Review of Corporate Finance, 2022, 2: 151–190

We Reddit in a Forum: The Influence of Message Boards on Firm Stability

Shaen Corbet[1,2], Yang (Greg) Hou[2], Yang Hu[2] and Les Oxley[2*]

[1] *DCU Business School, Dublin City University, Dublin 9, Ireland; shaen.corbet@dcu.ie*

[2] *School of Accounting, Finance and Economics, University of Waikato, New Zealand*

ABSTRACT

The use of messaging boards to instigate coordinated manipulation of stock prices is not a novel phenomenon. However, the growing breadth and sophistication of social media, the widespread availability of technological cloaking techniques, and the ready accessibility of leveraged derivatives have all significantly contributed to the growth of the infamous Reddit forum r/wallstreetbets. Using a lexicon of terminology designed to identify explicit and implicit manipulation attempts, this research presents several novel findings. The results indicate significantly positive, pronounced, and persistent abnormal returns in the aftermath of manipulative events, with the results being robust across several testing procedures. These abnormal returns have increased significantly in line with the growth of forum users and the reach of manipulation-related comments. Significant effects on market liquidity and analyst recommendations are further identified across tests. Such continued predatory headwinds are certainly of interest to market-makers, regulators, and policymakers alike, as the irrational exuberance incited by millennial meme stocks and sarcastic GIFs is found to have played a significant role in disrupting market functionality. The desired outcome of some of these five million monthly forum users is to create enough momentum to trigger algorithmic responses to move out-of-the-money options into a profitable state. Our results present evidence that this potential threat to corporate stability is concrete.

*Corresponding Author: Les Oxley, les.oxley@waikato.ac.nz.

Online Appendix available from:
http://dx.doi.org/10.1561/114.00000014_app
ISSN 2693-9312; DOI 10.1561/114.00000014

Keywords: Market activity, liquidity, volatility, pump-and-dump, manipulation, messaging boards

JEL Codes: G12, G13, G14, G41

1 Introduction

Throughout the period 2018 through 2021, social media channels began to discuss the presence of a new phenomenon of manipulative information dissemination for prospective day traders through the Reddit forum r/wallstreetbets. This development may seem, at first sight, difficult to differentiate from that highlighted in previous work showing that messaging boards can influence stock prices (Das and Chen, 2007; Sabherwal *et al.*, 2008). However, due to the increasingly important role of algorithmic trading and the increased speed of information dissemination, the capacity to manifest market influence is much more pronounced than it was during the dotcom bubble. When we consider broad technological availability, substantive financial education, and the widespread, immediate availability of platforms offering leveraged options, we can identify key differentials in comparative behavior between messaging board eras. While the content of many messaging boards is available for all to see, the scale of the number of Reddit users of this forum, which grew from 206,205 registered users in January 2018 to over 10 million now, substantially overshadows that of similar platforms and social media influencers.

The forum description reads "*Like 4chan found a Bloomberg Terminal*". Contained within the forum is, for the inexperienced trader, apparently infinite wisdom and teachings presented in a raft of sarcastic terminology, where self-described "autists" (referring to autistic persons or r/wallstreetbets users, who characterize themselves as having an abnormal and unhealthy focus or persistence or, in this case, an obsession with profits) seek "tendies" (described within the forum as gains earned from an investment, usually by amateur investors day trading on the platform Robinhood). Should the trade not go according to plan, there is always the "long $ROPE" investment alternative (a quite disturbing reference to suicide should a position lead to a loss), based on the motivation that "YOLO" ("you only live once"). The sarcasm of this millennial lexicon betrays an undertone of desperation—that is, the hope that risking everything can lead to exceptional profits. There is some evidence of partial successes in this regard; for example, in August 2020, Reddit user u/LaikaBowls commented, "Laid off due to COVID. Moved my 401k into an IRA. 'Invested' it all on SPX fd calls today . . .." along with a picture of the user's trading account showing profits for 1 day, 24 August 2020, of $270,508.45. A review of many of the posts reveals that they commonly contain inaccurate analyses (often referred to as DD or due diligence) and pseudomathematics alongside a range of memes and sarcastic statements; there are, for example, detailed explanations of the "kangaroo market" that is purported to exist between bull and bear markets and of the logic for why initiating a "pump-and-dump" on the CBOE VIX cannot generate losses. The forum often quickly descends into chaos, with contributors, for instance, moving from substantiated and somewhat educated debates about financial markets to stating that the word "trader" is an anagram of the slur "retard" or announcing as a badge of honor that they have been blocked by the Federal Reserve (@federalreserve) Twitter account. Some posts also directly attempt to instigate illicit behaviour:

> *"We need to more efficiently aim our autism, and we need a consensus on the next meme stock to buy mass amounts of otm calls on so we can pool our bet. Submit your pick for the ticker you feel is most likely to have unfounded growth soon so we can find our next rising star!"*
>
> *"Also if this happens to be illegal, I don't know about it so I'm safe."*
>
> *Tues. 18 Feb 2020, 23:20 GMT, U/AssPowers, comments available here*

In regards to the motivation and rationales of several traders on the r/wallstreetbets forum, many posts refer to similarities between the preferred tactics of "autists" and those of the perceived "True WSB legends," who in the 1980s were the inspiration behind the "O'Hare play," in which Chicago traders would bet heavily, often using all available resources, on a commodity of choice and then head to O'Hare airport. Should the trade lead to a profit, the trader would leave Chicago for an extravagant holiday. However, if the trade resulted in a total loss, the trader would buy a one-way ticket and leave to relocate and disappear.

There are many specific examples of premeditated, unethical behaviour. In one case, user u/SuXs posted an "update" on 6 February 2020 with the title "Pumping $MSFT to the moon with LOIC, and how to profit from it (Satire) (Fiction)." The user continues to state throughout that this is a fictional piece where "I pretended that my goal was to see if a collective of autists such as r/WSB can use the integrated leverage of Call contracts 100:1 ratios; Institutions algorithms risk hedging strategies; Social Media; WEAPONISED.

AUTISM." This post is based on a particular sequence of events that had occurred in the preceding days.[1]

2 Theoretical Development

In this research, we set out to establish whether the development and expansion of the now infamous r/wallstreetbets forum on Reddit influenced any or all of our selected measures in the form of abnormal returns, market liquidity and the research and recommendations of industry analysts with regards to companies selected by Reddit users as targets of both implicit and explicit pumps and dumps.[2] Such manipulation techniques have been identified to consist of "back and forth" trading to attract uninformed investors (Khwaja and Mian, 2005), while other manipulative tactics have been instigated through the creation of "short squeezes" (Jiang *et al.*, 2005). A brief discussion surrounding the legality of such behavior is presented in Section A of the Online Appendix. To date, many of these practices have been found to breach both SEC and FBI regulations, which forbid (a) buying or selling of a security in breach of a fiduciary duty or other relationship of trust and confidence, based on material, non-public information about the security, and (b) manipulation through increased trading volume generated by inducing unwitting investors to purchase shares of the targeted security through false or deceptive sales practices and/or public information releases. Within this context, it is possible that the mere mention of a company name in such a forum could be a catalyst

[1] On 4 February 2020, user u/SuXs stated, "LOL BLOOMBERG ADMITTING THAT AS LONG AS WE BUY THE CALLS THE STOCKS WILL GO UP BECAUSE OF HEDGING ALGORITHMS," explicitly pointing out that Bloomberg had presented an argument making clear that group purchases of options relating to stocks would generate further support through algorithmic purchases. /LifterPuller responded, "WSB found an infinite money cheat code. Literally can't go tits up." Less than 6 hours later, in a separate forum, /SuXs stated, "Bloomberg confirmed how to rig the market. Buy $MSFT for free tendies," a post identified in the following research as an implicit attempt to "pump" Microsoft's stock price. This post garnered substantial attention.

[2] In February 2020, a Bloomberg article presented a viewpoint that the actions of r/WallStreetBets investors were akin to "a kind of guerrilla warfare in the markets, trying to exploit what they see as weaknesses in the system to move prices where they want them." Considering a range of efficiency arguments (Antweiler and Frank, 2004), such organized behavior remains illegal throughout a range of international jurisdictions. When defending actions that could potentially constitute market manipulation. Such a context surrounding the potential irrational exuberance sourced within such forums can be observed in the period before the dot-com collapse and the growth of Internet chat rooms, where such sentiment acted as a significant preamble to forthcoming market volatility (Bhattacharya *et al.*, 2009). However, now, through the use of options and a variety of leverage-based products, such traders using coordinated action can theoretically manipulate stock prices to take advantage of further algorithmically driven processes through which exceptionally inexpensive options can sharply and rapidly appreciate.

for potential irrational exuberance, considering that over 10.0 million users now monitor the forum. To establish the degree of influence of such forums, we identify, based on a robust lexicon of search terms, an algorithmically verified set of online statements intended, either implicitly or explicitly, to instigate a uniform investment response from Reddit users. If the author of the post directly states her intention to open a position, now or at some time in the future, to instigate others to do the same, we regard this as a case of explicit instigation. Our identification of implicit instigation incorporates further use of a lexicon that uses common "slang" and techniques previously used in the r/wallstreetbets forum. For example, one technique involves users including terms such as "(satire)" and "(fiction)" in their statements to provide both a signal and cover for any future third-party independent inquiry into actions that could potentially constitute market manipulation. Reddit user u/recentlyunearthed stated in jest, "*How can we have insider knowledge when we don't have any knowledge*?"

