Place your bets? The market consequences of investment research on Reddit's Wallstreetbets[*]

Daniel Bradley[a], Jan Hanousek Jr.[b, c], Russell Jame[d], and Zicheng Xiao[e]

[a]*Department of Finance, University of South Florida, Tampa, FL 33620, danbradley@usf.edu*

[b]*Department of Finance, University of Memphis, Memphis TN 38152, jan.hanousek@memphis.edu*

[c] *Faculty of Business and Economics, Department of Finance, Mendel University in Brno, Czech Republic*

[d]*Gatton College of Business, University of Kentucky, Lexington, KY 40515, 859.218.1793, russell.jame@uky.edu*

[e]*Department of Finance, University of South Florida, Tampa, FL 33620, zicheng@usf.edu*

April 2023

Abstract

We examine the consequences of due diligence recommendations on Reddit's Wallstreetbets (*WSB*) platform. Before the Gamestop (GME) short squeeze, recommendations are significant predictors of returns and cash-flow news. This predictability is completely eliminated post-GME. Post-GME, the fraction of reports emphasizing price-pressure or attention-grabbing stocks dramatically increases, and the decline in informativeness is concentrated in these reports. Similarly, retail trade informativeness increases following DD reports in the pre-GME period, but not post-GME. Our findings are consistent with the view that the Gamestop event altered the culture of *WSB*, leading to a deterioration in investment quality that adversely impacted smaller investors.

Keywords: Reddit, Wallstreetbets, *WSB*, retail trading, social media, Gamestop, trading frenzies

JEL classifications: G20, G23

[*] We thank Gregory Eaton (discussant), John Campbell (discussant), Tyson Van Alfen (discussant), Paul Irvine (discussant), Andriy Shkilko (discussant), Michael Farrell and seminar participants at the FARS Midyear Meeting, the Stockholm Business School (Future of Financial Information Webinar), the Texas A&M Young Scholars Finance Consortium, the 2022 Financial Management Association Meeting, the 2022 Southern Finance Association Meeting, and the University of Kentucky.

Place your bets? The market consequences of investment research on Reddit's Wallstreetbets

1. Introduction

On February 18, 2021, the CEOs of Reddit and Robinhood along with a Reddit user testified before Congress for their role in the well-publicized Gamestop (GME) short squeeze that sent shares to almost $500 before plummeting to around $50 a few days later.[1] One of the main concerns among regulators can be summed up with the following quote during testimony: "The Reddit discussions are in many ways quite worrisome. They create volatility in the markets, and volatility is generally bad. It creates all kinds of dislocations."

While regulators have previously raised alarm about the impact of social media on stock market efficiency and investor welfare, several unique attributes of Reddit's *Wallstreetbets (WSB)* platform have exacerbated these concerns.[2] First, unlike other social media platforms, users' postings are anonymous. During testimony, several Congressional representatives questioned this policy. Reddit's CEO countered suggesting that *WSB* would not exist if users had to reveal their identity.[3] In addition, in contrast to other prominent investment-related social media platforms such as Seeking Alpha, *WSB* posts have virtually no editorial review and tend to focus on speculative strategies that emphasize small probabilities of large gains, possibly at the expense of lower expected returns (Bali, Cakici, and Whitelaw, 2011). Further, as evidenced by the Gamestop episode, *WSB* appears to have a distinct ability to induce trading frenzies that can potentially harm smaller investors, destabilize prices, and even impact the real economy (Goldstein, Ozdenoren, and Yuan, 2013).

We take a first look at the determinants and capital market consequences of the investment research provided on *WSB*. To disentangle the general effects of social media with the more unique features of *WSB*, we contrast *WSB* research with research provided by Seeking Alpha. Presumably, many people were attracted to *WSB* based on the extraordinary success of the Gamestop short squeeze. Given investors tendency to extrapolate from salient recent events, one concern is that new users will flood the site with research reports on attention-grabbing stocks that emphasize strategies unrelated to firm fundamentals (e.g., short squeezes).

[1] Representatives from hedge funds Melvin Capital and Citadel also testified. For the transcript of the testimony, see https://www.c-span.org/video/?508545-2/GameStop-hearing-part-2.
[2] For an example of previous regulator concerns, see https://www.nytimes.com/2013/04/29/business/media/social-medias-effects-on-markets-concern-regulators.html.
[3] https://finance.yahoo.com/finance/news/reddit-ceo-testimony-122526759.html.

To the extent that users overestimate the effectiveness of such strategies, *WSB* research after the GME event may be particularly uninformative.[4]

We focus primarily on single firm 'Due Diligence" (DD) reports, which are reports identified by the poster (and verified by moderators) as containing some type of analysis and a clear buy or sell signal. DD reports contain clear investment recommendations and potentially new value-relevant information, which makes them most comparable to other forms of crowdsourced investment research, such as Seeking Alpha reports. We also separately examine non-research related *WSB* posts (e.g., meme posts, bragging about recent gains and losses, etc.). Although non-research posts are unlikely to contain useful information, they could still influence prices through attention-based buying (Barber and Odean, 2008). Our sample includes 5,015 DD reports and 13,255 non-research related *WSB* posts issued between July 2018 and June 2021.

Consistent with the view that *WSB* emphasizes high-risk investments, we find that both DD reports and non-research posts on *WSB* tilt towards young, volatile stocks with high skewness and high short interest. *WSB* preference towards speculative investments also increases substantially in the post-GME period. For example, *WSB* tendency to cover more volatile stocks increases by 150%, while *WSB* coverage of stocks with heavy short interest increases by nearly 500%. The time-series patterns are consistent with the Gamestop event attracting even more risk-seeking users. While *SA* coverage also tilts towards volatile firms, *SA* coverage of speculative stocks does not increase in the post-GME period. These findings suggest that the GME event had an impact on *WSB* that did not generalize across all social media platforms.

In the pre-GME period, we find DD reports are significant predictors of future returns. For example, an incremental DD buy recommendation is associated with a 5.17% increase in one-month ahead returns for the full sample and a 2.33% increase after excluding GME and AMC.[5] However, the one-month return predictability is fully eliminated in the post-GME period. Consistent with prior work (e.g., Chen et al., 2014), we find that *SA* research report recommendations also predict future returns. However, we find no evidence

[4]Prior work suggests that strategies based on extrapolative expectations often earn lower expected returns (see, e.g., Greenwood and Shleifer, 2014; Barberis et al., 2015; Cassella and Gulen, 2018; and Da, Huang, and Jin, 2021).
[5] We have also considered excluding all meme stocks, defined as the 50 stocks for which Robinhood imposed a trading halt (https://www.theverge.com/2021/1/29/22256419/robinhood-limits-wall-street-bets-stock-buys). Excluding other meme stocks apart from GME and AMC has a negligible impact on the results.

that *SA* research reports exhibit a decline in informativeness in the post-GME period, which alleviates the concern that broad economic forces resulted in a deterioration in investment research across all social media platforms. We also find no evidence that non-research *WSB* postings are informative, even in the pre-GME period, which is consistent with our conjecture that DD reports are distinct from other *WSB* posts.

One concern is that the informativeness of DD reports in the pre-GME period is largely a consequence of the unusual pandemic market, which amplified market volatility and generated a surge in retail trading (Ozik, Sadka, and Shen, 2021). To explore this conjecture, we partition the pre-GME period into pre-pandemic and post-pandemic. We find that the informativeness of DD reports are slightly larger (albeit insignificantly so) for reports issued prior to the pandemic, which is inconsistent with the view that the results are driven by factors related to the pandemic (e.g., excess volatility, working from home, stimulus checks, etc.). More generally, we find that the predictive ability of DD reports was stable in 2020, and then immediately and permanently declines in the post-GME period. The lack of pre-trends in the pre-GME period and the sharp decline in informativeness in the post-GME period is consistent with the GME event itself contributing to the decline in report informativeness.

We also consider that the predictive ability of *WSB* research in the pre-GME period is not driven by reports containing value-relevant information (*information*), but rather because DD reports induce uninformed demand shocks that push prices beyond fundamentals (*price pressure*). However, we find no evidence that the return patterns reverse over longer-horizons (up to 12 weeks following the report), which is inconsistent with the joint hypothesis of price pressure and short-term downward sloping demand curves. We also document a positive relation in the pre-GME period between DD reports and future media sentiment, earnings surprises, and analyst earnings forecast revisions, which points to the possibility that at least a portion of the return predictability associated with DD reports in the pre-GME period is attributable to forecasting future cash-flow news. However, in the post-GME period, the positive relation between DD reports and cash-flow news is fully eliminated, and in some cases, it even becomes significantly negative.

Why does report informativeness decline in the post-GME period? We conjecture that the remarkable success of the GME short squeeze may have attracted new users who 1) place too much emphasis on coordinated trading strategies, possibly at the expense of analyzing firm fundamentals and, 2) have the tendency to trade attention-grabbing stocks, which has been shown to negatively impact retail traders (Barber et al., 2022). We find support for both conjectures. In particular, the fraction of reports focusing on price pressure strategies increases by 165% in the post-GME period, while the fraction of reports issued on attention-grabbing stocks increases by 75%. Further, the decline in return predictability of DD reports in the post-GME period is significantly stronger among these reports. In fact, while post-GME DD reports are, on average, uninformative, we find evidence of significant negative return predictability among the subset of price pressure or attention-grabbing reports.

Our last set of tests examine how investors of differing sophistication levels trade following DD reports. We consider three groups of investors: institutional investors, large retail investors (as proxied by volume-based measures of retail order imbalance), and small retail investors (as proxied by trade-based measures of retail order imbalance). We compute a measure of standardized abnormal imbalances for each investor type, which captures the intensity of the directional order imbalances for a firm relative to a benchmark period. We find that abnormal institutional imbalances are uncorrelated with DD report recommendations, abnormal large retail imbalances are modestly correlated, and abnormal small retail imbalances exhibit a substantial correlation. We also find that *SA* recommendations correlate with retail imbalances, particularly smaller retail investors. However, the increase in abnormal small retail imbalances following a DD buy recommendation is roughly four times larger than the corresponding increase following an *SA* buy recommendation.

Since small retail investor imbalances are highly correlated with DD report recommendations, the decline in DD report informativeness in the post-GME period may lead to a decline in small retail trade informativeness. Consistent with this view, we find that small retail trade informativeness is significantly larger following DD reports issued in the pre-GME period, but this relation is fully eliminated in the post-GME period. The finding is consistent with the decline in the informativeness of *WSB* adversely impacting less-sophisticated retail investors.

Our study contributes to the literature about the informativeness of investment research provided on social media. While some papers find a significant positive relation between investment opinions on social finance sites and future stock returns (e.g., Chen et al., 2014; Jame et al., 2016; Crawford et al., 2018; Bartov, Faurel, and Mohanram, 2018), others do not (e.g., Tumarkin and Whitelaw, 2001; Kim and Kim, 2014; Giannini, Irvine, and Shu, 2018; Ammann and Schaub, 2021). *WSB* has recently become the most influential social finance site by many metrics, and it offers many features that are distinct from other social media sites. These features give rise to its apparent ability to induce trading frenzies among smaller investors and coincidently significantly greater scrutiny from regulators. Consistent with these differences being important, we find that the influence of *WSB* on financial markets is very different from Seeking Alpha. In this sense, our finding echoes recent work by Cookson et al. (2022) who find that sentiment exhibits very little correlation across three different social media platforms (Twitter, StockTwits, and Seeking Alpha), and who caution readers from drawing general conclusions about social media based on evidence from one specific platform.

Our study also extends the nascent literature that explores the growing importance of *WSB* and its impact on financial markets. Several contemporaneous papers focus on the dynamics between *WSB* activity and one-day ahead returns, trading volume, short interest, volatility, and market quality (e.g., Aharon et al., 2021; Semenova and Winker, 2021; Hu et al., 2021; Long, Lucey, and Yarovaya, 2021; Eaton et al., 2022; and Allen et al., 2021). Relative to this literature, we make at least three contributions. First, we distinguish between DD reports and non-research posts. We show that DD reports were far more informative that non-research posts in the pre-GME period, and we encourage future research on *WSB* to separately analyze DD reports (i.e., investment research) from other *WSB* posts that are more likely to proxy for retail attention. Second, we benchmark our findings from *WSB* to Seeking Alpha. This contrast highlights several important differences between the two sites. For example, we find that small retail imbalances are much more strongly related to *WSB* reports, which is consistent with regulators' concerns that *WSB* posts are more likely to induce retail trading frenzies. Lastly, we contrast the value of DD reports in the pre-GME and post-GME period. Our findings highlight a dramatic decline in the informativeness of *WSB* in the post-GME period that is at least partially attributable to increased emphasis on coordinated trading strategies and attention-grabbing stocks.

Finally, we add to the literature on retail trading. Several recent papers highlight a surge in retail trading spurred by innovations by fintech brokerage firms (Barber et al., 2022) and pandemic-related disruptions (Ozik, Sadka, and Shen, 2021). Our results suggest that the growth in *WSB* is another factor that contributed to the recent growth in retail trading. A related literature examines the informativeness of retail trading. Early work finds that retail traders are uninformed 'dumb money' (e.g., Hvidkjaer, 2008; Frazzini and Lamont, 2008; Barber, Odean, and Zhu, 2009). However, more recent evidence suggests that retail traders are informed (e.g., Kaniel et al., 2012; Kelley and Tetlock, 2013; Boehmer et al., 2021; Welch, 2022). Farrell et al. (2022) suggest that the improvements in retail trading over time may be partially attributable to the increasing democratization of investment research. Consistent with Farrell et al. (2022), we find evidence consistent with *WSB* reports improving the informativeness of retail trading, but this relation only holds during the pre-GME period.

2. Data and Descriptive Statistics

2.1 Reddit and Wallstreetbets – Background

Reddit is a social media platform founded in June 2005. Like many other social media websites, contributors post content, and users can add comments in response to the original post. The Reddit community is a collection of forums, where each forum is dedicated to a particular topic called a subreddit. Each subreddit is then organized into several pages based on users' ranking criteria. For instance, the default page is the 'Hot Page,' which lists the currently most viewed posts or posts with the most active commentators. 'New Posts' lists posts based on the listing timestamp, and 'Top posts' lists posts with the most likes (upvotes) and comments for a specified period. When a new post is written, it is only visible in the new post category. The post can then move up to the hot page if it reaches sufficient traffic.

Wallstreetbets (*WSB*) is one of many subreddits within the Reddit community. It was created on January 31, 2012, with a particular emphasis on highly speculative trading strategies. While this is not the only subreddit dedicated to investing strategies (i.e., r/Investing, r/Personalfinance, r/Stocks, etc.), we focus on this particular subreddit for three primary reasons. First, with over 13 million subscribers, it is by far larger than other finance-related subreddits. Second, it is the subreddit that has recently attracted significant media and regulatory attention for its role in the GameStop short squeeze and ensuing trading frenzies targeting meme stocks. The conventional view is that this forums' userbase is predominantly unsophisticated retail investors who are more interested in gambling than investing.[6] There has also been significant concern that the "research" on *WSB* is at best uninformative, and at worse, a force that destabilizes stock prices and contributes to significant retail trading losses. Lastly, as we discuss in greater detail in the next section, *WSB* differs significantly from other social finance sites (e.g., *Seeking Alpha*), which suggests that prior work on social finance may not apply to *WSB*.

2.2. A Comparison of WSB to Seeking Alpha (SA)

WSB shares important similarities with other social finance platforms such as *Seeking Alpha* (*SA*). Both sites allow non-professional investors to share their investment research, and both sites allow readers to provide comments on the report and engage in discussions with other users. Prior work on *SA* suggests that a large fraction of *SA* reports and contributors are skilled (see, e.g., Chen et al., 2014; Farrell, Jame, and Qiu, 2020). However, there are several prominent differences between *WSB* and *SA,* many of which suggest that *WSB* research may be less informative than *SA* research.

First, while *SA* employs an editorial team to review all research reports to ensure quality there is very limited quality control on *WSB*. Relatedly, *WSB* allows users greater anonymity than *SA*.[7] Greater anonymity reduces the incentives to develop a strong reputation and potentially allows users with more nefarious motives (e.g., pump and dump schemes) to switch identities without accountability. *WSB* reports also tend to be considerably less in-depth than the average *SA* report, and the userbase of *WSB* is likely to have less financial sophistication.[8] Anecdotal evidence suggests that *WSB* also places a larger emphasis on highly speculative

[6] For example, William Gavin, suggested suspending trading in GameStop because "unsophisticated investors are probably going to get hurt by this", (https://www.cnbc.com/2021/01/27/gamestop-speculation-is-danger-to-whole-market-massachusetts-regulator.html), and John Coffee of Columbia Law School describes *WSB* users as a "mob of uninformed, unsophisticated retail traders" https://qz.com/1965494/are-wallstreetbets-reddit-traders-manipulating-gamestop-shares/.

[7] Seeking Alpha allows users to contribute using pseudonyms, but they still require private disclosure of their identities to Seeking Alpha, and they do not allow the same user to post under multiple pseudonyms (https://seekingalpha.com/page/policy_anonymous_contributors).

[8] With respect to article depth, we find that the average *WSB* report in our sample is 352 words, which is roughly half of the length of a typical *SA* report (675 words), as reported in Chen et al. (2014). With respect to investor sophistication, the average Seeking Alpha user has a household income of $321,000 and roughly $1.5 million in investable assets (see https://static.seekingalpha.com/uploads/pdf_income/sa_media_kit_04_2020.pdf). While these figures are unknown for *WSB* users, anecdotal evidence suggests that these estimates would be substantially smaller.

trading strategies. As a result, investment research on the site may gravitate towards strategies that tend to earn lower expected returns such as buying stocks with high volatility (Ang et al., 2006) or stocks with lottery-like features (Bali, Cakici, and Whitelaw, 2011).

Finally, in contrast to *Seeking Alpha* which has steadily grown over the past 15 years, much of *WSB* growth is attributable to the GME short squeeze event. For example, Figure 1 shows the forum grew from 500,000 users in July of 2018 to 10.7 million users as of June 2021, with a clear spike during the GameStop short squeeze in January of 2021. One concern is the dramatic increase in new users, most of whom were attracted to the site by the extraordinary price increases in Gamestop, can have a profound shift on the culture of the site. For example, given investors tendency to forecast expected returns based on recent performance (e.g., Greenwood and Shleifer, 2014), new users may overestimate the effectiveness of strategies that are unrelated to fundamentals, such as price pressure induced short squeezes. Indeed, ample anecdotal evidence suggests that new members of *WSB* tend to emphasize coordinated buy-and-hold strategies for a handful of meme stocks with little regard to the company's fundamentals.[9]

Given the significant shift in both the userbase and culture of *WSB* following the GME event, our analysis will separately examine *WSB* reports issued in the pre-GME period and post-GME period. We define the GME event as occurring on January 13, 2021 because this was the first day when GME experienced a dramatic increase in returns and trading volume (see Figure IA.1 for daily GME trading volume from December 2020 through January 2021). Accordingly, we define the pre-GME period as July 2018 – January 12, 2021, the post-GME period as January 14, 2021 – June 2021, and we exclude DD reports issued on the event day itself. We also confirm that our main results are robust to excluding the week prior to and after the GME event day (see Table IA.1).

