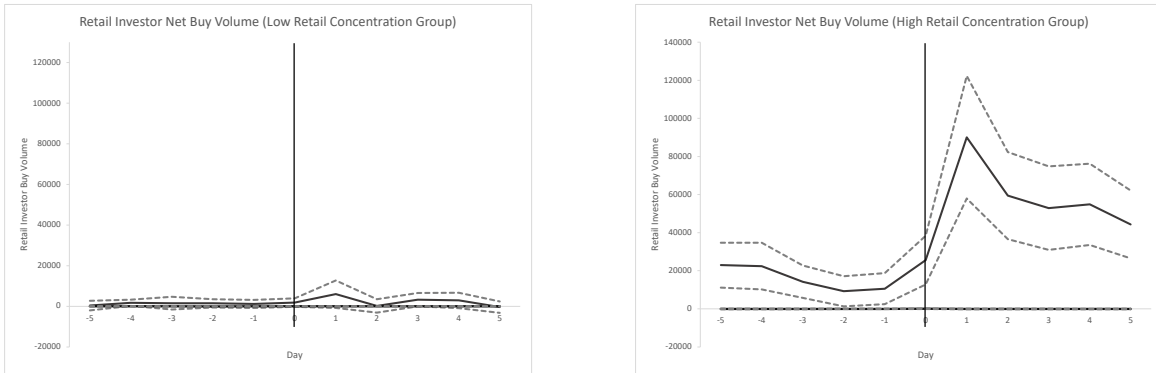


Figure A5: **Average Excess Retail Net Buy Volume (Attention Remained High at least 5 Days).**

This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess volume is measured controlling the average of the stock's respective beta group. The vertical line marks the day before the start of the abnormal attention event. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from TAQ following the BJZZ algorithm. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

Internet Appendix:

Social Media Attention, Stock Returns and Retail Trades

Weiting Hu and Rusi Yan

This internet appendix presents results for robustness tests of abnormal social media attention events exclude earnings announcement.

- Figure IA1. Average Excess Cumulative Return
- Table IA1. Cumulative Return Regression Results (Onset of abnormal attention)
- Table IA2. Cumulative Return Regression Results (After abnormal attention)
- Table IA3. Price Informativeness Regression Results
- Figure IA2. Average Excess Retail Buy Volume
- Figure IA3. Average Excess Robinhood Users Holding
- Table IA4. Retail Trading Activity Regression Results

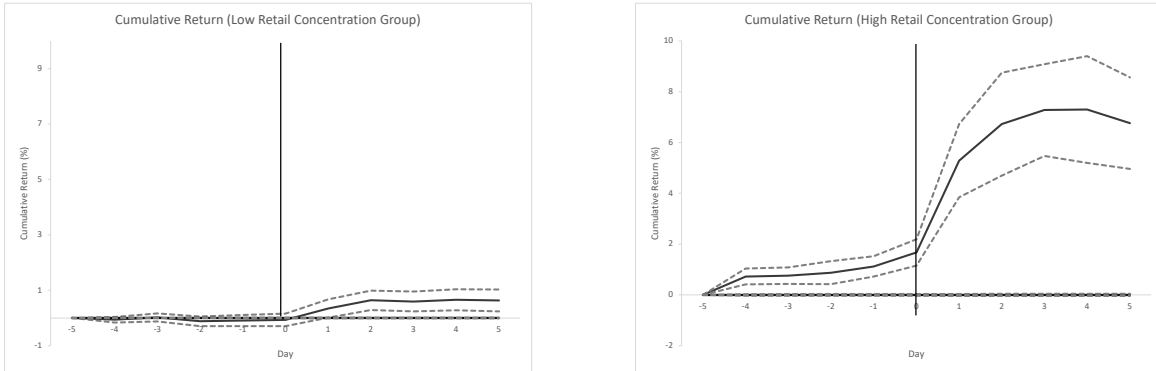


Figure IA1: Average Excess Cumulative Return (Exclude Earnings Announcement).

This figure plots the average of excess cumulative return of stocks on a 11-day window. The excess return is measured controlling the average of the stock's respective beta group. The cumulative return is calculated relative to stock price five days prior to the start of the abnormal attention event excluding those near earnings announcement. The dotted lines are the pointwise 95% confidence interval for the average return. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average cumulative return for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

Table IA1: **Cumulative Return Regression Results - Onset of Abnormal Attention (Exclude Earnings Announcement)**.