We utilize a broad range of specific research areas to create a foundation upon which to build our analysis. r/wallstreetbets, in essence, is very similar in structure to the messaging boards that developed throughout the growth and collapse of the dotcom bubble. Using classifier algorithms coupled with a voting scheme, Das and Chen (2007) extracted sentiment from stock message boards to identify a relationship between technology stock levels and both volume and volatility. This work built on that of Sabherwal *et al.* (2011) and Das *et al.* (2005), with the latter research building on the creation of sentiment and disagreement measures or, as the authors termed them, "eInformation" measures. Further, utilizing 400,000 S&P 500 stock-related Twitter messages, Sprenger *et al.* (2014) found that returns before good news events are more pronounced than those associated with bad news events, while Renault, 2017 provided evidence of sentiment-driven noise trading at the intraday frequency. We further develop upon the work of Akyildirim *et al.* (2020), Allen *et al.* (2021), Cioroianu *et al.* (2021), Philippi *et al.* (2021), Meegan *et al.* (2021) and Hu *et al.* (2021), which have focused on the growth of fintech and cryptocurrencies, some of which have also been subjected to the attention of message-board sentiment. Identifying links in such induced sentiment generation is important, but we must develop methods of measuring abnormal returns, as our research attempts to do. Manipulation of corporate perceptions was also studied by Dellarocas (2003), Dellarocas (2006), Jian and Sami (2012) and Luca and Zervas (2016). Ackert *et al.*, 2016 built a case that the use of messaging boards provides informational efficiency by targeting actively traded large firms. Whether such behavior can be deemed informationally efficient is somewhat contestable, considering the potential for jurisdictional and regulatory breaches (Comerton-Forde and Putniņš, 2015). In the research presented here, we focus specifically on issues relating to market manipulation and, in particular, the processes surrounding pumps and dumps.

3 Data

We collect data from multiple sources to analyze the interactions between r/wallstreetbets and associated sub-Reddit stock forums that have received attention with regard to both implicit and explicit intentions to pump and dump prices through both spot and options markets at daily frequency. We primarily develop a concise list of announcements that specifically constitute and state such intentions. To complete this task, we develop several strict rules to standardize the process across major international financial markets. The first rule is that the specified company must be a publicly traded company with an available stock ticker for the period of 1 January 2015 through 14 February 2021. The numbers of forum users and comments per day are presented in Figure 1. The identified comments were relatively infrequent in the period before 1 January 2015, leading us to select this as our sample start date. We elaborate these data based on a combined search of the Reddit forum r/wallstreetbets.

We devise two strategies to separate our list of observations to be investigated and further analyzed. We first set out to classify messages as positive ("bullish"), negative ("bearish") or indeed neutral (no distinct classification). Next, we explicitly search the groupings of positive and negative sentiment messages to extract those that distinctly refer to a pre-established lexicon[3] of terms relating to questionable manipulation.[4]

[3]We further verified the data search by using several other search tools (such as https://redditsearch.io/) and then using an iterative loop. We began by setting the epoch date to the start of the analysis, but due to limitations in terms of data collection (only 1,000 were available per analysis), we then utilize the created_utc function, as presented by user u/Watchful1/Sketchpad at the following link. Significant difficulty in collecting such data occurred in the period from January 24 though February 7, 2021, as daily comments exceeded 50,000 per day, peaking at 394,280 on January 28. This is in stark contrast to the period throughout the entirety of 2015 through 2020, where daily comments peaked at approximately 60,996 on March 18, 2020, averaging less than 10,000 per day throughout this period. We then developed a list of the most common users, denoted as the key influencers within the r/WallStreetBets forum, to search their own posts using the following search process, as provided by an adaptation of the searching blueprint provided by user u/Watchful1/Sketchp. There are numerous additional parameters that can be used when performing a comment search. A list (obtained from https://github.com/pushshift/api) of the common adapted terms that we utilized in our analysis is provided. To verify our final dataset in more detail, we further developed this code using individual reiterated time period searching analysis of the form available here, which then was reiterated on a continuous basis in 1,000 observation blocks. Datasets were merged, and duplicates were removed.

[4]Reddit data were collected using the pushshift.io Reddit API (https://github.com/pushshift/api), which provides enhanced functionality for searching Reddit comments and submissions and represents the main avenue for programmatically accessing and collecting Reddit data for academic research. The pushshift.io Reddit API was designed and created by the r/datasets mod team to help provide enhanced functionality and search capabilities for searching Reddit comments and submissions.

a) Number of r/wallstreetbets forum subscribers b) Number of r/wallstreetbets comments per day

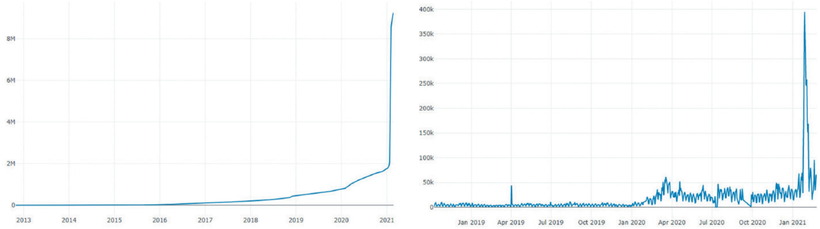


Figure 1: r/wallstreetbets Forum Statistics.