[9] For a summary of these competing views see: https://www.insider.com/wallstreetbets-reddit-forum-divided-as-new-users-flood-subreddit-2021-2.

2.3 The WSB Sample

We scrape all posts on *WSB* from July of 2018 through June of 2021 using the Pushshift API, which collects new posts in almost real-time.[10] Posts can be deleted by the original author, moderator of the subreddit, or an "automod" (which is a spam filter robot operated and constructed by moderators). Deletions by the automod typically occur in less than a minute. Deletions of posts by moderators take slightly longer if the post breaks the rules of the subreddit and was not already captured by the automod. Lastly, a post can be deleted by the author at any time. Importantly, the API retains posts deleted by authors, and these posts are included in our sample.

WSB contains more than 100,000 different posts spanning several different categories including: *News* (links to news stories *WSB* users found interesting), *Discussion* (open-ended discussions, frequently on macroeconomic forces such as proposed regulations, supply chain disruptions, etc.), *YOLO* (posts reporting large upcoming bets), *Gains/Losses* (posts highlighting major investment successes and failures), *Sh$tposts* (ironic investment theses that are meant to entertain rather than inform), and *Due Diligence* reports (posts that contain investment analysis and a clear investment recommendation). The contributor of the post selects the category, and the classification is immediately known to user participants. However, posts that are incorrectly classified by users, such as a DD report that does not provide any new investment research, are deleted by moderators/bots very quickly (generally within 2-3 minutes of posting). We limit the sample to posts that have more than one upvote (which is automatic from the user). This filter should minimize the impact of posts that were immediately removed by moderators/bots because they failed meet the criteria for a specific category.

Our analysis focuses primarily on Due Diligence (DD) reports. These reports are vetted by the moderators as containing information where 1) at least some analysis has been performed and 2) the author provides a clear investment recommendation (long or short). DD reports are most similar to other forms of social-finance investment research (e.g., *Seeking Alpha* research reports) that have been studied in the prior literature, and they are the most likely to contain value-relevant information.

[10] There is a period between April 13th and August 4th of 2020 where DD reports are missing. This is likely due to an issue with Reddit's API.

We limit the sample to DD reports focused on a single ticker (e.g., we eliminate DD reports that focus on market-wide or industry analysis) and to common stocks (CRSP share codes 10 and 11) with available data in the CRSP-Compustat merged database. For each DD report, we manually review the report to identify the investment recommendation and ticker. Although the author's investment recommendation is clear to anyone reading the report, there is no standardized format for listing the recommendations which necessitates a manual review of each report. The manual review of tickers is also needed for two reasons. First, users may place special characters before or after a ticker symbol that a program would misclassify. Second, users sometimes intentionally report a wrong ticker to misdirect hedge funds and other institutional investors that monitor message boards using algorithms.[11]

Appendix A provides an example of a DD report in our sample. A manual reading of the report indicates that the author is recommending a "Buy" for "BYND." The header of the report also includes the username and the timestamp of the report. For DD reports that occur outside of trading hours, we set the date of the report equal to the date on which an investor could have first traded on the report.[12]

Panel A of Table 1 provides summary statistics. The sample includes 5,015 DD reports covering 3,782 firm days and 906 different firms. The overwhelming majority of DD reports (88%) are buy recommendations. We also find the average contributor, as measured by the username on the DD report, issues only 1.32 reports during the sample period. These estimates likely significantly understate the number of reports per person since users often get temporary bans for violating moderator rules and circumvent the ban by joining the forum with a different username. Nevertheless, the lack of repeated posts by the same username suggests that *WSB* users are not especially concerned with developing and maintaining a reputation.

Summary statistics for the pre- and post-GME period are presented in the second and third rows of Panel A. Although the post-GME period is substantially shorter in calendar time, it accounts for a slight

[11] For an example of *WSB* users attempting to mislead hedge funds, see: https://www.reddit.com/r/wallstreetbets/comments/ly0d4m/how_to_beat_hedge_fund_algorithms_on_wsb_a/

[12] For example, if a report was issued at 5 pm on Wednesday, January 6, we would classify the date of the report as Thursday, January 7, and we would define the [1,5] day return as the return from Friday, January 8 through Thursday, January, 14. We exclude the Day [0] return to reduce the impact of potentially confounding news that could influence both the DD report and the Day [0] return.

majority (~51%) of all DD reports. DD reports in the post-GME are more likely to recommend a long position (95% versus 81%), and they are substantially more likely to discuss GME or AMC (19.6% versus 4.2%). The substantial differences in report characteristics in the pre- and post-GME period are consistent with the GME event resulting in a significant shift in the culture of *WSB*. We also provide summary statistics of the distribution of DD report coverage at the firm-month level. The average firm has only 0.04 DD reports per month, although there is considerable dispersion in the intensity of coverage, as evidence by the relatively large standard deviation (0.80).

We contrast *WSB* DD reports with non-research posts. We define a *WSB* post as non-research related if it belongs to one of the following categories: *News*, *Losses*, *Gains*, *Charts*, and *Shi$posts*. We focus on these categories because they are not designed to contain any new value-relevant information and they do not provide a clear directional recommendation (i.e. buy versus sell).[13] For example, *News* include links to articles and does not provide any analysis or interpretation of the news, *Gains* and *Losses* report previous successful and unsuccessful investments, *Charts* are typically graphs of past returns that could be found on any website, and *Shi$posts* are typically satirical posts that are not intended to be taken seriously. Panel B reports summary statistics for the non-research post sample. The sample includes 13,255 non-research posts, 57% of which are in the post-GME period. Non-research posts are much more likely to be on GME or AMC. In fact, in the post-GME period more than 70% of all non-research posts are on GME and AMC.

2.4 Other Variable Construction

We collect *Seeking Alpha* research reports over the same sample period (July 2018-June 2021). For each report, we collect the following information: a report ID assigned by *Seeking Alpha*, the date and time of the publication, the ticker (or tickers) assigned to each report, the author of the report, and the authors stated recommendation (e.g., "bullish", "neutral" or "bearish"). To parallel the *WSB* sample, we convert *SA* recommendations to a "buy" indicator, which equals one if the recommendation is "very bullish" or "bullish"

[13] We exclude *YOLO* (you only live once) and *Discussion* posts. Although *YOLO* posts, which simply detail large bets, do not provide any investment research, if *WSB* users are informed, then their directional trades could still provide value-relevant signals. Similarly, although *Discussion* posts do not provide a clear investment recommendation, they could provide useful contextual information. In untabulated results, we find that classifying *YOLO* and *Discussion* posts as non-research related posts yields similar results.

and zero otherwise. We exclude reports that do not have a stated recommendation, and we limit the sample to reports that are focused on single-ticker articles for US common stocks.

Panel C of Table 1 reports summary statistics for the *Seeking Alpha* sample. The sample of *SA* reports is much larger than DD reports (23,659 versus 5,015), and the *SA* sample spans a larger cross-section of firms. *SA* research also tends to be overwhelming bullish, with 85% of all reports being classified as buy recommendations. The average *SA* contributor writes 10.98 reports compared to only 1.32 reports for *WSB*. This finding is consistent with anecdotal evidence that contributors on *SA* are more interested in using the site to build a reputation.[14] Even in the post-GME period, less than 1% of all *SA* reports cover GME or AMC. This is consistent with the GME event having a less pronounced impact on *SA* research relative to *WSB* research.

We combine the data on social media research from *WSB* and *SA* with several additional data sources. We obtain financial statement data, including book value of equity, book value of debt, book value of assets, short interest, and total common shareholders from Compustat. We obtain financial market data, including daily data on share price, shares outstanding, volume, and stock returns from CRSP. Earnings announcement dates and sell-side analyst earnings forecast data are from the I/B/E/S unadjusted US detail history file. We collect the number of shares held by institutions from the Thomson Reuters Institutional Holdings database, and media coverage is collected from Bloomberg.

We identify retail trading from TAQ data using the approach of Boehmer, Jones, Zhang, and Zhang (2021) (hereafter BJZZ). Specifically, we classify trades with TAQ exchange code "D" and prices just below a round penny (fraction of a cent between 0.6 and one) as retail purchases, while trades with exchange code "D" and prices just above a round penny (fraction of a cent between zero and 0.4) are classified as retail sales. This classification is conservative in the sense that it has a low type 1 error (i.e., trades classified as retail are very likely to be retail). However, this classification omits retail trades that occur on exchanges as well as limit orders that are not immediately executable.

[14] For example, many of the testimonials of *Seeking Alpha* contributors emphasize the important reputational benefits associated with being a regular contributor on *Seeking Alpha* (https://seekingalpha.com/page/testimonials).

2.5 Determinants of WSB Coverage

In this section, we examine the determinants of *WSB* coverage. We expect that many of the firm characteristics that influence research coverage on other social finance sites (e.g., *Seeking Alpha*) are likely to be relevant on *WSB* as well. However, relative to *SA*, we expect that *WSB* users will tend to issue reports on more speculative stocks, and we expect that such effects may be amplified in the post-GME period.

We examine the determinants of coverage by estimating the following panel regression:

$$Coverage_{it} = \alpha + \beta_1 Chars_{it-1} + \beta_2 Chars_{it-1} \times Post_t + Month_t + \varepsilon_{it.} \quad (1)$$

The dependent variable, *Coverage*, is either equal to *WSB DD Coverage*, defined as the natural log of 1 plus the total number of DD reports issued for firm *i* during month *t*, *WSB Non-Research Coverage* defined as the natural log of 1 plus the total number of non-research reports issued for firm *i* during month *t,* or *SA Coverage*, defined as the natural log of 1 plus the total number of *SA* reports issued for firm *i* during month *t. Chars* contains the vector of firm characteristics used in Farrell et al. (2022) to explain *SA* coverage, namely the percentage of the firm's shares held by institutional investors at the end of the prior year (*Inst. Ownership*), the number of common shareholders (*Breadth of Ownership*), market capitalization (*Size*), book-to-market ratio (*BM),* return volatility (*Volatility*), share turnover (*Turnover*), past one-month returns ($Return_{m-1}$), past returns over the prior two to twelve months ($Ret_{m-2,\ m-12}$), the number of unique media articles mentioning the firm in the prior year (*Media Coverage*), and the number of sell-side analysts issuing a forecast for the firm in the prior year (*IBES Coverage*). In addition, given the ample anecdotal evidence that *WSB* users target stocks with lottery like features, stocks with heavy short interest, and stocks that recently went public, we add indicator variables equal to one if the firm is in the top quintile of the maximum daily return in the previous month (*High Max)*, the top quintile of short interest in the previous month (*Heavy Short*), or if the firm went public in the past six months *(Recent IPO*). Finally, given the large percentage of reports issued on GME and AMC, we include a separate GME/AMC indicator. We allow the coefficient on all the characteristics to vary in the pre- and post-GME period by interacting the characteristics with *Post,* an indicator equal to one in the post-GME period (February – June of 2021) and zero for the pre-GME period (July 2018-December 2020), and we exclude January 2021 since it spans both the pre- and post-GME period. We include month fixed effects and cluster standard errors by firm and month. We log all continuous variables other than past returns, and we standardize all continuous variables, including the dependent variables, to have zero mean and unit variance.

Specification 1 of Table 2 reports the results. Consistent with prior work on the determinants of *SA* coverage, we find that *WSB* coverage in the pre-GME period is increasing in firm size, volatility, and media coverage, and decreasing with institutional ownership. We also find that *WSB* coverage is significantly greater for lottery-like stocks (i.e., *High Max*), stocks with high short interest, stocks that recently went public, and GME/AMC. The coefficients on the post-GME interaction terms indicate that *WSB* coverage of speculative stocks, including stocks with higher volatility, higher max returns, higher shorter interest, recent IPO stocks, and GME/AMC are significantly greater in the post-GME sample. The magnitudes are economically large. For example, *WSB* coverage of stocks with high max returns increases by nearly 500% in the post-GME period (from 0.03 to 0.17), and we also observe a large increase in the coverage of stocks with high short interest and the coverage of GME/AMC. These dramatic increases are consistent with the extreme GME returns attracting even more speculative investors.

Specification 2 reports the determinants of *WSB Non-Research Coverage*. The findings in the pre-GME period are generally similar to the results for *DD Report Coverage*. However, the shift in *Non-Research Coverage* in the post-GME period is primarily driven by an increased emphasis on GME/AMC rather than a more general shift towards other more speculative stocks.

In Specification 3, we examine the determinants of *SA* coverage. Perhaps surprisingly, in the pre-GME period, we find that *SA* and *WSB* have similarly strong preferences for lottery-like stocks, stocks with heavy short-interest, and stocks that recently went public. However, in sharp contrast to *WSB DD Coverage*, *SA* coverage of speculative stocks does not dramatically increase in the post-GME period. In fact, *SA*'s coverage of volatile stock and stocks with heavy short interest declines in the post-GME period. The findings suggest that the effects of the GME event were not uniform across all social media platforms.

3. The Informativeness of *WSB* Research

3.1 WSB Research and the Cross-Section of Stock Returns

In this section, we examine whether DD report recommendations forecast future stock returns. Our null hypothesis is *WSB* reports are unable to forecast returns. This hypothesis is consistent with either *WSB* reports containing no useful information, or the market immediately incorporating all value-relevant information on the day of the report.[15] In contrast, it is possible that *WSB* reports could positively predict returns. While this prediction is inconsistent with efficient markets, it is in line with a large literature that documents longer-horizon predictability following investment research including drifts following sell-side analyst recommendations (Womack, 1996), Seeking Alpha research reports (Chen et al., 2014), and the tweets of professional investors on StockTwits (Cookson and Niessner, 2020). Finally, a more pessimistic view is that *WSB* either reflects or causes investor sentiment, in which case *WSB* reports could negatively predict future returns.

We estimate the following panel regression:

$$\begin{aligned} R_{it+1,t+x} = {} & \beta_1 Net\ DD_{it} + \beta_2 Net\ DD_{it} \times Post_t + \beta_3 NonResearch_{it} \\ & + \beta_4 NonResearch_{it} \times Post_t + \beta_5 Net\ SA_{it} + \beta_6 Net\ SA_{it} \times Post_t \\ & + Controls_{it} + Day_t + \varepsilon_{it}. \end{aligned} \tag{2}$$

The dependent variable is the stock return measured over the subsequent week (i.e., $x = 5$ trading days) or the subsequent month ($x = 21$ trading days). *Net DD* is the number of buy DD recommendations for stock *i* on day *t* less the number of sell DD recommendations for stock *i* on day *t*, and *Net DD*× *Post,* interacts *Net DD* with an indicator equal to one for the post-GME period. Thus, *Net DD* captures the average predictive ability of DD reports in the pre-GME period, and *Net DD × Post* captures the incremental predictive ability of DD reports in the post-GME period.

NonResearch is the is the number of non-research related postings on *WSB* for stock *i* on day *t, Net SA* is the number of *SA* reports issuing a buy recommendation for stock *i* on day *t* less the number of *SA* reports

[15] As emphasized in footnote 12, our primary analysis excludes the Day [0] return of the DD report. Thus, if the market immediately responds to the information in DD reports, we should not observe any future predictability.

issuing a sell recommendation for stock *i* on day *t, NonResearch* × *Post* interacts *NonResearch* with the post-GME indicator, and *Net SA* × *Post* is defined analogously. We winsorize *Net DD*, *NonResearch*, and *Net SA* at the 1st and 99th percentile of the distribution of firm-days where the variable is not equal to zero. Following Kelley and Tetlock (2013), the controls include *Size, Book-to-Market*, returns measured from days [0], [-5, -1], and [-26, -6] and media sentiment measured similarly. See Appendix C for detailed definitions. *Day* denotes calendar-day fixed effects. To account for the correlation in the residuals induced by the overlapping holding periods, we cluster standard errors by both firm and month.

Specifications 1 and 2 of Table 3 report the results for one-week and one-month holding periods for the full sample, and Specifications 3 and 4 report analogous results after excluding GME and AMC. Across all four specifications, the coefficient on *Net DD* is positive and statistically significant. The economic magnitudes are also sizeable. For example, an incremental buy DD report issued in the pre-GME period is associated with a 1.11% increase in one-week ahead returns and a 5.17% increase in one-month ahead returns. The extraordinary returns of GME and AMC contribute to the sizeable magnitudes, however the predictability of *WSB* reports is not driven purely by these two firms. In particular, after excluding GME and AMC, an incremental buy DD report issued in the pre-GME period is associated with a 0.91% increase in one-week ahead returns and a 2.33% increase in one-month ahead returns. While this finding is consistent with prior work that suggests that crowdsourced investment research can forecast future returns, it is perhaps surprising that this relation continues to hold in a setting with complete user anonymity, minimal oversight, and limited reputational incentives.

In contrast, the coefficient on *Net DD* × *Post* is significantly negative at the one-month holding period. In particular, in the full sample, the predictive ability of DD reports declines by 5.30% (to -0.13%), and in the sample that excludes GME and AMC the estimate declines by -3.45% (to -1.12%). Both the -0.13% and the -1.12% estimate are not significantly different from zero. Thus, the ability of *WSB* reports to forecast returns is completely eliminated in the post-GME period. We more deeply explore the factors driving the time-series decline in the predictability of *WSB* reports in Section 5.

The coefficient on *WSB NonResearch* is statistically insignificant, indicating that non-research *WSB* posts do not forecast returns in the pre-GME period. We also confirm that after excluding GME and AMC, the coefficient on *Net DD* is significantly larger than the coefficient on *NonResearch*. This finding alleviates the concern that any *WSB* post, independent of its content, can predict returns simply because it correlated with naïve investor demand. We also find some evidence that the informativeness of non-research reports declines in the post-GME period, and in Table IA.2 of the Internet Appendix, we confirm that the informativeness on non-research posts in the post-GME period (i.e., *WSB NonResearch* + *WSB NonResearch* × *Post*) is significantly negative after excluding GME and AMC. This finding is consistent with non-research posts in the post-GME period reflecting naïve investor sentiment that is associated with overpricing and subsequent reversals.

The coefficient on *Net SA* is positive and significant. This finding is consistent with prior work suggesting that *SA* research is informative (e.g., Chen et al., 2014; Dim, 2021). However, the estimates on *Net SA* are smaller than the estimates on *Net DD*, although the differences in the coefficients are significantly different from each other in only one of the four specifications. Nevertheless, at a minimum, the findings suggest that *WSB* research was at least as informative as *SA* research in the pre-GME period.[16] Finally, we find no evidence that the informativeness of *SA* research declined in the post-GME period. This finding is inconsistent with the view that broad macroeconomic forces contributed to a widespread decline in the informativeness of investment research across all social media platforms.

Lastly, we note that the coefficients on the control variables are consistent with prior work. In particular, the negative coefficient on prior day and prior week returns is in line with prior work on short-term return reversals (e.g., Nagel, 2012; Jame, 2018), and the positive coefficient on prior day media sentiment echoes the findings in Tetlock, Saar-Tsechanksy, and Macskassy (2008). The coefficient on the other control variables,

---

[16] In the Internet Appendix, we explore whether differences in risk can account for the somewhat stronger predictability of *WSB* research relative to *SA* research in the pre-GME period. We find that *WSB* recommendations are stronger predictors of future volatility than *SA* report recommendations (Specifications 1 and 2 of Table IA.4), and *WSB* superior performance relative to SA is eliminated when considering volatility-adjusted performance (Specifications 3 and 4 of Table IA.4) or when focusing on the left-tail of the distribution using quantile regressions (Table IA.3). However, the superior performance of *WSB* recommendations is robust to controlling for common measures of systematic risk such as size, book-to-market, and momentum (Specifications 5 and 6 of Table IA.4).

including size and book-to-market, exhibit no significant relation with returns, which is consistent with an insignificant size and value premium in more recent times (Smith and Timmerman, 2022).