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal mention day dummy for each stock excluding those near earnings announcement. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close	close(t-1) -to-close	close(t-3) -to-close	close(t-5) -to-close
Dummy	0.049 (0.041)	0.086 (0.073)	0.069 (0.095)	0.614*** (0.071)	1.087*** (0.126)	1.461*** (0.169)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

Table IA2: **Cumulative Return Regression Results - After Abnormal Attention (Exclude Earnings Announcement).**

This table presents regression results for cumulative excess return on stocks. Excess return is calculated by controlling the stock's respective beta group. Excess return is measured in percentage point. The independent variable is the abnormal mention day dummy for each stock excluding those near earnings announcement. The left panel reports results for the group of stocks which has below median retail investor concentration. The right panel reports results for the group of stocks which has above median retail investor concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration			High Retail Concentration		
	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)	close(t-1) to-close(t+1)	close(t-3) to-close(t+3)	close(t-5) to-close(t+5)
Dummy	0.036 (0.058)	0.156 (0.106)	0.235* (0.140)	0.882*** (0.101)	1.767*** (0.190)	2.390*** (0.464)
Firm Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Obs.	1,009,533	1,009,601	1,009,655	936,020	936,072	936,043

Table IA3: **Price Informativeness Regression Results (Exclude Earnings Announcement)**.

This table presents regression results for event study regression results of price informativeness around the abnormal mention event. Dependent variable is measured at the excess level by controlling the average value of the stock's beta group sorted in decile. The independent variable is the abnormal mention day dummy for each stock excluding those near earnings announcement. The left column presents result for stocks with below median retail trade concentration. The right column presents results for for stocks with above median retail trade concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Price Informativeness	High Retail Price Informativeness
day-5	0.00842* (0.00430)	0.00181 (0.00396)
day-4	0.00874** (0.00430)	0.00508 (0.00396)
day-3	0.00557 (0.00430)	0.00754* (0.00396)
day-2	0.00448 (0.00430)	0.00582 (0.00396)
day-1	0.00516 (0.00430)	0.00684* (0.00396)
day0	0.00479 (0.00430)	0.00729* (0.00396)
day1	0.0102** (0.00430)	0.0178*** (0.00396)
day2	0.00769* (0.00430)	0.0166*** (0.00396)
day3	0.00838* (0.00430)	0.0173*** (0.00396)
day4	0.0105** (0.00430)	0.0163*** (0.00396)
day5	0.0128*** (0.00430)	0.0194*** (0.00396)
Firm Controls	Y	Y
Industry FE	Y	Y
Time FE	Y	Y
Obs.	1,030,136	951,468

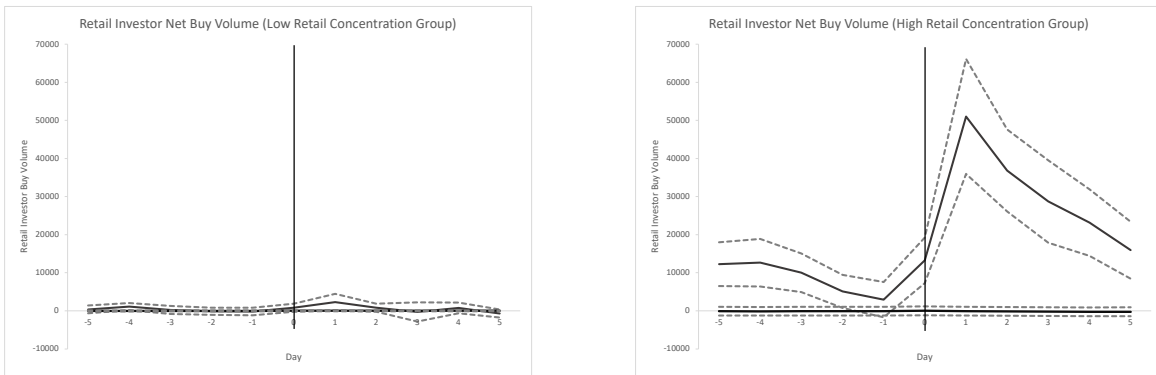


Figure IA2: **Average Excess Retail Buy Volume (Exclude Earnings Announcement).**

This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess volume is measured controlling the average of the stock's respective beta group. The vertical line marks the day before the start of the abnormal attention event excluding those near earnings announcement. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from TAQ following the BJZZ algorithm. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