Note: Data obtained from subredditstats.com.

To develop significant methodological robustness with regard to developing a suitable lexicon, a variety of types were considered in the aftermath of the collection of r/WallStreetBets posts and comments. As per Cioroianu *et al.* (2021), we initially expanded the Python package "pysentiment," which utilized the Harvard General Inquirer IV-4 dictionary and the Loughran and McDonald financial sentiment dictionary for sentiment analysis of the key posts and comments that had been identified. While working quite well with regard to sentiment analysis, it was identified that few these lexicons were representative of the crude and foul language commonly used by those posting in the r/WallStreetBets forum. To respond to this issue, we had to not only expand such well-known lexicons but also attempt to develop the predefined language that had already been identified as being representative of r/WallStreetBets users. To do so, we built upon the work of Rice and Zorn (2021) using a modified searching algorithm while expanding the lexicon generation method presented by Deng *et al.* (2017), who proposed a method for adapting the existing sentiment lexicons for domain-specific sentiment classification. The latter was developed based on 743,069 tweets, representing the usage of seeds and baselines. Upvotes and numbers of comments were used as a benchmark for assessing validity, that is, a representation of post attention. We expanded 69 previously identified pump and dumps within the r/WallStreetBets forum, where all posts are combined into a single file representative of the developing corpus. Using an adaptation of the work of Deng *et al.* (2017), we used candidate extraction and sentiment recognition via the above developing corpus and then identified the relationships between candidate and sentiment words to make a seed lexicon. We then added further posts from the r/WallStreetBets forum and attempted to add recognized words to the analysis while withdrawing words that were not in line with the seed lexicon. The entire database of posts was then stripped of punctuation, emojis and numbers. After a variety of methodological iterations, we then decided to remove the top 10% of the most frequently used terms in the sample and the bottom 10% of the most infrequently used terms from the combination of the search terms that came from the entire sample of r/WallStreetBets posts and further developed our selection based on the previously developed pump-and-dump lexicon. These terms are presented in Figure 2. We then reiterated the above methodology until the number of additional identified posts did not increase.[5]

Related stock market data are obtained from Thomson Reuters Eikon. All observations or comments made on either a Saturday or Sunday are denoted as being associated with the target company on the following Monday morning at market opening. The dataset incorporates 1,357 total announcements made during the selected time period and represents targeted actions upon 423 separate financial products. Descriptive statistics for these companies are presented in Table 1. These results are estimated using a dataset of 14,078,107 r/wallstreetbets comments made between January 2015 and February 2021, the growth of which is presented in Figure 3. Summary statistics on the number of occurrences for each company for each related event type are presented in Table 2 along with the number of occurrences by sentiment type. We observe that few of the identified pumps and dumps—indeed, few significant positive and negative sentiment events at all—occurred between 2015 and 2017. However, cases increased substantially between 2018 and 2021. When we focus on specific corporate statistics, GameStop is estimated to have experienced 45 separate pump and 10 separate dump events. Tesla experienced 27 pump attempts, while Advanced Micro Systems, Virgin Galactic and Nikola Systems experienced 27, 19, and 13 attempts, respectively. The same companies experienced a large number of dumps following GameStop. These companies are the most manipulated within the sample time frame. The

[5] After the first point at which the number of additional posts did not increase, it took four additional cycles before the number of identified posts found to be representative of pump-and-dump behavior began to decrease in line with the reduction in the lexicon. We must note the difficulties associated with coding such protocols, particularly where there exists minimal evidence of benchmarks on which to compare the results or to develop training classifiers. The methodology was also replicated with a variety of additional data included, such as the top 10,000 posts based on the largest number of upvotes or the top 10,000 posts based on the largest number of comments. However, the resulting identified posts based on pump-and-dump behavior remained within a $+/-5\%$ tolerance level of the original methodology, which was therefore identified to be the most suitable for the purposes of this research. In a similar manner as the development of positive and negative sentiment analyses, we built a co-occurrence matrix based on the posts found to be representative of behavior leading to pump-and-dump attempts. Within the identified posts, the top 300 positive and negative words were selected using cosine similarity, through which polarity was calculated. We then employed a zero cutoff point, where, quite simply, all posts that were scored above zero were found to represent pump attempts, while all those scored below zero were found to represent dumps. The remaining posts, which were found to not represent explicit attempts to pump or dump stocks, were found in some cases to be confused, that is, representative of explicit statements discussing terms representative of both pumps and dumps in the same post; in other words, this refers to buying the asset, increasing the price and then selling the asset in a timed manner.