3.2 WSB Research and the Cross-Section of Stock Returns – Robustness and Alternative Horizons

In Table IA.1 of the Internet Appendix, we examine whether the findings reported in Table 3 are robust to different research design choices. We briefly summarize the results below and offer a more detailed discussion in the Internet Appendix. In particular, we confirm that the results are similar if we 1) exclude the "Robinhood 50" stocks, defined as the 50 stocks that Robinhood imposed trading restriction on beginning on January 28th, 2021 and ending February 5th, 2021, from the sample; 2) interact all the control variables with a *Post* indicator; 3) exclude the five days prior to and after the GME-event or 4) replace *Net DD* with three indicator variables (*Heavy Buy, Light Buy*, and *Sell).* We also examine the relation between *Net DD* and stock returns over longer horizons. Specifically, we construct a monthly measure of *Net DD* by aggregating all reports over the calendar month, and then examine the one-month ahead return predictability. We find that the magnitudes of the estimates decline (in absolute value) which is consistent with the predictability of DD reports being stronger over shorter horizons. However, aggregating over longer windows also results in more precise estimates, and as a result, the statistical significance of the estimates remains similar.

We next examine the predictive ability of *WSB* reports for various alternative holding periods. Specifically, we repeat Equation (2) for the current day (i.e., Day 0), each of the first five days, weeks 2, 3, and 4, and weeks 5 through 12. Panel A of Table 4 reports the estimates on *Net DD, Net DD × Post GME,* and *Net DD + Net DD × Post GME* for the full sample, and Panel B reports the results after excluding GME and AMC.

In Panel A, we observe a significant positive relation between *Net DD* and Day 0 returns. In particular, one additional buy recommendation is associated with a 0.88% increase in same-day returns. We continue to find positive estimates ranging from 0.10% to 0.33% over the five days after the DD report, and most of these estimates are at least marginally significant ($p < 0.10$). We also observe that the returns continue to drift upwards over the subsequent 60 trading days. The absence of a return reversal is inconsistent with short-term price pressure being the primary driver of the informativeness of *WSB* research in the pre-GME period (additional price pressure tests in Section 4.3 reinforce this finding). Consistent with the return predictability declining in the post-GME period, the coefficient on *Net DD* × *Post* is negative in nine of the ten holding periods considered. Although the individual estimates are typically insignificant, as we show in Table 3, the cumulative estimates for the one-month holding period are significantly negative, and this effect does not reverse over the subsequent two months. The estimates of the combined coefficient (i.e., *Net DD* + *Net DD* × *Post*) are typically insignificantly different from zero. The results that exclude GME and AMC (Panel B) are generally smaller in magnitude but yield qualitatively similar conclusions.

To better visualize the cumulative returns, we also estimate Equation (2) for horizons ranging from one-week (i.e., $x = 5$) through 12 weeks (i.e., $x = 60$), and then plot the estimates. Figures 2A and 2B report the results for the full sample and the sample that excludes GME and AMC. Consistent with Table 4, we see that the predictive ability of *WSB* reports in the pre-GME period does not reverse over longer horizons. In addition, the decline in the predictive ability of DD reports in the post-GME period remains sizeable over longer horizons. For example, at the end of 12 weeks, the coefficient on *Net DD* × *Post* is -7.11% for the full sample and -3.91% for the sample that excludes GME and AMC.

4. Why Did *WSB* Reports Predict Returns in the Pre-GME period?

The results from the previous section suggest that *WSB* research reports were able to forecast returns in the pre-GME period. In this section, we seek to better understand the forces contributing to this return predictability.

4.1 The Role of the Pandemic

A significant portion of the pre-GME period coincides with the Covid-19 pandemic. Prior work finds that retail trading increased substantially during the pandemic (e.g., Ozik, Sadka, and Shen, 2021). Stay-at-home orders and the cancellation of sports and other forms of entertainment resulted in many retail investors having extra time that could potentially be devoted to understanding financial markets. Reduced spending and stimulus checks may have also provided investors a catalyst to trade financial securities. In addition, lockdowns limited institutional investors ability to meet privately with management, which may have reduced the informational advantage of institutional investors relative to retail investors (Bai and Massa, 2022; Bradley, Jame, and Williams, 2022). Thus, a natural question is whether the informativeness of DD reports in the pre-GME period is primarily attributable to the unusual pandemic period.

To test this idea, we partition the pre-GME period into *Pre-Pandemic* (July 2018 - March 15, 2020) and *Post-Pandemic* (March 16, 2020 - January 12, 2021).[17] The use of March 15 as the start of the pandemic follows Ozik, Sadka, and Shen (2021). It also results in a similar number of pre-GME DD reports in the pre-pandemic and post-pandemic periods (1,047 and 1,396 reports, respectively). We then repeat Equation (2) after partitioning *Net DD* into *Net DD* × *Pre-Pandemic* and *Net DD* × *Post-Pandemic*. We also perform analogous partitions for *NonResearch* and *Net SA*.

Table 5 reports the results. We find that the estimates on *Net DD* × *Pre-Pandemic* are statistically significant. In fact, after excluding GME and AMC, the point estimates on *Net DD* × *Pre-Pandemic* are larger (albeit insignificantly so) than the estimates on *Net DD* × *Post-Pandemic*. This evidence suggests that the informativeness of DD reports in the pre-GME period is not purely driven by the pandemic period.

To more generally understand how the predictive ability of *WSB* reports varies over the time series, in Figure 3, we estimate Specifications 2 and 4 of Table 3 over the following intervals: the-pre 2020 sample, quarters 1 and 2 of 2020, and each of the remaining quarters in 2020 and 2021. We combine the pre-2020 sample because there are a relatively small number of DD reports (606) prior to 2020, we combine quarters 1 and 2 of 2020 because of the small sample of reports in Q2 of 2020 due to missing data from the API, and we include the small number of pre-GME DD reports issued in January of 2021 in the Q4 2020 estimates. After excluding GME and AMC, the predictive ability of DD reports was stable in 2020, with point estimates ranging from 2.14% to 3.92%. We also observe a sharp decline in informativeness in Q1 of 2021 (-1.57%), which remained negative in Q2 of 2021 (-0.36%). Including GME and AMC yields similar results except that the return predictability in Q4 of 2020 is substantially larger (7.49%) due to the very large returns of GME in January of 2021.

[17] Post-pandemic is also equal to one for the post-GME period. Thus, the coefficient on *Net DD* × *Post* now measures the change in informativeness in the post-GME period relative to the post-pandemic portion of the pre-GME period.

Collectively, the results from this section suggests that informativeness of *WSB* was stable in the pre-GME period. In particular, it was of similar magnitude in both the pre-pandemic and post-pandemic periods, and it was economically sizeable for every quarter in 2020. One important caveat, however, is that the pre-GME time series is relatively short, and even the pre-pandemic period may not be representative of a "normal" environment. Given the short time-series of *WSB* reports, we are unable to assess whether the informativeness of *WSB* reports in the pre-GME period would generalize to other market environments. [18]

4.2 Information Processing versus Information Production

We next explore the economic channel underlying the investment value of *WSB* reports in the pre-GME period. If the return predictability following DD reports is primarily a consequence of DD reports piggybacking off of other news events (e.g., Altinkilic and Hansen, 2009) or skillfully interpreting public news (e.g., Engelberg, Reed, and Ringgenberg, 2012), then we would expect the results to be significantly stronger for reports that coincide with major information events (i.e., *Information Processing*). On the other hand, if *WSB* users independently produce novel information, then the return predictability results may be stronger for reports not issued during major news events (*Information Production*). While both channels are potentially valuable to users who rely on *WSB* for investment research, distinguishing these explanations provides insight into the source of *WSB* investment value in the pre-GME period.

We classify a report as *information processing* if the firm had an earnings announcement, analyst revision, or abnormally high media coverage (as defined in Appendix C) on the day prior to the DD report or the day of the DD report. Roughly 40% of all DD reports are categorized as *information processing*, with the remaining 60% of reports classified as *information production*. We then repeat Equation (2) after partitioning *Net DD* into *Net DD Processing* and *Net DD Production*. Table 6 reports the results. In the pre-GME period, we find that the coefficients on both *Net DD Processing* and *Net DD Production* are always positive, and the estimates are both

[18] In contrast to *WSB*, Seeking Alpha offers a relatively long time series, which allows us to explore whether the pre-GME period is unusually informative for *SA* reports. We find that the informativeness of *SA* reports in the pre-GME period is very similar to their average informativeness over the January 2005 through June 2018 sample period (see Table IA.5 of the Internet Appendix for additional details).

significantly different from zero for the one-month horizon. The evidence suggests that both information processing and information production contribute to the predictive ability of *WSB* in the pre-GME period.

4.3 Information versus Price Pressure

The existing evidence is consistent with DD reports issued in the pre-GME period containing value-relevant information that is impounded into prices over the subsequent month (*information*). However, an alternative view is that DD reports cause (or are correlated with) uninformed demand shocks that induce significant price pressure over the subsequent month (*price pressure*). The lack of reversal over the 12-week holding period is inconsistent with a temporary price pressure explanation, but it is still possible that *WSB* induces price pressure that persists for even longer holding periods.

To more directly test whether DD reports contain value-relevant information, we examine whether DD reports forecast cash-flow news over the subsequent week or month. Specifically, we repeat Equation (2) after replacing returns with one of three proxies for cash-flow news. The first measure is *Media Sentiment* obtained from Bloomberg. Specifically, for each firm day, Bloomberg assigns a sentiment score ranging from -1 (very negative news) to 1 (very positive news), with a median value of 0 (neutral articles). We assign firms with no media coverage a value of 0, and we sum the daily media sentiment over the five-day or 21-day holding period. Our second measure is *Positive Forecast Error*, which equals one if realized earnings exceed the median quarterly forecast across all I/B/E/S analysts as of day *t,* and zero otherwise. The five-day (21-day) sample is limited to firms that will announce earnings within five (21) trading days of day *t*, and we also require that the firm have at least one I/B/E/S earnings forecast. While *Positive Forecast Error* is a common proxy for cash-flow news (e.g., Kelley and Tetlock, 2013), one limitation is that it restricts the sample to firms that will shortly announce earnings. As a broader measure of earnings-related news, we also compute *Positive Forecast Revision*, which equals the total number of upward revisions scaled by the total number of revisions. In computing this measure, we consider both quarterly and annual earnings forecast revisions. We exclude firms with zero I/B/E/S coverage, and we set *Positive Forecast Revision* to 50%, the median value across the sample, for firms with I/B/E/S coverage but no forecast revisions over the holding period.[19]

[19] The results are robust to excluding all firm with zero forecast revisions.

Table 7 reports the results for the full sample.[20] In all six specifications, the estimates on *Net DD* are positive and three of the six estimates are statistically significant. The economic magnitudes are also sizeable. For example, the estimate in Specification 3 indicates that an incremental buy DD recommendation issued within five days of the earnings announcement is associated with a 4.22% percentage points increase in the likelihood of beating the sell-side earnings consensus forecast, which corresponds to roughly a 7% increase relative to the sample mean of 60%. These findings are consistent with DD reports containing value-relevant information related to a firm's future cash flows. An alternative view is that DD reports themselves cause changes in the cash-flow measures. The latter view is perhaps most plausible for media sentiment (i.e., news coverage is either influenced by DD reports or by the price increase caused by DD reports); however, it seems unlikely that DD reports will directly impact earnings over the subsequent week or month.[21] Thus, we believe the totality of evidence suggests that *WSB* reports in the pre-GME period contain value-relevant information. In contrast, we find no evidence that *WSB* reports in the post-GME period predict cash-flow news. Further, in the case of *Positive Forecast Error*, the post-GME estimate (i.e., *Net DD* + *Net DD* × *2021*) is significantly less than zero indicating that *WSB* reports in the post-GME period are negative predictors of forecast revisions.

5. Why Did Return Predictability of *WSB* Reports Decline in the Post-GME Period?

In this section, we explore factors that may have contributed to the decline in the informativeness of *WSB* research in the post-GME period. Our primary conjecture is that the success of the GME event altered the composition and culture of *WSB* research. As Figure 1 shows, *WSB* experienced a nearly 10-fold increase in the number of users during January of 2021. The shift in the userbase naturally altered the composition of DD contributors. In fact, we find that only 176 of the 2,717 post-GME DD reports (~6.5%) were written by contributors who issued a DD report in the pre-GME period.[22] In Sections 5.1 and 5.2 we document that the

[20] Neither GME nor AMC have extreme measures of cash-flow news, so excluding them from the analysis has a negligible impact on the results.

[21] This is not to suggest that *WSB* reports can never influence earnings. For example, in response to higher prices several meme stocks including AMC, GME, and American Airlines were able to raise capital at attractive prices, which naturally impacted the firms' investment decisions and subsequent cash flows. However, we would expect these potentially important "real effects" to manifest only over longer horizons.

[22] In Table IA.6 of the Internet Appendix, we compare the informativeness of post-GME DD reports issued by new versus existing contributors. The point estimates are consistent with post-GME reports issued by existing contributors

shift in the userbase also results in significant shifts in contributor strategy, and in Section 5.3 we link the shifts in strategy to the decline in *WSB* return predictability.

5.1 Shift in Contributor Strategy – Price Pressure versus Fundamentals

While the impact of the GME event on the culture of *WSB* is likely far-reaching and multifaceted, anecdotal evidence suggests that a particularly important change was that the site became more focused on identifying potential profit opportunities due to coordinated price pressure strategies, possibly because the massive (and salient) success of the GME short-squeeze resulted in upwardly biased expectations of the profitability of this strategy.[23] We begin by exploring whether there is an increase in *WSB* reports emphasizing coordination and/or price pressure following the GME event. We conduct textual analysis to identify whether the report focuses on strategies related to price pressure. We randomly read approximately 100 DD reports and wrote a list of words that appeared to be related to price pressure strategies (e.g., short selling) or company fundamentals (e.g., earnings). This word list is available in Appendix B. In Table IA.7 of the Internet Appendix, we also examine the relative importance of each word on the price pressure list and the impact each word has on the main results. We find that the three most frequent price pressure words are "squeeze", "short interest", and "float". We also confirm that the main findings are qualitatively unchanged if we exclude any particular word.

We define a report as focusing on price pressure if the number of price pressure words exceeds the number of fundamental words (*PP Report*). As a robustness check, we also classify a report as focusing on price pressure if there is at least one price pressure word in the report (*PP Report2*). To test whether the frequency of price pressure reports increase in the post-GME period, we regress the *PP Report* indicator on a *post-GME* indicator. The results of this analysis, reported in Panel A of Table 8, indicate that the fraction of price pressure

being more informative than the reports by new contributors, however the estimates are not reliably different from each other. Motivated by some of the findings of this paper, Cookson et al. (2022) also examine changes in informativeness of message-level data from StockTwits around the GME-event. They find that the informativeness of messages from new users' declines significantly while the informativeness of messages from established users does not change.

[23] For example, one user laments about the increasing frequency of posts discussing short squeezes here: https://www.reddit.com/r/wallstreetbets/comments/nujffg/not_every_stock_is_a_short_squeeze/.

reports increases by 19.4 percentage points in the post-GME period. This reflects a roughly 165% increase relative to the pre-GME sample mean of 11.6%. Figure 4A plots the fraction of PP Reports by quarter (where the construction of quarters follows Figure 3). We find that the fraction of price pressure reports are relatively stable in the pre-GME period and then jump to 33% and 30% in Q1 and Q2 of 2021, respectively. Specification 2 confirms that the results are robust to excluding GME and AMC, and Specifications 3 and 4 show the results are also very similar if we consider the alternative measure of price pressure reports (*PP Report 2*).

5.2 Increases in Behavioral Biases – Attention Based Trading

Our second conjecture is that composition of contributors on the site shifted towards a less sophisticated userbase, which could result in post-GME reports being more influenced by behavioral biases. One prominent behavioral bias that has be shown to adversely impact retail trading is their tendency to purchase stocks that catch their attention (e.g., Barber et al., 2022). Accordingly, we test whether the composition of DD reports in the post-GME period shifts towards more attention-grabbing stocks.

We consider two proxies for attention. The first is the absolute return on the day prior to the DD report. To facilitate comparison with the *PP Report* indicator variable, we convert the absolute return to an indicator based on whether the absolute return was in the top decile of the distribution (*High Absolute Return*). Our second measure is the number of non-research posts on *WSB* over the previous week. We again convert this to an indicator equal to one if there was more than one non-research report issued on *WSB* over the previous five trading days (*High WSB Posts*). Finally, we consider a composite attention measure, *High Attention*, which is equal to one if either *High Absolute Return* or *High WSB Posts* is equal to one.

Panel B of Table 8 reports the results for *High Attention*. The fraction of DD reports that are classified as high attention increases from 32.5% in the pre-GME period to 57.1% in the post-GME period. Figure 4B indicates a sharp increase in attention-based DD reports beginning in Q1 of 2021, and Specifications 2-4 confirm that the results are robust to excluding GME and AMC or considering either of the attention measures (*High Absolute Return* or *High WSB Posts*) in isolation.

5.3 The Impact of Price Pressure Reports and High-Attention Reports on DD Informativeness

The increase in *PP Reports* and *High-Attention Reports* in the post-GME period could contribute to the decline in report informativeness for at least two reasons. The first view is that *PP* and *High-Attention Reports* are always uninformative (even in the pre-GME period) and thus a higher percentage of these types of reports results in a decline in the average report informativeness. Alternatively, it is possible that *PP* and *High-Attention Reports* become significantly less informative in the post-GME period. Time-varying informativeness of *PP Reports* would be consistent with users overestimating the value of price pressure strategies in the post-GME period, perhaps because contributors extrapolate the investment value of price pressure signals (e.g., the level of short interest) based on the recent success of the GME event. Similarly, the informativeness of *High-Attention Reports* could decline if the dramatic shift in the userbase of *WSB* following the GME event shifted the motivation for issuing reports on high-attention stocks from primarily rational reasons (e.g., skillful interpretation of major news announcements) to primarily behavioral reasons (e.g., attention-induced trading).

To examine whether *PP Reports* (*High-Attention Reports*) are less informative than other reports, we repeat Equation (2) after including Net *DD PP* (*Net DD Attention*) defined as the *Net DD* measure computed for the subset of reports classified as a price pressure (high-attention) report. We also consider a composite measure *(PP/Attention*) which is equal to one if the report is classified as either a *PP Report* or a *High-Attention Report*. We continue to allow the informativeness to vary in the post-GME period by interacting the different measures of *Net DD* with the post-GME indicator.