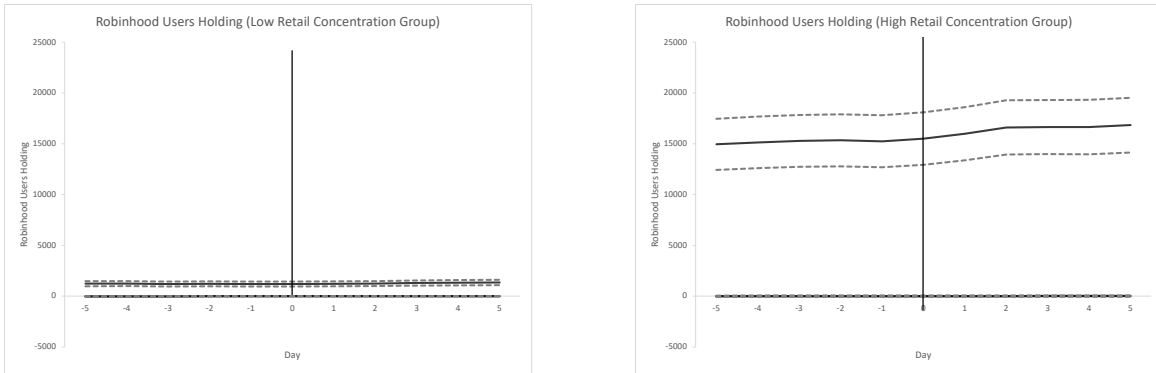


Figure IA3: **Average Excess Robinhood Users Holding (Exclude Earnings Announcement).**

This figure plots the average of excess retail investor buy volume of stocks on a 11-day window. The excess holding is measured controlling the average of the stock's respective beta group. The vertical line marks the day before the start of the abnormal attention event excluding those near earnings announcement. The dotted lines are the pointwise 95% confidence interval. Sample period is from Jan 2018 to Dec 2020. Retail buy volume is obtained from Robintrack. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Left figure is the average for the group of stocks which has below median retail investor concentration. Right figure is for the group has above median retail investor concentration.

Table IA4: **Retail Trading Activity Regression Results (Exclude Earnings Announcement)**.

This table presents regression results for event study regression results of retail trading activity around the abnormal attention events excluding those near earnings announcement. Dependent variable is measured at the excess level by controlling the average value of the stock's beta group sorted in decile. The first column uses retail buy trade volume as dependent variable. The second column uses robinhood users holding as the dependent variable. The left panel reports results for stocks with below median retail trade concentration. The right panel reports regression results for stocks with above median retail trade concentration. Sample period is from Jan 2018 to Dec 2020. Stocks are the list of stocks which has available data from TAQ, while having a CRSP share code of 10 or 11. Firm controls included are log book-to-market ratio and log market cap measured at the end of the previous month. Industry fixed effect at the 4-digit SIC level and time fixed effect at the month level are included. Levels of significance are denoted by *(10%), **(5%), and ***(1%).

	Low Retail Concentration		High Retail Concentration	
	Retail Net Buy	Robinhood	Retail Net Buy	Robinhood
	Volume	holding	Volume	holding
day-5	107.5 (355.3)	390.4*** (42.09)	9038.0*** (1958.8)	3790.6*** (620.3)
day-4	767.2** (355.6)	399.5*** (42.17)	8267.5*** (1959.2)	3981.5*** (620.0)
day-3	732.9** (355.2)	406.9*** (42.16)	6794.1*** (1960.1)	4172.9*** (621.9)
day-2	2.884 (355.3)	419.4*** (42.16)	2034.7 (1957.1)	4263.9*** (621.2)
day-1	138.7 (355.0)	428.4*** (42.25)	-1209.0 (1955.3)	4177.9*** (620.9)
day0	869.1** (355.0)	430.1*** (42.32)	9151.3*** (1957.1)	4387.4*** (621.6)
day1	954.3*** (355.8)	445.3*** (42.33)	44063.7*** (1957.9)	4786.0*** (621.6)
day2	1067.3*** (355.5)	467.7*** (42.39)	34720.6*** (1959.3)	5303.4*** (622.3)
day3	-630.3* (355.2)	485.0*** (42.48)	27941.8*** (1958.3)	5339.9*** (622.8)
day4	841.7** (355.3)	485.3*** (42.37)	20956.8*** (1956.6)	5360.8*** (623.0)
day5	-321.3 (355.2)	487.3*** (42.46)	11645.0*** (1957.5)	5446.0*** (622.8)
Firm Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Obs.	1,030,141	707,606	951,508	610,061