Figure 2: Common and Infrequent Words Removed.

Note: After a variety of methodological iterations, we then removed the top 10% of the most frequently used terms in the sample and the bottom 10% of the most infrequently used terms from the combination of the search terms from the entire sample of r/WallStreetBets posts and further developed our selection based on the developed pump-and-dump lexicon. The top figure represents the common and most infrequent words removed during this process. The bottom figure represents the most common words used in both pump and dump searches.

largest identified organized pump of share prices occurred in January 2021, where substantial sentiment was identified on 11 January, when shares were valued at \$19.95. In the following days, momentum grew, with considerable evidence of coordinated behavior: on 27 January, shares were valued at \$347.51. Two further coordinated dumps and one further pump are identified in the following 2 weeks,[6] such that on 19 February, GameStop (GME) shares were valued at \$40.69.

All times are in GMT, with the official end-of-day closing price treated as the listed observation for each company in our analysis of the associated effects. Excess returns are defined as the daily log returns beyond the risk-free rate, proxied by the 1-month domestic sovereign bond rate of the country in which the domestic exchange of the target company is located. We measure market activity using the daily average traded volume. Market liquidity is proxied by the bid-ask spreads for each trading day. Bid and ask prices are calculated as

[6] It is important to note that Gamestop were not the only company to issue statements outlining significant financial abnormality after being the focus of r/wallstreetbets attention. The company, Hometown International (HWIN), owner of a single deli in outside of New Jersey, was reported in April 2021 to have generated sales of \$21,772 in 2019 and \$13,976 in 2020 respectively. Between January 2020 and April 2021, the company's share price increased from \$1.75 to \$13.75, where the company was estimated to possess a market capitalisation level above \$100 million. During the same period, the average daily liquidity in terms of value was less than \$5,000. Most trading had taken place OTC.

Table 1: Descriptive Statistics.

	Mean	St. Dev.	Skewness	Kurtosis	25th Perc.	Median	75th Perc.
$[-3, +1]$	0.0187	0.1551	3.2635	32.4899	−0.0346	0.0056	0.0474
$[-2, +1]$	0.0172	0.1510	3.6589	37.4176	−0.0346	0.0057	0.0426
$[-1, +2]$	0.0192	0.1560	3.6242	35.5646	−0.0327	0.0036	0.0433
$[-1, +3]$	0.0189	0.1623	3.1090	27.6607	−0.0367	0.0034	0.0451
$[-1, +1]$	0.0170	0.1492	3.7713	39.4051	−0.0302	0.0036	0.0413
$[-2, +2]$	0.0194	0.1580	3.4822	33.6468	−0.0370	0.0054	0.0452
$[-3, +3]$	0.0207	0.1669	2.7697	24.0178	−0.0384	0.0047	0.0531
$[0, +20]$	0.0085	0.1996	−0.0549	4.6994	−0.0610	0.0030	0.0737
$[0, +40]$	0.0324	0.2373	2.5048	25.9317	−0.0629	0.0193	0.1251
$[0, +60]$	0.0599	0.2719	1.2492	8.9075	−0.0592	0.0335	0.1936
Firm Size	20.7620	9.3784	−1.6588	1.0134	20.9587	25.0037	26.2105
Tobin's Q	0.3284	0.2909	0.2581	−0.5887	0.1024	0.2764	0.4962
Op Profit	0.3874	0.2921	0.2547	−0.9606	0.1146	0.3648	0.5991
Leverage	0.5399	0.3460	−0.1030	−0.7830	0.2505	0.5720	0.8345
Cash Holdings	0.2329	0.2354	1.2054	0.6729	0.0320	0.1860	0.3445
R&D Dummy	0.6019	0.5103	−0.3447	−1.9581	0.0000	1.0000	1.0000
Div. Dummy	0.4056	0.5052	0.4570	−1.8708	0.0000	0.0000	1.0000

Note: This table reports the descriptive statistics for the characteristics of target firms used in the main empirical analysis: the cross-sectional mean, standard deviation, and quartiles for the month in which each company received attention in the /wallstreetbets forum between January 2015 and February 2021.