Specifications 1 and 2 of Table 9 report the results for *PP Reports* before and after excluding GME and AMC, respectively. We find that the coefficient on *Net DD PP* is positive, albeit statistically insignificant after excluding GME and AMC. The positive coefficient is inconsistent with the view that *PP Reports* are always uninformative. In contrast, the coefficient on *Net DD PP* × *Post* is significantly negative, although the estimate is only marginally significant ($p < 0.10$) for the sample that excludes GME and AMC. Specifications 3 and 4 reveal that the informativeness of *High-Attention Reports* declines in the post-GME period. In particular, the coefficient *Net DD Attention* × *Post* is negative and statistically significant at a 1% level in both specifications. Furthermore, the results using the composite measure (*PP/Attention*) reported in Specifications 5 and 6, yield even stronger statistical support. Collectively, the evidence supports the view that the shift in contributor focus to high-attention and price pressure reports contributed to the decline in informativeness in the post-GME period.[24]

An important question is whether price pressure reports and high-attention reports in the post-GME period were simply uninformative or whether such reports are associated with mispricing. In Table 10, we formally test for the significance of *PP/Attention* reports return predictability in the pre-GME period, the post-GME period, and the difference between the two estimates for the sample that excludes GME and AMC (i.e., Specification 6 of Table 9). We find that *PP/Attention Reports* are associated with a 3.87% increase in one-month ahead returns in the pre-GME period, and this effect is reliably different from zero. This effect declines by a significant 5.53% in the post-GME period. The estimate for the post-GME period only (i.e., *PP/Attention Reports* + *PP/Attention Reports* × Post) is -1.67%, which is marginally significant ($p < 0.10$). The one-month return might mask interesting intra-month return dynamics. Accordingly, we also examine the returns for the other holding periods considered in Table 4. In the 11-15 days after the report, *PP/Attention* reports are associated with a -0.54% return in the post-GME period. This estimate is statistically significant, although some caution is needed in interpreting the statistical significance since the tests in Table 10 consider multiple holding periods. We also find that *PP/Attention* reports are associated with a further decline of -1.02% in the fourth week, and an additional -1.08% decline over weeks 5-12, but neither estimate is reliably different from zero. Collectively, the results from Table 10 provides evidence consistent with *PP/Attention Reports* negatively forecasting returns in the post-GME period over certain horizons, although the imprecision of the estimates prevents us from drawing very strong conclusions.

6. *WSB* Reports and Retail Trading

A concern among regulators is that *WSB* induces uninformed trading frenzies that can destabilize prices and potentially harm investors, particularly less sophisticated investors. In this section, we explore this concern by investigating how *WSB* reports correlate with the trading direction and trade informativeness of

[24] In Table IA.8 of the Internet Appendix, we examine whether the results in Table 9 are robust to replacing *PP Report* with *PP Report2* and replacing the *High Attention* measure with either *High Absolute Return* or *High WSB Posts*. The results are qualitatively similar.

investors with differing sophistication levels. In conducting this analysis, we acknowledge that we cannot cleanly distinguish the extent to which *WSB* directly impacts retail trading versus simply proxies for retail sentiment. Nevertheless, the absence of a strong relation between *WSB* research and retail trading would cast doubt on regulators concerns that *WSB* induces trading frenzies, while the presence of such a relation would be consistent with this concern.

6.1. Investor Order Imbalances following DD Reports

We begin by examining the relation between *WSB* reports and investor order imbalances.[25] We consider three groups of investors: small retail investors, large retail investors, and institutional investors. We proxy for small retail traders by equally weighting retail trades, which tend to be dominated by relatively smaller traders, and we proxy for large retail traders by examining retail share volume, which is heavily influenced by large trades. Finally, any trade not classified as retail is classified as an institutional trade. We sign retail trades using the algorithm of BJZZ, and we sign institutional trades using the Lee and Ready (1991) algorithm.

For each investor type, we consider two measures of order imbalances. First, following BJZZ, we define *Percent Imbalance* as the difference between buys and sales (*imbalance)* scaled by the sum of buys and sells. For example, *Inst Percent Imbalance* is defined as the difference between institutional buy share volume and institutional sell share volume scaled by total institutional share volume, and *Large Retail Percent Imbalance* and *Small Retail Percent Imbalance* are defined analogously. While *Percent Imbalance* is commonly used in the literature, extreme values (e.g., *Percent Imbalance* = 1) could be associated with very small trading volume, and presumably relatively small price impact. Accordingly, we construct an alternative measure, *Std Abnormal Imbalance* defined as the *imbalance* on day *t* less the average *imbalance* over days *t-120* through *t-240*, scaled by the standard deviation of *imbalance* estimated over the same six-month estimation window.[26] In contrast to *Percent Imbalance, Std. Abnormal Imbalance* captures the intensity of the imbalance, and is thus arguably a better proxy for the buying

[25] We focus on trading direction, rather than unsigned trading volume, because regulators' concerns have largely focused on price volatility, which is likely induced by extreme order imbalances (e.g., the GME buying frenzy). That being said, reports that are not associated with large order imbalances, but are associated with sizeable increases in retail trading could result in large trading losses for retail investors due to high trading costs (Barber and Odean, 2000; Barber, Lin, and Odean, 2022). Accordingly, we also examine unsigned trading in Table IA.9 of the Internet Appendix.

[26] We skip six-months when computing the benchmark order imbalance to ensure that the post-GME period measures are always benchmarked to the pre-GME period (e.g., GME order imbalances in Q2 of 2021 are benchmarked to GME order imbalances in Q4 of 2020).

frenzies commonly associated with *WSB*.[27] To facilitate comparison across the two measures, we convert both *Percent Imbalance* and *Std. Abnormal Imbalance* into percentile rankings.

We next estimate the following panel regression:

$$Y_{it} = \beta_1 Net\ DD_{it} + \beta_2 Net\ DD_{it} \times Post_t$$
$$+ \beta_3 NonResearch_{it} + \beta_4 NonResearch_{it} \times Post_t + \beta_5 SA_{it} \quad (3)$$
$$+ \beta_6 SA_{it} \times Post_t + Controls_{it} + Day_t + \varepsilon_{it.}$$

The dependent variable is one of the six order imbalance variables (i.e., three investor types × two order imbalance measures), and the key independent variables, *Net DD, NonResearch*, and *SA*, and *Post* are defined as in Equation (2). Controls include a similar set of controls as in Equation (2) with a few modifications. First, we add the lag of all the order imbalance variables measured over the previous trading day, which helps control for persistence in order imbalances (BJZZ, 2021). Second, we exclude the contemporaneous return and contemporaneous media sentiment since they are measured at the same time as investor order imbalances.[28] Third, we include the absolute return and the absolute news sentiment measured over days [-1], [-5, -2], and [-26,-6] to control for attention-grabbing events that could potentially influence imbalances (Barber and Odean, 2008). Standard errors are double-clustered by firm and month.

Table 11 reports the results. Specifications 1-3 report the results for *Percent Imbalance* for institutional investors, large retail investors, and small retail investors, respectively, and specifications 4-6 report analogous results for *Std Abnormal Imbalance*. The results using the two order imbalance measures are typically similar, so the remaining discussion will focus on the *Std Abnormal Imbalance* measure.

In the pre-GME period, we find that *Inst Std Abnormal Imbalance* is uncorrelated with *Net DD*. Both *Large Retail Std Abnormal Imbalance* and *Small Retail Std Abnormal Imbalance* are significantly correlated with *Net DD*, but the estimate for *Small Retail Std Abnormal Imbalance* is more than twice as large (6.09 versus 2.43). This

[27] As an example, consider GME on the day of January 27th, when the stock price had it largest increase of 137%. On this day, *Small Retail Percent Imbalance* was 0.15% which corresponds to the 78th percentile of the distribution, while *Small Retail Std. Abnormal Imbalance* was 61.09, which corresponds to the 99.9th percentile.

[28] Including contemporaneous returns and/or contemporaneous media sentiments yields virtually identical estimates.

finding is consistent with *WSB* recommendations influencing retail investors, particularly smaller retail investors, during the pre-GME period.

The correlation between *Small Retail Std Abnormal Imbalance* and *Net DD* remains economically large in the post-GME period (*Net DD* + *Net DD* × *Post* = 4.82). In contrast, the relation between *Large Retail Std Abnormal Imbalance* and *Net DD* declines in the post-GME period, and there is no significant relation between *Large Retail Std Abnormal Imbalance* and *Net DD* in the post-GME period. These findings point to the possibility that larger, and presumably more sophisticated retail investors are better able to recognize the decline in report quality in the post-GME period.[29]

Consistent with Farrell et al. (2022), we also find that retail order imbalances are correlated with the direction of *SA* recommendations. However, the magnitude of the effect is noticeably weaker. Specifically, for small retail traders the estimated coefficient on *Net DD* (6.09) is more than four times as large as the estimated effect on *Net SA* (1.46). This finding is consistent with regulators' concern that *WSB* has a particularly pronounced effect on retail trading.

6.2. The Informativeness of Retail Trading following DD Reports

We next examine whether the informativeness of small retail investor trading changes following DD reports. We focus on small retail investors because, as shown in Table 11, their trading is most strongly related to DD reports.[30] Given that DD reports are informative in the pre-GME period, but not the post-GME period (Table 3), and that small retail traders closely follow DD reports (Table 11), we conjecture that small retail trade informativeness increases following DD reports in the pre-GME period, but not the post-GME period. However, this need not be the case. For example, if smaller retail investors are less likely to trade following lower quality reports than the decline in DD report informativeness would not necessarily translate to a decline in the informativeness of small retail investor trading.

[29] In Table IA.10 of the Internet Appendix, we find that the decline in the relation between *Large Retail Std Abnormal Imbalance* and *Net DD* in the post-GME period is concentrated in *High-Attention* reports which were shown to be particularly uninformative in the post-GME period (see Table 9).

[30] We find no evidence that trade informativeness significantly changes following DD reports for either large retail investors (Table IA.11) or institutional investors (Table IA.12) in either the pre- or post-GME period.

To examine how the informativeness of small retail trading changes following DD reports we estimate the following regression:

$$
\begin{aligned}
R_{it+1,t+21} = {} & \beta_1 Imb_{it} + \beta_2 Imb_{it} \times Post_t + \beta_3 Imb_{it} \times DD_{it} + \beta_4 Imb_{it} \times DD_{it} \times Post_t \\
& + \beta_5 Imb_{it} \times NR\ Indicator_{it} + \beta_6 Imb_{it} \times NR\ Indicator_{it} \times Post_t \\
& + \beta_7 Imb_{it} \times SA_{it} + \beta_8 Imb_{it} \times SA_{it} \times Post_t + Controls_{it} + Day_t \\
& + \varepsilon_{it.}
\end{aligned}
\tag{4}
$$

The dependent variable is the one-month ahead return. *Imb* is *Small Retail Std Abnormal Imbalance,* and we interact *Imbalance* with *Post*, which tests whether the informativeness of small retail trading changes in the post-GME period. We also interact *Imbalance* with an indicator for whether there was a DD report and an indicator for a DD report in the post-GME period. Thus, *Imbalance* × *DD* tests whether trade informativeness following DD reports is different from trade informativeness on non-DD report days during the pre-GME period, and *Imbalance* × *DD* × *Post* examines whether this relation changes in the post-GME period. We also contrast the effects of DD reports on trade informativeness with the effects of non-research postings and *SA* reports by interacting *Imbalance* with *NR Indicator, NR Indicator* × *Post, SA,* and *SA*× *Post,* where *NR Indicator* and SA are indicators equal to one if there was at least one non-research post or at least one SA report, respectively. Finally, *Controls* and *Day* are defined in equation (2). We also consider specifications that control for the directional recommendation of the reports on the day (e.g., *Net DD* and *Net DD* × *Post*). Comparing the estimates across specifications that do and do not control for the recommendation of the DD report allows us to examine the importance of *Net DD* in explaining changes in retail trade informativeness following DD reports.

The results are reported in Table 12. Specifications 1 and 2 reports the results for the full sample and the sample that excludes GME and AMC, prior to controlling for the recommendation of the DD reports. We find some evidence that small retail imbalances in the pre-GME period are a stronger predictor of one-month ahead returns on days with DD reports. For instance, on days when a DD report is issued, a one decile increase in small retail order imbalance is associated with a 0.14% (1.39% x .10) increase in one-month ahead returns for the full sample and a 0.05% (0.54% x .10) increase for the sample that excludes GME and AMC. Both estimates are marginally significant ($p < 0.10$). We also find that these effects are fully eliminated in the post-

GME period. In particular, the coefficient on *Imbalance* × *DD* × *Post* is negative, and larger (in absolute value) than the coefficient on *Imbalance* × *DD*. Below the regression estimates, we also formally test for whether retail trade informativeness following DD reports in the post-GME period (i.e., *Imbalance* × *DD* + *Imbalance* × *DD* × *Post)* is significantly negative. We find that the estimates are not reliably different from zero for the full sample, but they are significantly negative for the sample that excludes GME and AMC. We also find that small retail trade informativeness following non-research reports is significantly negative for the sample that excludes GME and AMC. Overall, these results provide some support for regulators' concern that *WSB* can result in significant trading losses for retail investors.[31]

In Specifications 3 and 4 we explore the extent to which the decline in report informativeness contributes to the decline in small retail trade informativeness by including *Net DD* and *Net DD* × *Post* as controls. The inclusion of these variable attenuates the estimates. For example, for the full sample, the coefficient on *Imbalance* × *DD* falls by more than half (from 1.39 to 0.62), while the estimate on *Imbalance* × *DD* × *Post* falls by roughly one-third. We observe qualitatively similar patterns after excluding GME and AMC. These findings suggest that the decline in DD report informativeness contributes to some, but not all, of the decline in small retail trading informativeness following DD reports. One potential explanation for remaining decline is that the GME event shifted the composition of traders following DD reports towards less sophisticated investors.

7. Conclusion

Wallstreetbets (*WSB*) has become an increasingly prominent source of investment research, particularly for risk-seeking retail investors. This paper offers a first look at the investment value of *WSB* due-diligence (DD) reports. We find that prior to the GME short squeeze event, *WSB* was a source of valuable investment research. In particular, in the pre-GME period, *WSB* DD reports positively forecasted one-month ahead returns. *WSB* research also positively forecasted media sentiment, earnings surprises, and earnings forecast

[31] We acknowledge, however, that our analysis is limited to the informativeness of retail investor equity trading. Whether this translates into better (or worse) trading performance is an empirical question which can only be addressed with more granular account-level data.

revisions suggesting that *WSB* research contained useful information about future cash-flows news. However, all of the above findings are eliminated following the GME event. In the post-GME period, we find a dramatic increase in reports emphasizing price pressure strategies and reports focusing on attention-grabbing stocks, and we show that these shifts in strategy contributed to the significant decline in report informativeness. We also find that the informativeness of smaller retail trading increases following DD reports in the pre-GME period, but this relation fails to persist post-GME. Collectively, the evidence is consistent with the surge in new users stemming from the GME short squeeze event significantly altering the content of reports, deteriorating the informativeness of *WSB* research, and consequently, its potential benefits to less sophisticated investors.

Our findings should be of relevance to both regulators and investors. From a regulatory perspective, our evidence suggests that the impact of *WSB* research on financial markets is likely to be more nuanced than many of the black-and-white views expressed during the congressional hearings. For example, the pre-GME results suggest that the *WSB* can provide informative research despite the fact that *WSB* caters to a risk-seeking userbase, promises complete anonymity, offers minimal oversight, and provides limited reputational incentives. On the other hand, the post-GME results suggests that the culture and informativeness of *WSB* can change quickly, and such changes can have adverse consequences. For example, in the post-GME period, we find that *WSB* research was associated with trading frenzies that had potentially negative implications for smaller retail traders.

From an investor's perspective, the decline in the informativeness of *WSB* research in the post-GME period should provide caution to the 10+ million *WSB* subscribers who turn to *WSB* for investment research. Indeed, our evidence casts doubt on the view that simply following all DD report recommendations will generate significant abnormal returns going forward. However, *WSB* may still be a useful source of information for investors who are adept enough to discern between higher and lower quality *WSB* research. Our findings suggest that users should be particularly cautious of reports that focus on price pressure strategies or reports issued on attention-grabbing stocks. Identifying additional attributes of *WSB* reports that are associated with better performance, particularly in the post-GME period, is a potentially interesting area for future research.

References

Aharon, D. Y., Kizys, R., Umar, Z., and Zaremba, A., 2021. 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.

Allen, F., Haas, M., Nowak, E., Pirovano, M., & Tengulov, A., 2021. Squeezing shorts through social media platforms. Working paper.

Altınkılıç, O. and Hansen, R.S., 2009. On the information role of stock recommendation revisions. *Journal of Accounting and Economics 48*(1), 17-36.

Ammann, M., & Schaub, N. 2021. Do Individual Investors Trade on Investment-related Internet Postings? *Management Science* 67(9), 5679-5702.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X., 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61(1), 259-299.

Bai, J., and Massa, M., 2022. Is human-interaction-based information substitutable? Evidence from lockdown Working paper.

Bali, T. G., Cakici, N., and Whitelaw, R. F, 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99(2), 427-446.

Barber, B., Huang, X., Odean, T., & Schwarz, C., 2022. Attention Induced Trading and Returns: Evidence from Robinhood Users. *Journal of Finance* 77(6), 3141-3190.

Barber, B., Lin, S., and Odean, T., 2022. Resolving a paradox: Retail trades positively predict returns but are not profitable. Working paper.

Barber, B. and Odean, T. 2000. Trading is hazardous to your wealth. The common stock investment performance of individual investors. *Journal of Finance* 55(2), 773-806.

Barber, B. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785-818.

Barber, B., Odean, T., & Zhu, N., 2009. Do retail traders move markets? *Review of Financial Studies* 22(1), 151-186.

Barberis, N.,Greenwood, R., Jin, L., & Schleifer, A., 2015. X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1) 1-24.

Bartov, E., Faurel, L., and Mohanram, P. S., 2018. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93(3), 25-57.

Boehmer, E., Jones, C.M., Zhang, X. and Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance* 76(5), 2249-2305.

Bradley, D., Jame, R., and Williams, J. 2022. Non-deal roadshows, informed trading, and analyst conflicts of interest. *Journal of Finance* 77(1), 265-315.

Cassella, S., and Gulen, H., 2018. Extrapolation bias and the predictability of stock returns by price-scaled variable. *Review of Financial Studies* 31(11), 4345-4397.

Chen, H., De, P., Hu, J., and Hwang, B.H., 2014. Wisdom of the crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27(5), 1367-1403.

Cookson, J.A., Lu, R., Mullins, W., and Niessner, M., 2022. The social signal. Working paper.

Cookson, J.A., and Niessner, M., 2020. Why don't we agree? Evidence from a social network of investors. *Journal of Finance* 75(1), 173-228.

Crawford, S., Gray, W., Johnson, B., and Price, R., 2018. What motivates buy-side analysts to share recommendation online? *Management Science* 64(6), 2473-2972.

Da, Z., Huang, X., and Jin, L. 2021. Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics* 140(1), 175-196.

Dim, C., 2021. Should retail investors listen to social media analysts? Evidence from text-implied beliefs. Working paper.

Eaton, G.W., Green, T.C., Roseman, B. and Wu, Y., 2022. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics* 146(2), 502-528.

Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C., 2012. How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics* 105(2), 260-278.

Farrell, M., Green, T.C., Jame, R. and Markov, S., 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics* 145(2), 614-641.

Farrell, M., Jame, R. and Qiu, T., 2020. The cross-section of non-professional analyst skill. Working paper.