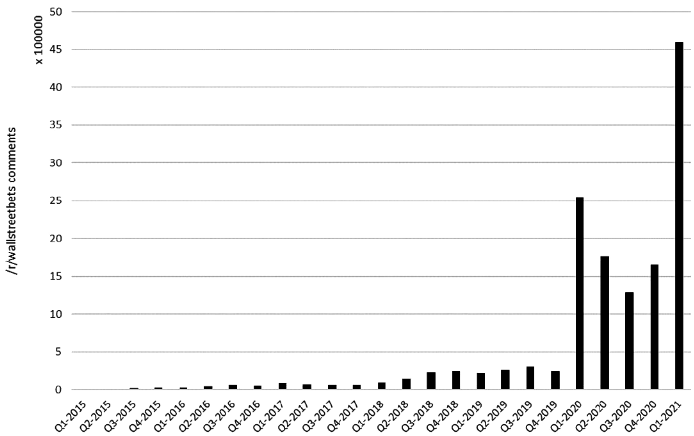


Figure 3: r/wallstreetbets Comment Growth.

Note: The above figure presents the growth in r/wallstreetbets comments on a quarterly basis between January 2015 and February 2021. Each of the above 14,078,107 comments is obtained through the specified methodologies.

Table 2: Corporate Descriptive Statistics.

Year	Pos. Sent.	Neg. Sent.	Pump	Dump	Total
2015	4	0	3	0	7
2016	1	1	9	0	11
2017	5	6	12	4	27
2018	84	67	42	11	204
2019	98	112	62	15	287
2020	111	149	260	78	598
2021	39	24	123	37	223
Total	342	359	511	145	1357
% of Total	25.2%	26.5%	37.7%	10.7%	

Note: This table reports the identified event within the r/wallstreetbets forum between January 2015 and February 2021.

the average bid and ask within a given day rescaled by the end-of-day stock price. We use several aggregate risk factors, such as market risk, size, value, momentum, investment opportunities and profitability, similarly to Fama and French (2015), to investigate the effect of investor attention in r/wallstreetbets on excess daily returns. We obtain factor-mimicking portfolios that proxy for these risk factors daily from Kenneth R. French's online library. We control for aggregate market behavior by using both the value-weighted daily market

Table 2: Descriptive Statistics (continued)

Companies	**Pump**	5-day return	Reddit rating
Gamestop (GME)	45	10.3%	5,043
Tesla (TSLA)	27	5.3%	6,768
Ad. Micro Syst (AMD)	19	9.1%	4,423
Virgin Galactic (SPCE)	17	7.9%	2,784
Nikola Corp (NKLA)	13	3.2%	2,908
Companies	**Dump**	5-day return	Reddit rating
Gamestop (GME)	10	−18.21%	3,274
Nikola Corp (NKLA)	8	−13.09%	5,237
Virgin Galactic (SPCE)	8	−7.32%	7,198
Ad. Micro Syst (AMD)	6	−15.60%	3,456
Moderna (MRNA)	3	−12.33%	529
Companies	Positive sentiment	Companies	Negative sentiment
Tesla (TSLA)	41	Tesla (TSLA)	13
Gamestop (GME)	13	Apple (AAPL)	11
Nikola Corp (NKLA)	13	Lyft Inc (LYFT)	9
Ad. Micro Syst (AMD)	12	Micron Tech (MU)	9
Virgin Galactic (SPCE)	11	Uber Tech (UBER)	9

Note: This table reports the identified corporate events by estimated sentiment within the r/wallstreetbets forum between January 2015 and February 2021. Both estimated returns and Reddit ratings reflect the average values in the 5-day period after each event.

returns presented in trillions of US dollars. Additionally, we collect a set of variables that capture firms' fundamentals.[7]

4 Methodology

4.1 Analyzing the Effects of r/wallstreetbets' behavior on Financial Markets

We begin our analysis by testing whether the identified cases in the r/wallstreetbets forum instigate abnormal share price responses. To this

[7] We collect data on firm size, calculated as the natural logarithm of total assets; Tobin's Q, calculated as the market value (year-end stock price multiplied by common shares outstanding) of the firm over total assets; operating profit; leverage; cash holdings; R&D, defined as a dummy variable that equals one if the firm invests in research and development and zero otherwise; and dividends, defined as a dummy variable that equals one if the firm pays dividends and zero otherwise.

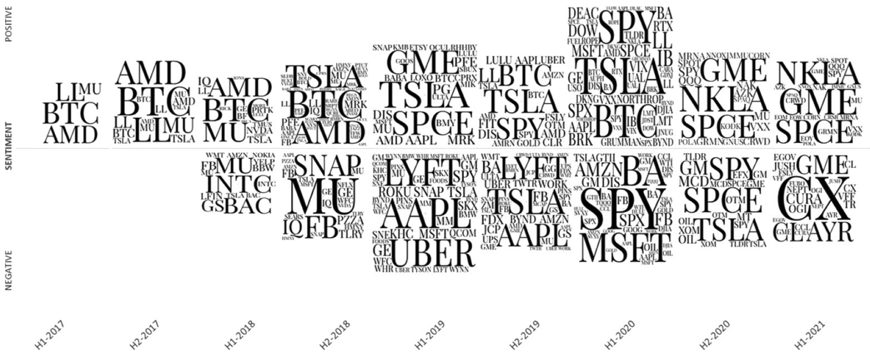


Figure 4: Time-varying Sentiment Identified on r/wallstreetbets by Company (Defined by Positive and Negative Sentiment).

Note: The presented word clouds represent the key companies identified as the subjects of significant positive and negative r/wallstreetbets sentiment. The text size represents the scale of the identified sentiment.

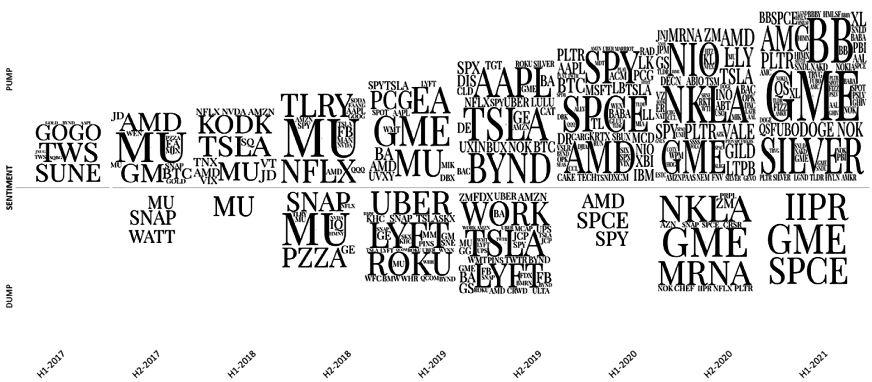


Figure 5: Time-varying Sentiment Identified on r/wallstreetbets by Company (Defined by Identified Pumps and Dumps).

Note: The presented word clouds represent the key companies identified as subjects of significant r/wallstreetbets pump-and-dump behavior. Text size represents the scale of the identified sentiment.

end, we analyze direct, short-term stock market responses around each of the algorithmically identified attempts to instigate crowd-driven pumping of specified stocks. The time-varying corporate results identified based on the sentiment classification search algorithms are presented in Figure 4, while corporate pumps and dumps explicitly identified as instigated by r/wallstreetbets are presented in Figure 5. These figures are scaled to clearly present the largest attempted price manipulations in each half-year.

To assess robustness, we measure abnormal returns using multiple estimation windows of one through 6 months surrounding each identified event. We then estimate abnormal returns using a variety of different event windows, including $[-3, +1]$, $[-2, +1]$, $[-1, +1]$, $[-2, +2]$, $[-3, +3]$, $[-1, +2]$, and $[-1, +3]$, to test the pricing response both before and after the dates on which the respective pumping behavior is identified. Each number refers to the specific trading days relative to each identified r/wallstreetbets-driven event. Note that for a variety of identified events as described, prospective traders state on the forums that they are seeking to open leveraged positions using call or put options of a variety of contract lengths. For these specifically identified cases, we further test the periods $[0, +20]$, $[0, +40]$ and $[0, +60]$ to reflect abnormal returns for the periods 1, 2 and 3 months after each identified event, reflecting the influence of the opening and closing of these options positions. The identified sentiment also reflects whether the proposed positions are of a positive or negative nature, that is, whether the forums attempt to pump or dump the stock under observation.

Based on the listed events, during the period under analysis, the expected returns are then estimated using the Fama and French three-factor model (Fama and French, 1993), the five-factor model (which adds the profitability and investment factors to the three-factor model) (Fama and French, 2015), and the subsequent six-factor model (which adds a momentum factor to the five-factor model) (Fama and French, 2018). We assume that the attempted pumping and dumping of stocks by the r/wallstreetbets forum is not in any way correlated with a firm's expected returns after we control for tradeable risk factors. The process through which abnormal returns (AR) and cumulative abnormal returns are generated can be found in Section B of the Online Appendix.

To understand the causal effect of investor attention through the r/wallstreetbets forum on daily abnormal returns and liquidity, we run a difference-in-difference analysis by estimating a set of panel regressions of the form:

$$y_{it} = \alpha + \gamma' D_{i\tau} \mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it} \quad \text{where} \quad t = \tau - T_N, \ldots, \tau + T_N \tag{1}$$

where τ identifies the event date, y_{it} represents the variable of interest for firm i at time t, C_i is a group dummy that takes a value of one if the company is affected by Reddit and zero otherwise, $D_{i\tau}$ is a $(k+1)$-dimensional vector of dummy variables that takes a value of one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} is a set of control variables, for example, mimicking risk factor portfolios, and μ_i represents the identified fixed firm effects. The null hypothesis that r/wallstreetbets investors' attention influences the outcome of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N.