Frazzini, A., and Lamont, O. A., 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88(2), 299-322.

Giannini, R., Irvine, P., and Shu, T., 2018. Nonlocal disadvantage. An examination of social media sentiment. *Review of Asset Pricing Studies* 8(2), 293-336.

Goldstein, I., Ozdenoren, E., and Yuan, K., 2013. Trading frenzies and their impact on real investment. *Journal of Financial Economics* 109(2), 566-582.

Greenwood, R., and Shleifer, A., 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27(3), 714-746.

Hu, D., Jones, C. M., Zhang, V., & Zhang, X., 2021. The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery. Working paper.

Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. *Review of Financial Studies*, 21(3), 1123-1151.

Jame, R., 2018. Liquidity provision and the cross-section of stock returns. *Management Science* 64(7), 3288-3312.

Jame, R., Johnston, R., Markov, S., and Wolfe, M., 2016. The value of crowdsourced earnings forecasts. *Journal of Accounting Research* 54(4), 1077-1110.

Kaniel, R., Liu, S., Saar, G., & Titman, S., 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67(2), 639-680.

Kelley, E.K. and Tetlock, P.C., 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68(3), 1229-1265.

Kim, S.H., and Kim, D., 2014. Investor sentiment from internet message positives and the predictability of stock returns. *Journal of Economic Behavior and Organization* 107, 728-729.

Lee, C. M., & Ready, M. J., 1991. Inferring trade direction from intraday data. *Journal of Finance* 46(2), 733-746.

Long, C., Lucey, B. M., & Yarovaya, L., 2021. " I just like the stock" versus "fear and loathing on main street": The role of reddit sentiment in the GameStop short squeeze. Working paper.

Nagel, S., 2012. Evaporating liquidity. *Review of Financial Studies* 25(2), 2005-2039.

Ozik, G., Sadka, R., and Shen., S., 2021. Flattening the illiquidity curve: Retail trading during the COVID-19 Lockdown. *Journal of Financial and Quantitative Analysis* 56(7), 2356-2388.

Semenova, V., and Winkler, J., 2021. Reddit's self-organized bull runs: Social contagion and asset prices. Working paper.

Smith, S. and Timmermann, A., 2021. Have risk premia vanished? *Journal of Financial Economics* 145(2), 553-576.

Tetlock, P.C., Saar-Tsechansky, M., and Macskassy, S., 2008. More than words: Quantity language to measure firms' fundamentals. *Journal of Finance* 63(3), 1437-1467.

Tumarkin, R. and Whitelaw, R.F., 2001. News or noise? Internet postings and stock price, *Financial Analyst Journal* 57, 41-51.

Welch, I., 2022. The wisdom of the robinhood crowd. *Journal of Finance* 77(3), 1489-1527.

Womack, K. 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51(1), 137-167.

Appendix A: Sample report

Posted by u/ MikeThePutz
Post time: Wednesday, Jul 10, 2019, 02:55:11 PM EST.

BYND is at Costco DD

.

DD

Costco is now carrying Beyond Burgers. I don't see this in any press releases by either Costco or BYND. There was a vegan blog that mentioned this information (it was brought to my attention by **relevant_pet_bug**) and I called the stores to confirm it. The item # is 1338620. It is in approximately 15 stores nationwide. A store in San Diego has it for sure, 2345 Fenton Parkway, and the Costco on 1890 S University Drive in Davie, Florida has it. I don't know the other 13 stores. When I called a purchasing manager in the regional area to find out why the stores only carried a limited supply of the Beyond Burgers, he said that Costco would buy as many Beyond Burgers as they could get their hands on, but that BYND only sold a limited amount because they couldn't keep up with demand. He said that Beyond Burgers are selling really well and they are selling out in "just a few days". Apparently, BYND will be installing new manufacturing lines by the end of 2019 to increase supply and supposedly they will be able to provide Costco nationwide with Beyond Burgers by next year.

I asked why Costco would sell Beyond Burgers when they already sell Morning Star and Don Lee Farm's Veggie Burgers and the purchasing manager said that Beyond Burgers just taste different and customers want them. AGAIN: This man is responsible for buying items for Costco in a large region of the US and he said that they would buy as many BYND Burgers as they could, but that supply was limited and that they will stock Beyond Burgers nationwide once BYND can meet demand in early 2020. This is hugely positive news and I don't see any news reports about it or analyst reports mentioning it (please let me know if I am wrong!)

Photos taken from other groups about it: https://imgur.com/a/FcalTI9

If any reporters want my sources for this story feel free to PM me.

I want to thank u/**relevant_pet_bug** for pointing me in the right direction and bringing the vegan blog post, where this was first mentioned, to my attention.

Edit: fixed an error with the years. I legitimately forgot we are living in 2019.

153 comments
95% Upvoted

Appendix B: List of Keywords in Price Pressure Analysis

This table reports the list of keywords assigned as "price pressure" words or "fundamental" words.

Price Pressure Words	Fundamental Words
Squeeze	Earnings
Short Interest	EPS
Short Sellers	Revenue
Short volume	Sales
Gamma	Growth Rate
Float	Cash Flow
Hedge	Net Income
Melvin	Customers
Citadel	Competitors
Robinhood (RH)	Market Share
Dealers	Store Visits
"HODL"[1]	P/S Ratio
	P/E Ratio
	Guidance
	Analysts

[1] HODL originated as misspelling of "Hold" in a WSB post, and it has become a popular inside joke on the site. Many users now also view HODL as an acronym for Hold On for Dear Life.

Appendix C: Variable Definitions

C.1 Outcome Variables

- *WSB DD Coverage (Table 2)* – the natural log of 1 plus the total number of Wallstreetbets (*WSB*) due diligence (DD) reports written for a firm during the calendar month. (Source: *WSB*).
- *WSB Non-Research Coverage* (Table 2) – the natural log of 1 plus the total number on *WSB* non-research posts reports for a firm during a calendar month, where a non-research post includes post belonging to one of the following *WSB* categories: *News, Losses, Gains, Charts,* and *Sh$tpost.* (Source: *WSB*).
- *SA Coverage (Table 2)* – the natural log of 1 plus the total number of Seeking Alpha *(SA)* reports written for a firm during the calendar month. (Source: Seeking Alpha).
- $Ret_{t+1,t+x}$ *(Tables 3,4,5,6, 9,10, and 12)* – the buy and hold return starting on day $t+1$ and ending on day $t+x$, where x typically equals five or 21 trading days.
- *News Sentiment*$_{t+1,t+x}$ *(Table 7)* – the sum of a daily sentiment score starting on day $t+1$ and ending on day $t+x$. The sentiment scores are obtained from Bloomberg and range from -1 (very negative news) to 1 (very positive news), with a median value of 0 (neutral articles). We assign firms with no media coverage a value of 0. (Source: Bloomberg).
- *Positive Forecast Error*$_{t+1,t+x}$ *(Table 7)* – An indicator equal to one if the realized quarterly earnings reported between day $t+1$ and day $t+x$ exceed the median forecast across all I/B/E/S analysts. The value is set missing for firms that do not have I/B/E/S coverage or for firms that will not announce earnings over the forecast horizon being analyzed (i.e. five or 21 trading days). (Source: I/B/E/S).
- *Positive Forecast Revision*$_{t+1,t+x}$ *(Table 7)* – the total number of upward revisions issued between day $t+1$ and days $t+x$, scaled by the total number of revisions issued over the same period. In computing this measure, we consider both quarterly and annual earnings forecast revision. This value is set to missing for firms that do not have I/B/E/S coverage, and the value is set to 50%, the median value across the sample, for firms with I/B/E/S coverage but no forecast revisions over the holding period. (Source: I/B/E/S).
- *PP* (Table 8) – an indicator equal to one if the number of price pressure words in the report exceeds the number of fundamental words, where the list of price pressure and fundamental words is available in Appendix C. (Source *WSB*).
- *PP2* (Table 8) – an indicator equal to one if the number of price pressure words is greater than zero, where the list of price pressure words is available in Appendix C. (Source *WSB*).
- *High Absolute Return* (Table 8) – an indicator equal to one if the absolute return on the day prior to the DD report was in the top decile. (Source CRSP).
- *High WSB Posts* (Table 8) – an indicator equal to one the firm had more than one non-research post issued on *WSB* over the previous five trading days. (Source *WSB*).
- *High Attention* (Table 8) – an indicator equal to one if either *High Absolute Return* or *High WSB Posts* is equal to one. (Source CRSP and *WSB*).
- *Institutional Percent Imbalance (*Table 11) – institutional buy share volume less institutional sell share volume scaled by total institutional share volume. Institutional trades are assigned as buys or sells based on the Lee and Ready (1991) algorithm. (Source: TAQ Intraday Indicators).
- *Large Retail Percent Imbalance (*Table 11) – retail buy share volume less retail sell share volume scaled by total retail share volume. Retail trades are assigned as buys or sells based on the BJZZ algorithm. (Source: TAQ Intraday Indicators).
- *Small Retail Percent Imbalance (*Table 11) – the number of retail buy trades less the number of retail sell trades scaled by total retail trades. Retail trades are assigned as buys or sells based on the BJZZ algorithm. (Source: TAQ Intraday Indicators).
- *Institutional Std Abnormal Imbalance* (Table 11) – institutional buy share volume less institutional sell share volume (*institutional imbalance*) less the average *institutional imbalance* over days *t-120* through *t-240*, scaled by the standard deviation of *institutional imbalance* estimated over days *t-120* through *t-240*. Institutional trades

are assigned as buys or sells based on the Lee and Ready (1991) algorithm. (Source: TAQ Intraday Indicators).

- *Large Retail Std Abnormal Imbalance* (Table 11) – retail buy share volume less retail sell share volume (*large retail imbalance*) less the average *large retail imbalance* over days *t-120* through *t-240*, scaled by the standard deviation of *large retail imbalance* estimated over days *t-120* through *t-240*. Retail trades are assigned as buys or sells based on the BJZZ algorithm. (Source: TAQ Intraday Indicators).
- *Small Retail Std Abnormal Imbalance* (Table 11) – the number of retail buy trades less the number of retail sell trades (*small retail imbalance*) less the average *small retail imbalance* over days *t-120* through *t-240*, scaled by the standard deviation of *small retail imbalance* estimated over days *t-120* through *t-240*. Retail trades are assigned as buys or sells based on the BJZZ algorithm. (Source: TAQ Intraday Indicators).

C.2 Main Independent Variables

- *Net DD* – the total number of *WSB* due diligence (DD) reports that recommend buying the firm over a time period (e.g., one day) less the total number of DD reports that recommend selling the firm during the time period. We winsorize this value at the 1st and 99th percentile of the distribution of firm days where *Net DD* is not equal to zero. (Source: *WSB*).
- *NonResearch Posts* – the total number of non-research posts where a ticker is mentioned in the title over a time period (e.g., one day). We classify posts in the following *WSB* categories as non-research related: *News*, *Gains, Losses, Charts,* and *Shi$posts*. We winsorize this value at the 1st and 99th percentile of the distribution of firm days where *NonResearch Posts* is not equal to zero. (Source: *WSB*).
- *Net SA* – the total number of Seeking Alpha research reports that recommend buying the firm over a time period (e.g., one day) less the total number of Seeking Alpha reports that recommend selling the firm during the time period. We winsorize this value at the 1st and 99th percentile of the distribution of firm days where *Net SA* is not equal to zero. (Source: SA).
- *Post GME* (*Post*) – an indicator equal to one for the post-GME period (January 14, 2021 – June 30, 2021) and zero for the pre-GME period (July 1, 2018 – January 12, 2021).
- *Pre-Pandemic* – an indicator equal to one for the pre-GME period that coincides with the period prior to the pandemic (July 2018 – March 15, 2020) and zero otherwise.
- *Post-Pandemic* – an indicator equal to one for the pre-GME period that coincides with the period that coincides with the pandemic (March 16, 2020 – January 12, 2021) and zero otherwise.
- *Processing* – an indicator equal to one if the report is issued around a major information event, defined as an earnings announcement issued on the previous or current day (-1, 0), an analyst revision on the previous or current day (-1, 0) or abnormal media coverage on the previous or current day (-1, 0).
    - *Earning Announcement* –a quarterly or annual earnings announcement (Source: I/B/E/S).
    - *Analyst Revision* – a quarterly or annual earnings forecast revision (Source: I/B/E/S).
    - *Abnormal Media Coverage* – an indicator equal to one if the number of articles on the firm is in the top 20% relative to the firm's average media coverage over the previous 60 days [-60, -1]. (Source: Bloomberg).
- *Net DD Processing* – *Net DD* computed using only the subset of reports where *Processing* =1. (Source: *WSB*, I/B/E/S, and Bloomberg).
- *Net DD Production* –*Net DD* computed using only the subset of reports where *Processing* =0. (Source: *WSB*, I/B/E/S, and Bloomberg).
- *Net DD PP* – *Net DD* computed using only the subset of reports where *PP Report = 1*. (Source: *WSB*).
- *Net DD Attention* – *Net DD* computed using only the subset of reports where *High Attention = 1*. (Source: *WSB* and CRSP).
- *Net DD PP/Attenttion* – *Net DD* computed using only the subset of reports where either *PP Report* =1 or *High Attention = 1*. (Source: *WSB* and CRSP).
- *DD* – an indicator equal to one if a *DD* report was issued for firm *i* on day *t* and zero otherwise. (Source: *WSB*).

- *NR Indicator* – an indicator equal to one if a *Non-Research* post was issued for firm *i* on day *t* and zero otherwise. (Source *WSB*).
- *SA* – an indicator equal to one if an SA report was issued for firm *i* on day *t* and zero otherwise. (Source: Seeking Alpha).

C.3 Other Control Variables

- *Size* – the market capitalization computed as share prices times total shares outstanding at the end of the year. (Source: CRSP).
- *Book-to-Market (BM)* – the book-to-market ratio computed as the book value of equity during the calendar year scaled by the market capitalization at the end of the calendar year. Positive values are winsorized at the 1st and 99th percentile. Negative value and missing values are set equal to zero and we include a corresponding "Missing BM" indicator. (Source: CRSP/Compustat).
- *Volatility* – the standard deviation of daily returns during the month (Source: CRSP).
- *Turnover* – the average daily turnover (i.e., share volume scaled by shares outstanding) during the month.
- *Ret [0]* – the buy-and-hold return on the current day. (Source: CRSP).
    - *Ret [-5, -1]* - the buy-and-hold return over the previous five trading days.
    - *Ret [-26, -6]* - the buy-and-hold return over the previous six to 26 trading days.
    - *Return (m-1)* – the buy-and-hold return in the previous calendar month.
    - *Return (m-2, m-12)* – the buy-and-hold return over the previous two to twelve calendar months. (Source: CRSP).
    - *Abs. Ret* – the absolute value of *Ret*.
- *Sentiment [0]* – The average sentiment scores across all news articles on the current day, where the score ranges from -1 (very negative news) to 1 (very positive news), with a median value of 0 (neutral articles). Firms with no media coverage are assigned a sentiment score of 0. (Source: Bloomberg).
    - *Sentiment [-5, -1]* – the sum of the sentiment score over the previous 1 to 5 trading days.
    - *Sentiment [-26, -6]* – the sum of the sentiment score over the previous six to 26 trading days.
    - *Abs. Sentiment* – the absolute value of *Sentiment*.
- *Institutional Ownership* – the percentage of the firm's shares held by institutions at year end. (Source: Thomson Reuters Institutional Holdings S34).
- *Breadth of Ownership* – the total number of common shareholders (Source: Compustat).
- *IBES Coverage* – the number of unique brokerage houses issuing earnings forecast for a firm during the calendar year. (Source: I/B/E/S).
- *Media Coverage* – the total number of media articles about a firm during the calendar year. (Source: Bloomberg).
- *High Max* – an indicator equal to one if the maximum daily return of the firm in the prior month was in the top quintile of the distribution (Source: CRSP).
- *Heavy Short* – an indicator equal to one if the firm is in the top quintile of short interest, defined as the number of shares that have been sold short scaled by shares outstanding. (Source: Compustat).
- *Recent IPO* – an indicator equal to one if the firm went public in the past six months. (Source: CRSP).
- *GME/AMC* – an indicator equal to one for GME or AMC and zero otherwise.

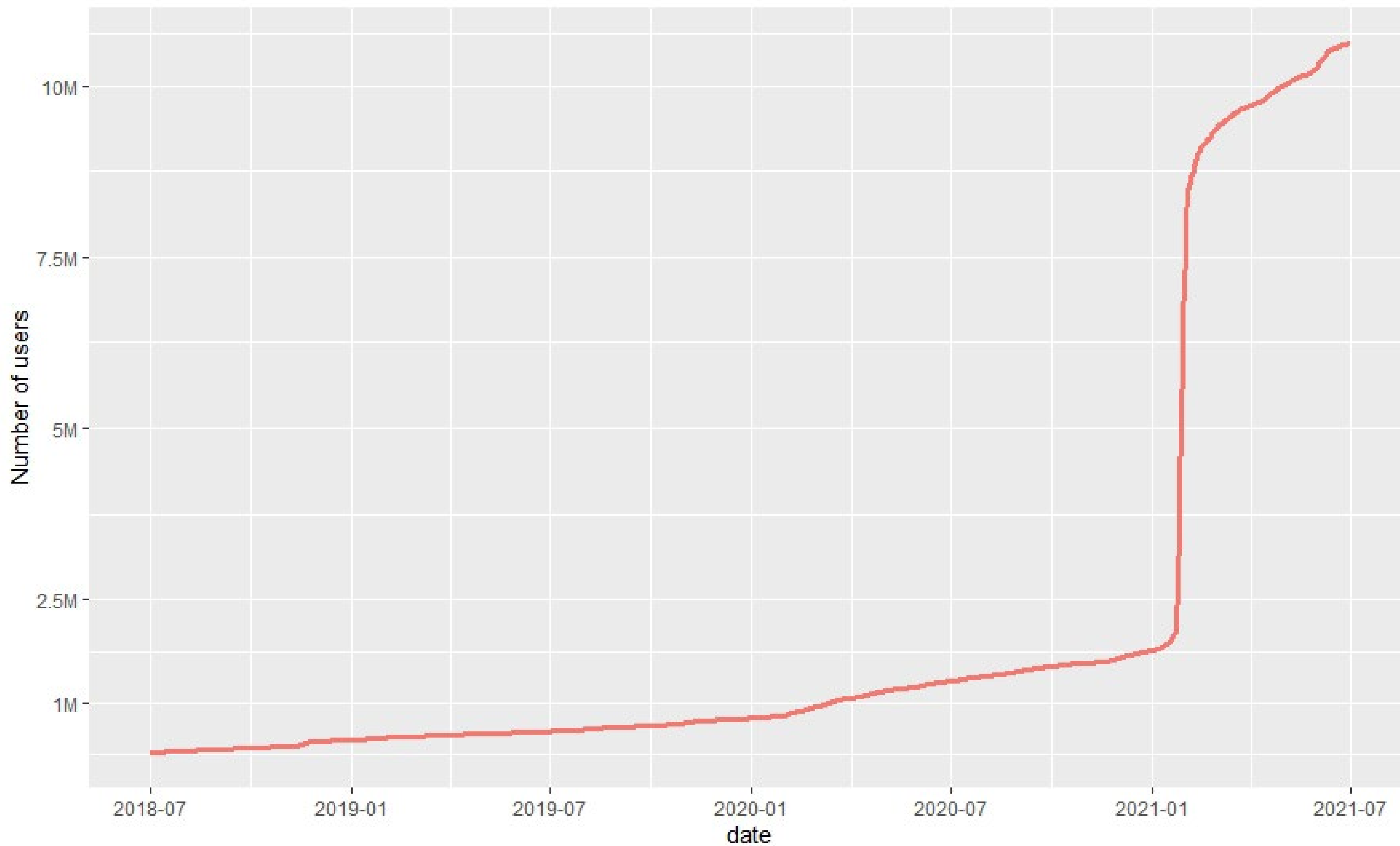
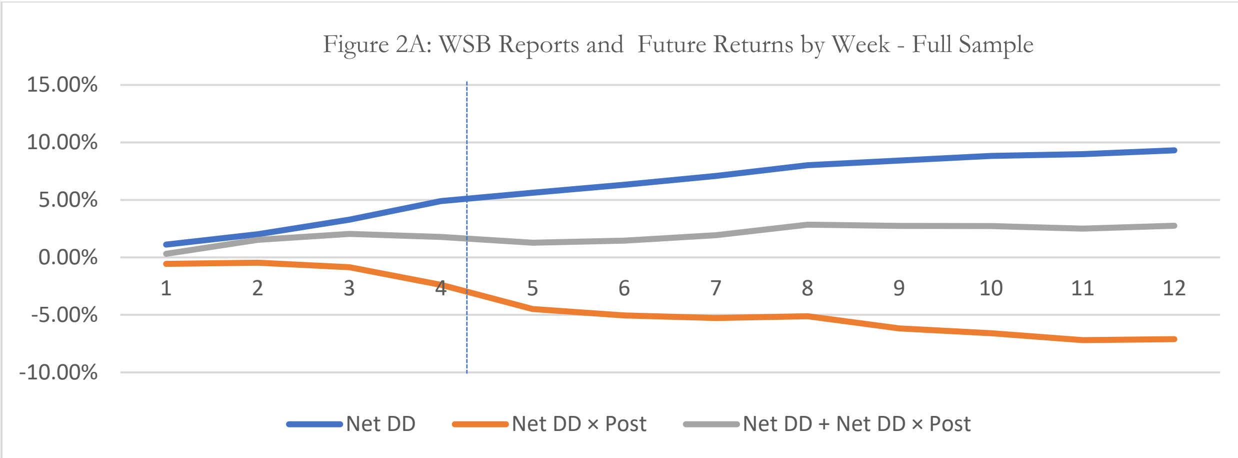


Figure 1: Growth in Reddit's Wallstreetbets (WSB)
This figure plots the total number of users on WSB from July 2018 through June 2021. This data can be found at https://subredditstats.com/r/wallstreetbets.



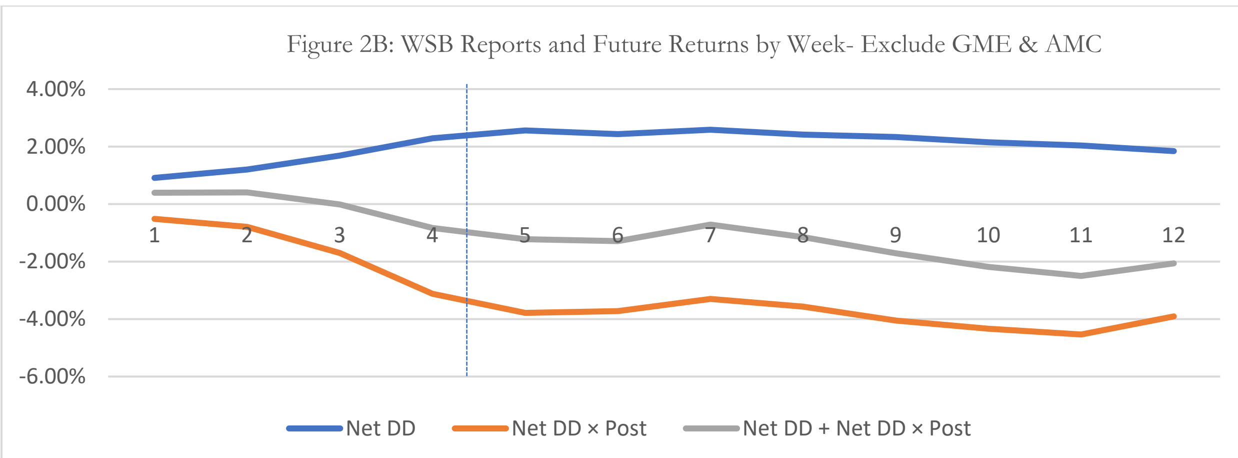


Figure 2: *WSB* Reports and Future Returns by Week
This table repeats the estimates from Table 3 for horizons ranging from one-week (i.e., x =5) through 12 weeks (i.e., x = 60). We report the coefficient estimates on *Net DD, Net DD × Post,* and *Net DD + Net DD × Post* for each horizon. The dashed-blue line corresponds to the 21-day holding period studied in Specifications 2 and 4 of Table 3. Figures 2A and 2B report the results for the full sample and the sample that excludes GME and AMC, respectively.

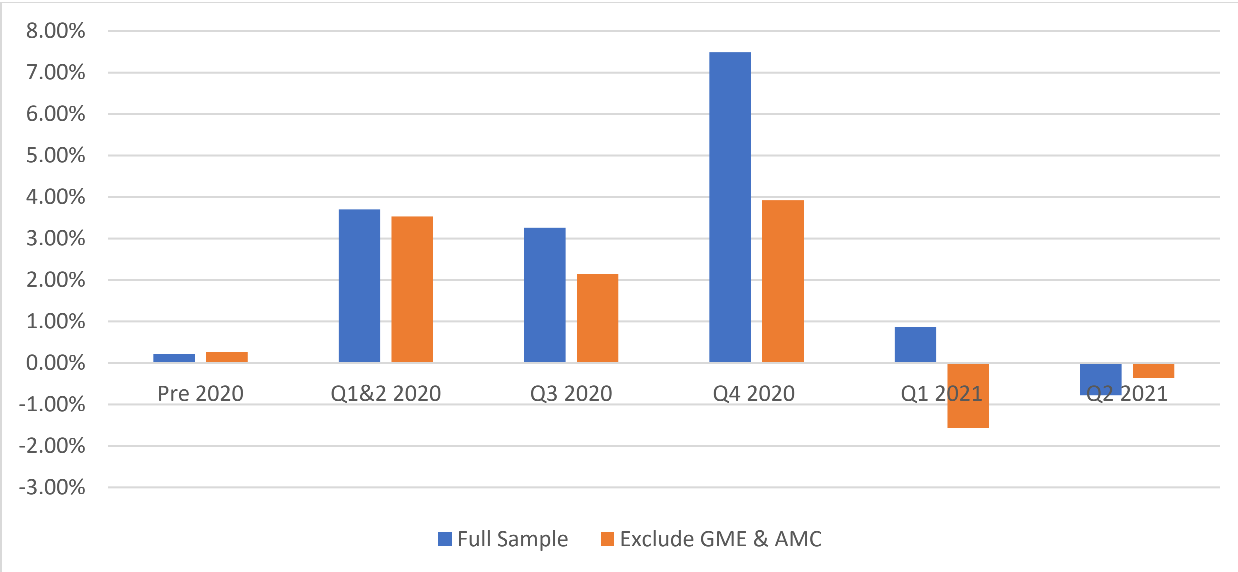


Figure 3: WSB Reports and Future Returns – Results by Calendar Quarter
This figure reports the estimates from Specifications 2 and 4 of Table 3 over the following intervals: the-pre 2020 sample, quarters 1 and 2 of 2020, and each of the remaining quarters in 2020 and 2021. We include the small number of pre-GME DD reports in January of 2021 in the Q4 2020 estimates. The blue bars report the estimates for the full sample of stocks (i.e., Specification 2 of Table 3), and the orange bars report the estimates for the sample that excludes GME and AMC (i.e., Specification 4 of Table 3).

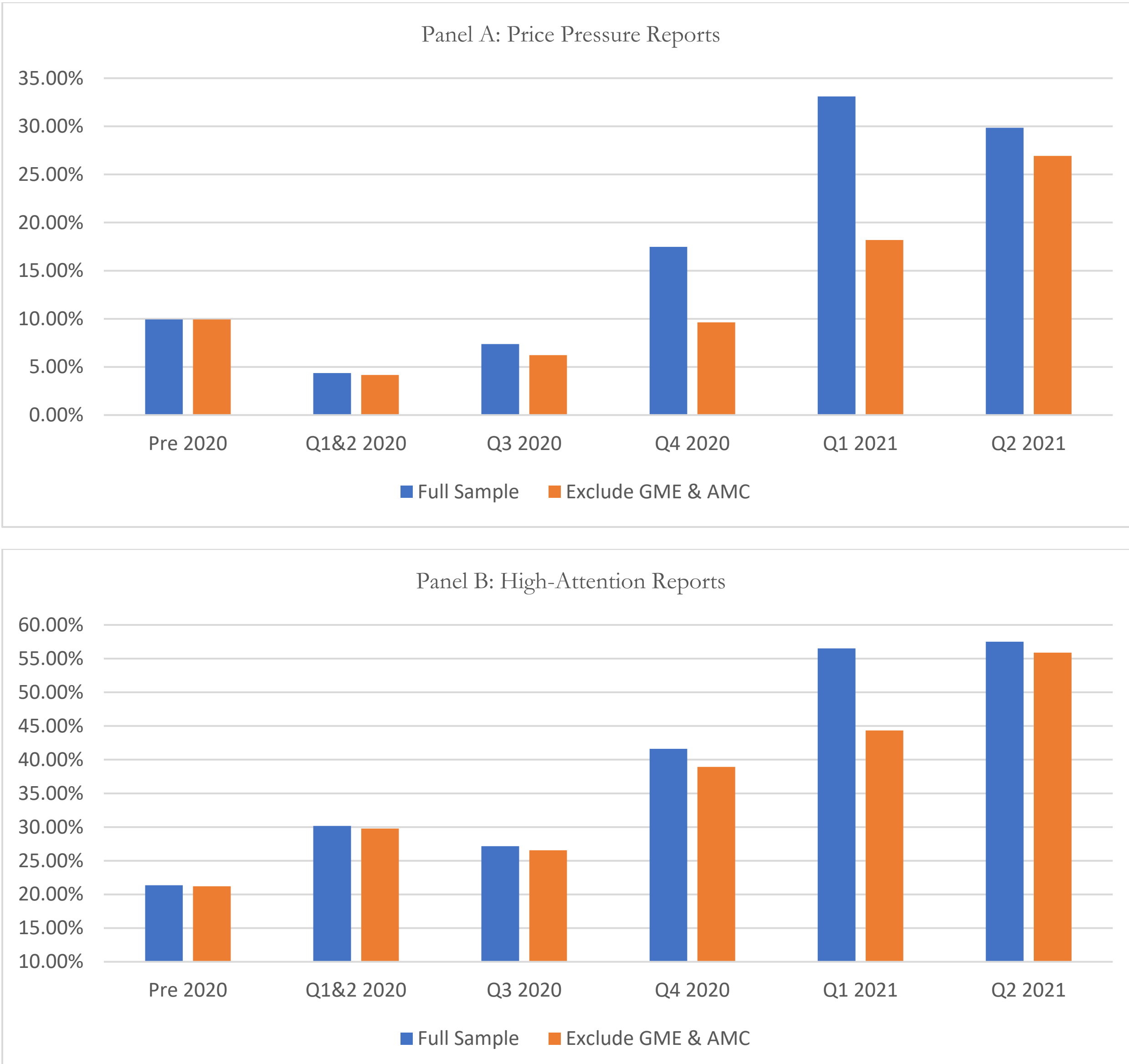


Figure 4: The Frequency of Price Pressure and High Attention Reports by Calendar Quarter
This figure reports the frequency of price pressure and high-attention reports over the following intervals: the-pre 2020 sample, quarters 1 and 2 of 2020, and each of the remaining quarters in 2020 and 2021. We include the small number of pre-GME DD reports in January of 2021 in the Q4 2020 estimates. The blue bars report the estimates for the full sample of stocks and the orange bars reports the estimates for the sample that excludes GME and AMC. We classify a report as a price pressure report if the number of price pressure words exceed the number of fundamental words (see Appendix C for the list of price pressure and fundamental words). We classify as a report as *High Attention* if either 1) the absolute return on the day prior to the DD report was in the top decile (*High Absolute Return*) or if the firm had more than one non-research post issued on WSB over the previous five trading days (*High WSB Posts*).

Table 1: Descriptive Statistics

This table reports summary statistics on the sample of social media posts. Panel A reports the results for Due Diligence (DD) reports on Reddit's Wallstreetbets (*WSB*). DD reports are reports identified by the poster (and verified by the moderator) as containing some analysis and offering a clear buy or sell signal. We limit the sample to DD reports that focus on a single common stock. We report the number of DD reports for the full sample (July 2018-June 2021), the *Pre-GME* period (July 2018 – January 12, 2021), and the *Post-GME* period (January 14, 2021-June 2021). For each period, we report the number of firm-days and firms with at least one DD report, the percentage of reports recommending a long position (*% Buys*), the percentage of DD reports that are written on GME or AMC *(GME/AMC),* and the average number of DD reports issued by each username (*Posts per Contributor*). We also report the mean and standard deviation of DD reports across all firm-months. The sample includes 90,175 firm-months in the pre-GME period and 27,344 firm-months in the post-GME period. Panel B reports results for *WSB* Non-Research posts defined as a post belonging to one of the following WSB categories: *News, Losses, Gains, Charts,* and *Sh$tpost,* and Panel C reports analogous results for the sample of research reports provided by Seeking Alpha.

Panel A: WSB DD Reports -Full Sample

			DD Report Statistics				Firm-Month Coverage	
	DD Reports	Firm-Days	Firms	% Buys	% GME or AMC	Posts per Contributor	Mean	Std Dev
Full Sample	5,015	3,782	906	88%	12.08%	1.32	0.04	0.80
Pre-GME	2,443	2,096	635	81%	4.18%	1.33	0.02	0.31
Post-GME	2,572	1,686	501	95%	19.60%	1.24	0.08	1.57

Panel B: WSB Non-Research Reports

			DD Report Statistics				Firm-Month Coverage	
	Non-Research Reports	Firm-Days	Firms	% Buys	% GME or AMC	Posts per Contributor	Mean	Std Dev
Full Sample	13,255	4,656	710	N/A	42.31%	1.21	0.10	8.37
Pre-GME	5,585	3,344	546	N/A	4.33%	1.34	0.05	1.49
Post-GME	7,589	1,312	355	N/A	70.01%	1.10	0.27	17.15

Panel C: Seeking Alpha Reports

			DD Report Statistics				Firm-Month Coverage	
	SA Reports	Firm-Days	Firms	% Buys	% GME or AMC	Posts per Contributor	Mean	Std Dev
Full Sample	23,659	22,460	2,955	85%	0.48%	10.98	0.17	0.71
Pre-GME	19,902	18,846	2,652	85%	0.39%	10.26	0.19	0.76
Post-GME	3,757	3,614	1,630	88%	0.96%	5.10	0.13	0.50

Table 2: Determinants of WSB Coverage

This table presents the estimates from Equation (1):

$$Coverage_{it} = \alpha + \beta_1 Chars_{it-1} + \beta_2 Chars_{it-1} \times Post_t + Month_t + \varepsilon_{it}.$$

The dependent variable, *Coverage,* is either *WSB DD Coverage* defined as Log (1 + DD Reports) for firm *i* during month *t, WSB Non-Research Coverage*, defined as Log (1+ Non-Research posts) for firm *i* during month *t*, or *SA Coverage* defined as Log (1 + SA Reports) for firm *i* during month *t*. *Chars* include the following firm characteristics: the percentage of the firm's shares held by institutional investors at the end of the prior year (*Inst. Ownership*), the number of common shareholders (*Breadth of Ownership*), market capitalization (*Size*), book-to-market ratio (*BM),* return volatility (*Volatility*), share turnover (*Turnover*), returns over the prior month (Ret_{m-1}), returns over the prior two to twelve months ($Ret_{m-2, m-12}$), the number of media articles mentioning the firm in the prior year (*Media Coverage*), the number of sell-side analysts issuing a forecast for the firm in the prior year (*IBES Coverage*), an indicator equal to one if the maximum daily return of the firm in the prior month is in the top quintile (*High Max),* an indicator equal to one if the firm is in the top quintile of short interest (*Heavy Short*), an indicator equal to one if the firm went public in the past six months *(Recent IPO*), and an indicator equal to one for GME or AMC (*GME/AMC*). We allow the loadings on firm characteristics to vary in the pre-GME and post-GME period by interacting the firm characteristics with *Post,* an indicator equal to one for the Post-GME period (February 2021 – June of 2021) and zero for the Pre-GME period (July 2018 – December 2020), with the month of the GME-event (January 2021) excluded. We also include calendar-month fixed effects, and we standardize all continuous variable to have mean zero and unit variance. More detailed variable definitions are in Appendix C. Standard errors are clustered by firm and month, and *t*-statistics are reported in parentheses.

	WSB DD Coverage	*WSB Non-Research Coverage*	*SA Coverage*
	[1]	[2]	[3]
Inst Ownership	-0.03	-0.04	-0.06
	(-3.84)	(-4.57)	(-5.87)
Inst Ownership × Post GME	-0.04	0.01	0.04
	(-1.55)	(0.72)	(3.94)
Log (Breadth of Own.)	0.01	0.01	0.03
	(0.80)	(0.31)	(2.12)
Log (Breadth of Own.) × Post GME	-0.01	-0.01	-0.02
	(-0.58)	(-0.83)	(-1.46)
Log (Size)	0.26	0.38	0.84
	(4.56)	(4.26)	(12.28)
Log (Size) × Post GME	0.50	-0.03	-0.10
	(3.79)	(-0.23)	(-1.36)
Log (BM)	-0.02	-0.03	-0.03
	(-2.39)	(-2.05)	(-2.42)
Log (BM) × Post GME	0.03	0.02	-0.02
	(1.96)	(1.97)	(-1.71)
Negative BM	0.02	0.03	0.02
	(1.21)	(0.95)	(0.53)
Negative BM × Post GME	-0.03	0.00	0.01
	(-0.27)	(-0.03)	(0.28)
Log (Vol)	0.11	0.15	0.47
	(4.74)	(3.71)	(9.51)
Log (Vol) × Post GME	0.16	-0.03	-0.26
	(2.42)	(-0.44)	(-4.46)
Log (Turn)	0.00	0.01	-0.03
	(0.20)	(0.53)	(-1.98)
Log (Turn) × Post GME	0.12	0.05	0.08
	(3.01)	(1.84)	(5.23)
Return [m-1]	0.01	0.02	0.01

	(2.08)	(3.08)	(2.12)
Return [m-1] × Post GME	0.04	0.04	0.01
	(1.50)	(1.66)	(1.01)
Return [m-2, m-12]	0.30	0.33	0.23
	(3.01)	(2.43)	(4.95)
Return [m-2, m-12] × Post GME	-0.27	-0.31	-0.21
	(-2.85)	(-2.36)	(-4.79)
Log (Media Coverage)	0.17	0.25	0.38
	(3.68)	(3.60)	(6.24)
Log (Media Coverage) × Post GME	0.01	-0.19	-0.18
	(0.12)	(-3.41)	(-3.45)
Log (IBES Coverage)	-0.01	-0.02	0.04
	(-0.76)	(-1.12)	(2.87)
Log (IBES Coverage) × Post GME	-0.08	-0.02	-0.05
	(-2.59)	(-0.74)	(-3.02)
High Max	0.03	0.05	0.06
	(2.87)	(3.35)	(3.62)
High Max × Post GME	0.14	-0.02	0.01
	(6.36)	(-0.56)	(0.66)
Heavy Short	0.05	0.06	0.15
	(2.69)	(2.19)	(5.53)
Heavy Short × Post GME	0.27	0.07	-0.06
	(3.46)	(0.71)	(-1.64)
Recent IPO	0.57	0.75	0.97
	(3.80)	(4.15)	(6.92)
Recent IPO × Post GME	1.17	0.01	0.41
	(3.49)	(0.03)	(2.92)
GME/AMC	3.38	2.76	1.77
	(8.52)	(2.96)	(4.14)
GME/AMC × Post GME	15.38	21.71	0.76
	(12.43)	(18.64)	(1.53)
Obs. (Firm-Months)	113,728	113,728	113,728
Fixed Effects	Month	Month	Month
R-square	3.28%	3.94%	13.34%

Table 3: WSB Reports and Future Returns

This table reports results from the estimation of Equation (2):

$$R_{it+1,t+x} = \beta_1 Net\ DD_{it} + \beta_2 Net\ DD_{it} \times Post_t + \beta_3 NonResearch_{it} + \beta_4 NonResearch_{it} \times Post_t + \beta_5 Net\ SA_{it} + \beta_6 Net\ SA_{it} \times Post_t + Controls_{it} + Day_t + \varepsilon_{it}.$$

The dependent variable, *R*, is the stock return measured over the subsequent week (i.e., x = 5 trading days) or the subsequent month (x = 21 trading days). *Net DD* is the number of buy DD recommendations less the number of sell DD recommendations for stock *i* on day *t*, and *Net DD × Post,* interacts *Net DD* with an indicator for the *Post-GME* period as defined in Table 1. *NonResearch* is the number of non-research posts (as defined in Table 1)*,* and *Net SA* is the number of *SA* reports issuing a buy recommendation less the number of *SA* reports issuing a sell recommendation for stock *i* on day *t*. *Controls* include size, book-to-market, prior returns and prior media sentiment measured on the current day, the previous five days, and the previous six to 26 days. *Day* denotes calendar-day fixed effects. More detailed variable definitions are in Appendix C. Standard errors are clustered by firm and month, and *t*-statistics are reported below each estimate. Below the regression estimates, we report formal tests for whether certain pairs of coefficients are significantly different from zero.

	Ret [1,5] [1]	Ret [1,21] [2]	Ret [1,5] [3]	Ret [1,21] [4]
Net DD	1.11%	5.17%	0.91%	2.33%
	(2.17)	(2.77)	(2.12)	(2.46)
Net DD × Post GME	0.11%	-5.30%	-0.51%	-3.45%
	(0.09)	(-3.27)	(-0.83)	(-2.52)
WSB NonResearch	0.20%	3.61%	0.14%	0.40%
	(1.44)	(1.33)	(1.08)	(1.03)
WSB NonResearch × Post GME	1.06%	-2.84%	-0.44%	-1.68%
	(7.05)	(-1.25)	(-2.89)	(-3.80)
Net SA	0.33%	0.59%	(0.00)	0.65%
	(4.24)	(2.48)	(4.26)	(2.68)
Net SA × Post GME	0.11%	0.39%	(-0.00)	0.36%
	(0.69)	(0.95)	(-0.09)	(0.91)
Log (Size)	-0.08%	-0.27%	-0.08%	-0.27%
	(-1.61)	(-1.27)	(-1.61)	(-1.26)
Log (BM)	-0.07%	-0.24%	-0.07%	-0.25%
	(-0.90)	(-0.80)	(-0.92)	(-0.83)
Ret [0]	-7.32%	-9.54%	-7.32%	-9.44%
	(-5.19)	(-4.70)	(-5.20)	(-4.58)
Ret [-5, -1]	-2.82%	-4.11%	-2.77%	-3.99%
	(-2.66)	(-2.78)	(-2.58)	(-2.70)
Ret [-26, -6]	-0.34%	-0.56%	-0.32%	-0.70%
	(-1.18)	(-0.47)	(-1.06)	(-0.56)
News Sentiment [0]	0.10%	0.10%	0.07%	0.08%
	(3.25)	(1.09)	(2.12)	(0.97)
News Sentiment [-5, -1]	0.02%	0.03%	0.01%	0.02%
	(0.56)	(0.34)	(0.23)	(0.30)
News Sentiment [-26, -6]	0.01%	0.05%	0.01%	0.06%
	(0.38)	(0.79)	(0.64)	(1.04)
Net DD + Net DD × Post	1.22%	-0.13%	0.40%	-1.12%
	(0.96)	(-0.11)	(0.81)	(-1.05)
Net DD − NonResearch	0.91%	1.56%	0.77%	1.93%
	(1.84)	(0.87)	(1.77)	(2.24)
Net DD × Post – NonResearch × Post	-0.95%	-2.45%	-0.07%	-0.78%
	(-1.63)	(-1.19)	(-0.12)	(-1.34)
Net DD − Net SA	0.79%	4.58%	0.59%	1.68%
	(1.55)	(2.47)	1.35%	(1.72)
Net DD × Post − Net SA × Post	0.00%	-5.68%	-0.50%	-3.81%
	(0.00)	(-3.11)	(-0.77)	(-3.07)
Obs. (Firm-Days)	2,772,053	2,772,053	2,770,545	2,770,545
Day FE	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	Yes	No	No

Table 4: WSB Reports and Future Returns - Alternative Horizons

This table repeats the analysis in Table 3 after replacing the dependent variable with the return on the current day (i.e., Day 0), each of the subsequent five days, the subsequent 2, 3, and 4 weeks, and the subsequent 5 through 12 weeks. We report the estimate on *Net DD, Net DD × Post*, and the sum of the two estimates. Panel A reports the results for the full sample, and Panel B reports analogous results after excluding GME and AMC. Standard errors are clustered by firm and month, and *t*-statistics are reported next to each estimate.

Panel A: Include GME & AMC

	Net DD		*Net DD × Post GME*		*Net DD + Net DD × Post GME*	
Ret. Period	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*
0	0.88%	(6.75)	-0.57%	(-3.19)	0.31%	(2.52)
1	0.23%	(1.61)	0.32%	(1.13)	0.55%	(2.36)
2	0.33%	(1.89)	-0.20%	(-0.84)	0.13%	(0.72)
3	0.18%	(1.80)	-0.29%	(-1.62)	-0.11%	(-0.72)
4	0.10%	(0.87)	0.36%	(0.76)	0.45%	(0.97)
5	0.20%	(2.32)	-0.14%	(-1.36)	0.05%	(0.57)
[6-10]	0.90%	(1.33)	-0.39%	(-0.76)	0.51%	(0.72)
[11-15]	1.27%	(2.08)	-1.53%	(-2.01)	-0.26%	(-0.72)
[16-20]	1.61%	(1.46)	-2.11%	(-1.73)	-0.50%	(-0.88)
[21-60]	4.41%	(3.11)	-2.46%	(-1.01)	1.95%	(0.94)

Panel B: Exclude GME & AMC

	Net DD		*Net DD × Post GME*		*Net DD + Net DD × Post GME*	
Ret. Period	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*
0	0.89%	(6.81)	-0.40%	(-2.36)	0.49%	(4.45)
1	0.24%	(1.53)	0.24%	(1.21)	0.48%	(4.33)
2	0.25%	(1.94)	-0.13%	(-0.71)	0.12%	(0.89)
3	0.15%	(1.57)	-0.21%	(-1.32)	-0.06%	(-0.51)
4	0.11%	(0.97)	0.03%	(0.12)	0.14%	(0.76)
5	0.18%	(2.73)	-0.22%	(-1.82)	-0.05%	(-0.44)
[6-10]	0.29%	(1.28)	-0.28%	(-0.75)	0.01%	(0.03)
[11-15]	0.49%	(1.76)	-0.91%	(-2.34)	-0.42%	(-1.83)
[16-20]	0.60%	(1.32)	-1.42%	(-1.86)	-0.82%	(-1.29)
[21-60]	-0.44%	(-0.35)	-0.32%	(-0.18)	-0.76%	(-0.54)

Table 5: WSB Reports and Future Returns - The Role of the Pandemic

This table repeats Table 3 after partitioning *Net DD* into *Net DD* × *Pre-Pandemic* and *Net DD* × *Post Pandemic,* where *Pre-Pandemic* is an indicator equal to one for July 2018 – March 15, 2020 and zero otherwise, and *Post-Pandemic* is an indicator equal to one for the pre-GME period that coincides with the pandemic (March 16, 2020- January 12, 2021) and zero otherwise. We also create analogous *Pre-Pandemic* and *Post-Pandemic* variables for *WSB Non-Research* and *Net SA*. The regression includes all the controls from Table 3, but in the interest of brevity, the estimates on controls are unreported. We also include calendar-day fixed effects, and we standardize all continuous variables to have mean zero and unit variance. Standard errors are clustered by firm and month, and *t*-statistics are reported in parentheses Below the regression estimates, we report formal tests for whether certain pairs of coefficients are significantly different from zero.

	Ret [1,5]	Ret [1,21]	Ret [1,5]	Ret [1,21]
	[1]	[2]	[3]	[4]
Net DD × *Pre-Pandemic*	0.96%	3.34%	1.00%	3.35%
	(1.93)	(2.22)	(2.03)	(2.16)
Net DD × *Post-Pandemic*	1.19%	5.44%	0.85%	1.70%
	(1.69)	(3.01)	(1.48)	(1.66)
Net DD × *Post GME*	0.03%	-5.58%	-0.45%	-2.82%
	(0.03)	(-3.84)	(-0.61)	(-2.00)
WSB NonResearch × *Pre-Pandemic*	0.14%	0.72%	0.13%	0.71%
	(0.91)	(1.15)	(0.86)	(1.13)
WSB NonResearch × *Post-Pandemic*	0.25%	5.99%	0.15%	0.09%
	(1.40)	(1.35)	(0.90)	(0.30)
WSB NonResearch × *Post GME*	1.02%	-5.23%	-0.45%	-1.37%
	(3.46)	(-1.37)	(-2.40)	(-3.77)
Net SA × *Pre-Pandemic*	0.33%	0.75%	0.33%	0.75%
	(3.44)	(2.64)	(3.47)	(2.65)
Net SA × *Post-Pandemic*	0.32%	0.39%	0.31%	0.44%
	(3.33)	(1.14)	(3.70)	(1.33)
Net SA × *Post GME*	0.12%	0.59%	0.00%	0.57%
	(0.70)	(1.22)	(0.01)	(1.19)
Net DD (Pre-Pandemic − Post-Pandemic)	-0.23%	-2.10%	0.16%	1.65%
	(-0.30)	(-0.67)	(0.23)	(0.95)
WSB NonResearch (Pre-Pandemic − Post-Pandemic)	-0.11%	-5.28%	-0.02%	0.62%
	(-0.53)	(-1.16)	(-0.12)	(1.23)
Net SA (Pre-Pandemic − Post-Pandemic)	0.01%	0.37%	0.02%	0.31%
	(0.07)	(1.14)	(0.22)	(0.93)
Obs. (Firm-Days)	2,772,053	2,772,053	2,770,545	2,770,545
Day FE	Yes	Yes	Yes	Yes
Table 3 Controls	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	Yes	No	No

Table 6: WSB Reports and Future Returns - Information Processing vs. Information Production

This table repeats Table 3 after partitioning all DD reports into information processing versus information production reports. *Processing* is an indicator equal to one if the report is issued around a major information event, defined as an earnings announcement issued on the previous or current day (-1, 0), abnormal media coverage on the previous or current day, or an analyst revision on the previous or current day. *Net DD Processing* is the *Net DD* measure computed for the subset of reports where *Processing = 1,* and *Net DD Production* is the *Net DD* measure computed for the subset of reports where *Processing = 0.* All other variables are defined as in Table 3. The regression includes all the controls from Table 3, but in the interest of brevity, the estimates on controls are unreported. We also include calendar-day fixed effects, and we standardize all continuous variable to have mean zero and unit variance. Standard errors are clustered by firm and month, and *t*-statistics are reported in parentheses. Below the regression estimates, we also report a formal test of whether certain pairs of coefficients are significantly different from zero.

	Ret [1,5]	*Ret [1,21]*	*Ret [1,5]*	*Ret [1,21]*
	[1]	[2]	[3]	[4]
Net DD Processing	0.69%	3.95%	0.32%	2.27%
	(0.84)	(2.09)	(0.50)	(2.05)
Net DD Processing × *Post GME*	-0.09%	-2.00%	0.07%	-2.52%
	(-0.06)	(-0.88)	(0.08)	(-1.69)
Net DD Production	1.41%	5.99%	1.35%	2.39%
	(2.93)	(2.99)	(2.79)	(2.52)
Net DD Production × *Post GME*	0.09%	-7.03%	-0.94%	-3.87%
	(0.06)	(-5.39)	(-1.37)	(-2.66)
WSB NonResearch	0.19%	3.22%	0.14%	0.42%
	(1.38)	(1.32)	(1.02)	(1.03)
WSB NonResearch × *Post GME*	1.11%	-2.56%	-0.43%	-1.70%
	(5.38)	(-1.26)	(-2.74)	(-3.67)
Net SA	0.31%	0.56%	0.31%	0.63%
	(4.05)	(2.33)	(4.10)	(2.54)
Net SA × *Post GME*	0.12%	0.44%	-0.01%	0.39%
	(0.75)	(1.09)	(-0.03)	(0.99)
Net DD (Processing – Production)	-0.72%	-2.03%	-1.03%	-0.13%
	(-0.90)	(1.44)	(1.46)	(0.14)
Net DD (Processing – Production) × *Post GME*	-0.18%	5.04%	1.01%	1.35%
	(-0.09)	(3.80)	(-1.18)	(1.06)
Obs. (Firm-Days)	2,772,053	2,772,053	2,770,545	2,770,545
Day FE	Yes	Yes	Yes	Yes
Table 3 Controls	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	Yes	No	No

Table 7: WSB Reports and Future Cash-Flow News

This table repeats the analysis in Table 3 after replacing the dependent variable, returns, with one of three proxies for cash-flow news, measured over the subsequent week (i.e., x = 5 trading days) or the subsequent month (x = 21 trading days). Cash-flow news is measured as either *Media Sentiment (Media)*, computed as the sum of the daily Bloomberg sentiment score; *Positive Forecast Error* (*Pos FE*), an indicator equal to one if the realized earnings exceed the median quarterly forecast across all I/B/E/S analysts as of day *t,* and *Positive Forecast Revision* (*Pos FR*) computed as the number of upward revisions by I/B/E/S analysts scaled by the total number of revisions. All other variables are defined as in Table 3. More detailed variable definitions are available in Appendix C. Specifications 1 and 2 report the results for the full sample for five-day and 21-day measures of *Media Sentiment*. Specifications 3 and 4 (5 and 6) report analogous results for *Positive Forecast Error (Positive Forecast Revision*). Standard errors are clustered by firm and month, and *t*-statistics are reported below each estimate. Below the regression estimates, we report formal tests for whether certain pairs of coefficients are significantly different from zero.

	Media [1,5]	*Media [1,21]*	*Pos FE [1,5]*	*Pos FE [1,21]*	*Pos FR [1,5]*	*Pos FR [1,21]*
	[1]	[2]	[3]	[4]	[5]	[6]
Net DD	5.38%	18.49%	4.22%	3.51%	2.22%	1.92%
	(2.37)	(2.62)	(2.37)	(1.62)	(1.73)	(1.52)
Net DD × Post GME	-4.71%	-25.37%	-17.11%	-13.86%	-3.00%	-2.83%
	(-2.05)	(-3.24)	(-5.94)	(-5.65)	(-2.34)	(-1.96)
WSB NonResearch	-1.19%	-2.70%	0.34%	-2.16%	0.55%	0.14%
	(-1.27)	(-0.53)	(0.16)	(-1.40)	(0.95)	(0.19)
WSB NonResearch × Post GME	0.10%	-1.75%	-0.58%	1.64%	-1.06%	-1.17%
	(0.11)	(-0.35)	(-0.26)	(0.79)	(-1.76)	(-1.74)
Net SA	-0.84%	-7.94%	-1.83%	-0.16%	-0.15%	-0.10%
	(-1.26)	(-2.83)	(-1.65)	(-0.18)	(-0.57)	(-0.28)
Net SA × Post GME	0.62%	2.39%	1.66%	1.11%	0.58%	0.42%
	(0.67)	(0.82)	(0.42)	(0.74)	(0.75)	(0.31)
Log (Size)	0.65%	3.21%	3.00%	3.52%	-0.09%	0.14%
	(5.47)	(5.66)	(10.15)	(11.04)	(-0.44)	(0.49)
Log (BM)	-0.85%	-4.26%	-1.37%	-0.90%	-0.53%	-1.17%
	(-6.62)	(-6.75)	(-2.05)	(-1.13)	(-3.37)	(-3.80)
Ret [0]	18.90%	31.06%	15.55%	13.40%	9.18%	13.10%
	(7.93)	(7.73)	(3.15)	(6.01)	(7.67)	(8.72)
Ret [-5, -1]	2.94%	6.21%	10.75%	11.18%	5.94%	10.28%
	(4.37)	(2.54)	(2.35)	(4.74)	(8.35)	(9.76)
Ret [-26, -6]	0.48%	3.49%	8.19%	5.89%	4.07%	7.27%
	(1.37)	(1.98)	(3.97)	(2.21)	(5.35)	(6.32)
News Sentiment [0]	33.26%	87.09%	2.17%	2.08%	2.43%	2.49%
	(34.13)	(25.06)	(3.18)	(4.01)	(7.93)	(9.86)
News Sentiment [-5, -1]	15.53%	56.24%	0.98%	1.61%	1.03%	1.35%
	(22.72)	(20.96)	(1.85)	(4.00)	(10.51)	(10.23)

News Sentiment [-26, -6]	7.90%	30.76%	0.95%	0.75%	0.33%	0.61%
	(18.59)	(16.15)	(3.93)	(3.20)	(5.49)	(6.25)
Net DD + Net DD × Post GME	0.67%	-6.88%	-12.89%	-10.35%	-0.78%	-0.91%
	(1.33)	(-1.83)	(-4.94)	(-7.67)	(-1.50)	(-1.21)
Net DD − NonResearch	6.57%	21.19%	3.88%	5.67%	1.67%	1.78%
	(2.75)	(2.94)	(1.27)	(2.40)	(1.26)	(1.25)
Net DD × Post −NonResearch × Post	-4.81%	-23.62%	-16.53%	-15.50%	-1.94%	-1.66%
	(-2.01)	(-3.03)	(-4.64)	(-4.63)	(-1.50)	(-1.07)
Net DD − Net SA	6.22%	26.43%	6.05%	3.67%	2.37%	2.02%
	(2.63)	(-3.37)	(3.00)	(1.56)	(1.79)	(1.62)
Net DD × Post − Net SA × Post	-5.33%	-27.76%	-18.77%	-14.97%	-3.58%	-3.26%
	(-2.18)	(-3.05)	(-4.34)	(-5.62)	(-2.11)	(-1.62)
Obs. (Firm-Days)	2,772,053	2,772,053	164,018	642,522	1,964,351	1,964,351
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: Changes in Contributor Strategies in the Post-GME Period

This table reports estimates from the following regression:

$$Y_{it} = \alpha + \beta_1 Post_t + \varepsilon_{it.}$$

The dependent variable, *Y*, is an indicator that classifies a DD report as either a price pressure report (Panel A) or an attention-based report (Panel B). We classify a report as a price pressure report if either 1) the number of price pressure words exceed the number of fundamental words (*PP*) or 2) the report contains at least one price pressure word (*PP2*). Appendix C provides the list of price pressure and fundamental words. We classify a report as attention-based (*High Attention*) if either 1) the absolute return on the day prior to the DD report was in the top decile (*High Absolute Return*) or 2) the firm had more than one non-research post issued on *WSB* over the previous five trading days (*High WSB Posts*). We also report the results separately for *High Absolute Return* and *High WSB Post.* The independent variable, *Post,* is an indicator equal to one for the *Post-GME* period, as defined in Table 1, and zero otherwise. Standard errors are clustered by firm and date, and *t*-statistics are reported below each estimate.

Panel A: Price Pressure Reports

	PP	*PP*	*PP2*	*PP2*
	[1]	[2]	[3]	[4]
Intercept	11.63%	7.91%	21.74%	17.31%
	(3.29)	(6.00)	(5.35)	(11.63)
Post GME	19.40%	17.25%	28.51%	29.67%
	(7.45)	(5.67)	(10.07)	(9.44)
Obs. (DD Reports)	5,015	4,409	5,015	4,409
Include GME & AMC	Yes	No	Yes	No

Panel B: Attention-Based Reports

	High Attention	*High Attention*	*High Absolute Return*	*High WSB Posts*
	[1]	[2]	[3]	[4]
Intercept	32.53%	30.96%	21.16%	17.47%
	(10.00)	(10.10)	(11.88)	(5.21)
Post GME	24.55%	20.46%	12.25%	25.49%
	(5.70)	(4.96)	(3.39)	(4.85)
Obs. (DD Reports)	5,015	4,409	5,015	5,015
Include GME & AMC	Yes	No	Yes	Yes

Table 9: WSB Reports and Future Returns - Price Pressure and Attention Reports

This table examines the incremental informativeness of price pressure reports and high-attention reports. We classify a report as a price pressure report if the number of price pressure words exceed the number of fundamental words (see Appendix C for the list of price pressure and fundamental words). We classify as a report as *High Attention* if either 1) the absolute return on the day prior to the DD report was in the top decile (*High Absolute Return*) or if the firm had more than one non-research post issued on *WSB* over the previous five trading days (*High WSB Posts*); and we classify a report as *PP/Attention* if the report is classified as either *Price Pressure* or *High Attention*. *Net DD PP* is the *Net DD* measure computed for the subset of reports classified as price pressure, and *Net DD Attention* and *Net DD PP/Attention* are computed analogously. Other variables are defined as in Table 3. The regression includes all the controls from Table 3, but in the interest of brevity, the estimates for controls are unreported. We also include calendar-day fixed effects, and we standardize all continuous variables to have mean zero and unit variance. Standard errors are clustered by firm and month, and *t*-statistics are reported in parentheses.

	Ret [1,21]	*Ret [1,21]*	*Ret [1,21]*	*Ret [1,21]*	*Ret [1,21]*	*Ret [1,21]*
	[1]	[2]	[3]	[4]	[5]	[6]
Net DD	2.11%	2.17%	1.88%	1.41%	1.41%	1.31%
	(1.98)	(2.19)	(2.35)	(2.12)	(2.03)	(2.10)
Net DD × Post GME	0.05%	-2.53%	-0.61%	-0.81%	-0.15%	-0.45%
	(0.03)	(-1.46)	(-0.36)	(-0.54)	(-0.14)	(-0.35)
Net DD PP	31.73%	1.93%			8.88%	
	(2.03)	(1.33)			(2.40)	
Net DD PP × Post GME	-35.80%	-3.24%				
	(-2.15)	(-1.85)				
Net DD Attention			8.57%	2.55%		
			(2.25)	(1.60)		
Net DD Attention × Post GME			-10.58%	-5.16%		
			(-3.10)	(-2.84)		
Net DD PP/Attention					8.88%	2.57%
					(2.40)	(1.82)
Net DD PP/Attention × Post					-10.59%	-5.08%
					(-3.24)	(-3.22)
WSB NonResearch	3.44%	0.40%	3.31%	0.33%	3.32%	0.33%
	(1.40)	(1.03)	(1.30)	(0.89)	(1.29)	(0.89)
WSB NonResearch × Post GME	-2.71%	-1.64%	-2.45%	-1.51%	-2.50%	-1.54%
	(-1.27)	(-3.75)	(-1.14)	(-3.57)	(-1.17)	(-3.63)
Net SA	0.60%	0.65%	0.59%	0.65%	0.60%	0.65%
	(2.58)	(2.69)	(2.52)	(2.69)	(2.52)	(2.70)
Net SA × Post GME	0.36%	0.35%	0.38%	0.36%	0.38%	0.35%
	(0.88)	(0.89)	(0.94)	(0.91)	(0.93)	(0.90)
Obs. (Firm-Days)	2,772,053	2,770,545	2,772,053	2,770,545	2,772,053	2,770,545
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Table 3 Controls	Yes	Yes	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	No	Yes	No	Yes	No

Table 10: WSB Price Pressure and Attention Reports - Alternative Return Horizons

This table repeats the analysis in Specification 6 of Table 9 (the sample that excludes GME and AMC) after varying the return horizon. For reference, the first row, reports the baseline 21-day holding period. We also report results for returns measured on the current day (i.e., Day 0), each of the subsequent five days, the subsequent 2, 3, and 4 weeks, and the subsequent 5 through 12 weeks. Standard errors are clustered by firm and month, and *t*-statistics are reported next to each estimate.

	Net DD PP/Attention		*Net DD PP/Attention × Post GME*		*Net DD PP/Attention + Net DD PP/Attention × Post GME*	
Ret. Period	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*
[1,21]	3.87%	(1.99)	-5.53%	(-2.86)	*-1.67%*	(-1.66)
0	*1.11%*	(3.24)	-0.57%	(-1.56)	*0.54%*	(4.45)
1	*0.29%*	(1.06)	0.26%	(0.87)	*0.55%*	(6.83)
2	*0.65%*	(2.19)	-0.50%	(-1.43)	*0.15%*	(0.90)
3	*0.33%*	(2.09)	-0.41%	(-1.82)	*-0.08%*	(-0.67)
4	0.04%	(0.32)	0.06%	(0.27)	0.10%	(0.56)
5	0.36%	(1.98)	-0.42%	(-1.98)	-0.06%	(-0.54)
[6-10]	*0.31%*	(0.87)	-0.54%	(-1.40)	*-0.23%*	(-0.78)
[11-15]	*0.39%*	(0.59)	-0.93%	(-1.24)	*-0.54%*	(-2.28)
[16-20]	*1.43%*	(1.37)	-2.45%	(-2.11)	*-1.02%*	(-1.69)
[21-60]	*-0.39%*	(-0.16)	-0.69%	(-0.25)	*-1.08%*	(-0.70)

Table 11: WSB Reports and Investor Order Imbalances

This table reports estimates from the following regression:

$$Y_{it} = \beta_1 Net\ DD_{it} + \beta_2 Net\ DD_{it} \times Post_t + \beta_3 NonResearch_{it} + \beta_4 NonResearch_{it} \times Post_t + \beta_5 SA_{it} + \beta_6 SA_{it} \times Post_t + Controls_{it} + Day_t + \varepsilon_{it}.$$

The dependent variable, *Y*, is one of six order imbalance variables. *Institutional % Imbalance is* defined as institutional buy share volume less inst. sell share volume (*Institutional Imbalance*) scaled by total inst. share volume. *Large Retail % Imbalance* and *Small Retail % Imbalance* are defined analogously after replacing *Inst. Share Volume* with *Retail Share Volume* and *Retail Number of Trades,* respectively. *Inst. Std Abnormal Imbalance is* defined as the *institutional imbalance* on day *t* less the average *institutional imbalance* over days *t-120* through *t-240*, scaled by the standard deviation of institutional imbalance estimated over the same window. *Large* (*Small) Retail Std Abnormal Imbalance* are defined analogously. All imbalance measures are converted to percentile rankings. Detailed definitions for all variables are in Appendix C. Standard errors are clustered by firm and month, and t-statistics are reported below each estimate. Below the regression estimates, we report formal tests for whether certain pairs of coefficients are significantly different from zero.

	[1] *Institutional % Imbalance*	*[2]* *Large Retail % Imbalance*	*[3]* *Small Retail % Imbalance*	*[4]* *Institutional Std. Abn. Imb.*	*[5]* *Large Retail Std. Abn. Imb.*	*[6]* *Small Retail Std. Abn. Imb.*
Net DD	-0.19	1.62	4.80	-0.67	2.43	6.09
	(-0.60)	(3.82)	(6.30)	(-1.12)	(2.86)	(6.09)
Net DD × Post GME	0.08	-1.02	0.95	-0.09	-2.10	-1.28
	(0.18)	(-2.25)	(0.90)	(-0.09)	(-2.13)	(-1.04)
WSB NonResearch	-0.22	0.10	1.55	0.24	-1.17	1.74
	(-1.50)	(0.66)	(3.80)	(0.64)	(-3.12)	(3.99)
WSB NonResearch × Post GME	0.32	-0.12	-1.43	0.26	1.21	-1.09
	(2.24)	(-0.70)	(-2.43)	(0.58)	(2.87)	(-1.71)
Net SA	0.02	0.76	2.12	0.22	0.31	1.46
	(0.09)	(5.53)	(9.71)	(0.88)	(1.01)	(5.14)
Net SA × Post GME	0.32	0.22	0.75	0.15	0.68	-0.80
	(1.29)	(0.91)	(1.40)	(0.32)	(1.37)	(-1.24)
Log (Size)	0.94	0.28	0.22	0.05	0.28	0.41
	(14.84)	(9.57)	(3.54)	(0.80)	(5.12)	(5.54)
Log (BM)	0.01	-0.05	-0.19	0.05	0.06	0.18
	(0.19)	(-1.22)	(-2.69)	(0.87)	(1.20)	(2.54)
Ret [-1]	-7.53	10.30	9.03	-15.06	16.04	13.43
	(-6.48)	(8.46)	(6.71)	(-11.47)	(10.83)	(9.61)
Ret [-5, -2]	6.53	-5.57	-8.68	7.19	-4.38	-6.90
	(6.07)	(-14.28)	(-9.14)	(6.11)	(-6.84)	(-7.40)
Ret [-26, -6]	1.12	-1.94	-4.17	0.95	-1.42	-3.05
	(2.53)	(-9.70)	(-12.69)	(1.88)	(-4.25)	(-9.11)
News Sentiment [-1]	0.31	0.12	0.11	0.50	0.16	0.19
	(3.38)	(1.23)	(1.07)	(3.80)	(1.20)	(1.41)
News Sentiment [-5, -2]	0.14	0.17	0.14	0.15	0.19	0.18
	(2.88)	(3.02)	(2.08)	(2.43)	(2.56)	(1.99)
News Sentiment [-26, -6]	-0.03	0.03	0.01	-0.06	0.03	0.03
	(-1.52)	(1.35)	(0.29)	(-2.51)	(0.88)	(0.76)

Abs. Ret [-1]	6.89	-5.73	-3.29	5.31	-13.30	-10.05
	(6.12)	(-4.24)	(-2.53)	(4.19)	(-6.82)	(-6.99)
Abs. Ret [-5, -2]	-4.04	4.55	8.75	-7.77	2.57	8.23
	(-3.43)	(15.61)	(13.16)	(-5.69)	(4.18)	(14.00)
Abs. Ret [-26, -6]	-1.49	1.41	4.70	-3.06	0.21	4.06
	(-3.71)	(6.07)	(11.45)	(-6.35)	(0.64)	(8.92)
Abs. News Sentiment [-1]	0.52	-0.47	-0.54	0.57	-0.88	-0.71
	(3.80)	(-3.14)	(-3.40)	(3.66)	(-4.23)	(-4.15)
Abs. News Sentiment [-5, -2]	0.11	-0.03	0.04	0.01	-0.17	-0.01
	(1.62)	(-0.53)	(0.53)	(0.07)	(-2.37)	(-0.09)
Abs. News Sentiment [-26, -6]	0.05	0.00	0.09	-0.02	-0.01	0.03
	(1.68)	(0.10)	(1.88)	(-0.42)	(-0.32)	(0.57)
Inst. % Imbalance [-1]	0.14	-0.01	-0.01	-0.12	-0.01	0.00
	(44.52)	(-3.01)	(-4.25)	(-16.10)	(-5.43)	(0.34)
Large Retail % Imbalance [-1]	0.00	0.01	-0.04	0.00	-0.12	-0.03
	(-1.39)	(2.83)	(-14.66)	(0.36)	(-19.51)	(-14.50)
Small Retail % Imbalance [-1]	-0.01	0.04	0.12	0.00	-0.03	-0.17
	(-3.60)	(16.44)	(19.25)	(1.49)	(-8.03)	(-25.14)
Inst. Std. Abn. Imbalance [-1]	0.02	0.00	0.00	0.30	0.00	-0.01
	(8.28)	(-2.75)	(2.13)	(34.70)	(0.36)	(-1.80)
Large Retail Std. Abn. Imbalance [-1]	-0.01	0.02	0.02	-0.01	0.16	0.02
	(-6.34)	(7.82)	(9.21)	(-5.06)	(20.49)	(6.66)
Small Retail Std. Abn. Imbalance [-1]	0.01	0.00	0.01	-0.01	0.07	0.33
	(3.34)	(0.52)	(1.77)	(-1.79)	(19.35)	(36.89)
Net DD + Net DD × Post	-0.11	0.60	5.76	-0.76	0.33	4.82
	(-0.41)	(3.99)	(7.18)	(-1.05)	(0.76)	(7.24)
Net DD − NonResearch	0.03	1.52	3.26	-0.92	3.61	4.36
	(0.04)	(3.31)	(3.81)	(-0.84)	(4.40)	(3.80)
(Net DD − NonResearch) × Post	-0.24	-0.90	2.38	-0.35	-3.32	-0.19
	(-0.35)	(1.69)	(2.05)	(-0.27)	(-3.21)	(-0.14)
Net DD − Net SA	-0.21	0.87	2.69	-0.90	2.13	4.63
	(-0.61)	(1.92)	(3.23)	(-1.04)	(2.38)	(4.17)
(Net DD 21 − Net SA) × Post	-0.24	-1.24	0.21	-0.24	-2.79	-0.47
	(-0.30)	(-2.25)	(-0.17)	(-0.18)	(-2.69)	(-0.33)
Obs. (Firm-Days)	2,494,475	2,494,475	2,494,475	2,523,578	2,523,578	2,523,578
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	Yes	Yes	Yes	Yes	Yes

Table 12: Informativeness of Small Retail Trading Following WSB DD Reports

This table reports estimates from the following regression:

$$R_{it+1,t+21} = \beta_1 Imb_{it} + \beta_2 Imb_{it} \times Post_t + \beta_3 Imb_{it} \times DD_{it} + \beta_4 Imb_{it} \times DD_{it} \times Post_t + \beta_5 Imb_{it} \times NR\ Indicator_{it} + \beta_6 Imb_{it} \times NR\ Indicator_{it} \times Post_t + \beta_7 Imb_{it} \times SA_{it} + \beta_8 Imb_{it} \times SA_{it} \times Post_t + Controls_{it} + Day_t + \varepsilon_{it}.$$

The dependent variable is the one-month ahead return. *Imb* is *Std Abnormal Small Retail Imbalance* (as defined in Table 11). *DD* is an indicator equal to one if a *DD* report was issued for firm *i* on day *t*. *NR Indicator* and *SA* are indicators equal to one if a non-research post or SA report was issued for firm *i* on day *t*, respectively. *Controls* and *Day* are defined as in Table 3, and in the interest of brevity, the estimates on controls are unreported. We standardize all continuous variable to have mean zero and unit variance. Standard errors are clustered by firm and month, and *t*-statistics are reported in parentheses Below the regression estimates, we report formal tests for whether certain pairs of coefficients are significantly different from zero.

	[1]	[2]	[3]	[4]
Small Retail Imbalance	0.05%	0.05%	0.05%	0.05%
	(2.26)	(2.27)	(2.29)	(2.28)
Small Retail Imbalance × Post	-0.09%	-0.10%	-0.09%	-0.10%
	(-1.50)	(-1.61)	(-1.53)	(-1.62)
Small Retail Imbalance × DD	1.39%	0.54%	0.62%	0.35%
	(1.88)	(1.79)	(2.65)	(1.57)
Small Retail Imbalance × DD × Post	-1.50%	-1.11%	-1.01%	-0.76%
	(-1.33)	(-3.05)	(-1.13)	(-2.37)
Small Retail Imbalance × NR Indicator	0.80%	0.08%	0.09%	0.00%
	(1.05)	(0.38)	(0.38)	(0.02)
Small Retail Imbalance × NR Indicator × Post	-1.03%	-0.73%	-0.68%	-0.46%
	(-0.69)	(-2.11)	(-0.56)	(-1.96)
Small Retail Imbalance × SA	0.03%	0.02%	0.00%	0.00%
	(0.51)	(0.39)	(-0.04)	(0.03)
Small Retail Imbalance × SA × Post	-0.15%	-0.11%	-0.16%	-0.10%
	(-1.66)	(-1.20)	(-1.65)	(-1.06)
Net DD			4.97%	1.88%
			(2.46)	(2.07)
Net DD × Post			-4.28%	-2.44%
			(-3.20)	(-1.68)
NonResearch			4.92%	0.53%
			(1.24)	(1.08)
NonResearch × Post			-3.99%	-1.75%
			(-1.15)	(-3.03)
Net SA			0.58%	0.65%
			(2.29)	(2.64)
Net SA × Post			0.54%	0.48%
			(1.33)	(1.18)
Imb. × *DD + Imb.* × *DD × Post*	-0.11%	-0.56%	-0.40%	-0.41%
	(-0.16)	(-2.67)	(-0.50)	(-1.69)
Imb. × *NR Indicator + Imb.* × *NR Indicator × Post*	-0.23%	-0.65%	-0.59%	-0.46%
	(-0.23)	(-2.36)	(-0.53)	(-2.86)
Imb. × *SA + Imb.* × *SA × Post*	-0.13%	-0.09%	-0.16%	-0.10%
	(-1.54)	(-1.16)	(-1.84)	(-1.23)
Obs. (Firm-Days)	2,523,578	2,522,075	2,523,578	2,522,075
Day FE	Yes	Yes	Yes	Yes
Table 3 Controls	Yes	Yes	Yes	Yes
Include GME & AMC	Yes	No	Yes	No