The target company is compared to the control companies through the interaction with $D_{i\tau}$. To provide additional insight and robustness, we proceed to analyze not only the effects on pricing and liquidity but also the scale of investor attention, the year in which the algorithmically generated pump and dump attempt occurred and whether the attempt was explicit or implicit. Further robustness checks are presented based on the sector in which the company trades to analyze specific corporate vulnerabilities. Finally, several placebo tests yielding no significant results are presented to provide considerable structural and methodological robustness.

4.2 Did r/wallstreetbets Generate Deviations from Analysts' Expectations

In a secondary analysis, we explore whether the r/wallstreetbets forum influenced the differentials between the market price of the stock under identified manipulative pressures from the forum and the estimated fair market value identified by professional market analysts. As such market analysts closely monitor particular stocks and sectors, it is of interest to identify whether the added forces generated in online messaging forums were responsible for any additional mispricing, which in this section is identified as price differentials between targeted stocks and their respective I/B/E/S-based estimated prices.[8] Analyses at both pricing and volatility levels are conducted.

To do so, we employ an extended autoregressive (AR) model to examine the effects of different types of online discussions and comments around one

[8]Institutional Broker Estimated System (I/B/E/S) estimates are provided by Refinitiv and provide a record of data and the most complete global view of analyst forecasts on company performance. I/B/E/S estimates are found to provide deeper insight and analysis on a product/segment basis. Data are shown at the sector level, "screened with the most rigorous quality control methods, across 22,000 active companies in 90 countries, and sourced from over 18,000 analysts. These are expressed via 260+ measures, including generic measures, such as EPS, and industry-specific KPIs, such as oil production per day". I/B/E/S estimates provides analyst forecasts on company performance expressed via over 260 consensus measures, including earnings per share, sales, and net income. I/B/E/S guidance provides comments directly from corporate management about future company expectations. These forward-looking statements focus on sales or earnings expectations, allowing for the detailed valuation of a company's earnings potential and future outlook. Finally, I/B/E/S global aggregates provide bottom-up sector and industry earnings forecasts and related data. The I/B/E/S-based estimated prices for one stock are the mean price targets of the stock, which are the statistical average of all broker estimates determined on the majority accounting basis. The price target is the projected price level forecast by an analyst within a specific time horizon, and I/B/E/S-based estimated prices are updated on a daily basis. Henceforth, I/B/E/S-based estimates reflect the expectations of trading professionals/analysts on the fundamental values of a stock. The differences between the actual market price and I/B/E/S-based estimate of a stock mirror the deviations in the actual price of the stock in the market from its fair value. Based on this, we are able to examine whether Reddit commenting/betting activities contribute to pricing inefficiencies.

stock on the first-order differences of price differentials between the stock's daily actual prices and its I/B/E/S estimates. The model is as follows:

$$\mathrm{DPD}_t = a + \sum_{i=1}^{p} b_i \mathrm{DPD}_{t-i} + cD_t + e_t, \quad e_t \sim \mathrm{GED}(0, h_t, \lambda) \tag{2}$$

where DPD_t is the first-order difference of daily price differentials between one stock's market prices and its I/B/E/S estimates, that is, PD_t. Further, we have $\mathrm{PD}_t = P_t - \mathrm{IBES}_t$, where P_t is the natural logarithm of daily stock prices and IBES_t is the natural logarithm of daily I/B/E/S-based estimates. Unit root tests suggest that price differentials, PD_t, are not stationary for any sample under analysis. D_t is a dummy variable that takes value 1 when the period is 20 days after one specific commenting/betting activity takes place on Reddit and zero otherwise. D_t takes value 0 when the period is 20 days before one specific commenting/betting activity takes place on Reddit, examined specifically by the coefficient c. The lag order p is chosen according to the Akaike information criterion (AIC).

Furthermore, we employ an exponential generalized autoregressive conditional heteroscedasticity (EGARCH) model developed by Nelson (1991) to specify the conditional variance h_t of the innovations. The EGARCH model has the advantage of ensuring the positivity of estimated conditional variance without any parameter restrictions, in contrast to the alternative GARCH specifications. It also imposes fewer parameter restrictions to guarantee stationarity of the conditional variance. The EGARCH(1,1) model is specified as:

$$\log(h_t) = \omega + \alpha \left| \frac{e_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \frac{e_{t-1}}{\sqrt{h_{t-1}}} + \beta \log(h_{t-1}) + \phi D_t \tag{3}$$

where $\log(\cdot)$ is the natural logarithm. α examines the effect of recent shocks, while β measures persistence. To ensure that h_t is stationary, β should be less than one. γ examines the asymmetric response of conditional variance to recent negative shocks in comparison to that of recent positive shocks. D_t is the manipulative event dummy. ϕ examines whether one specific manipulative event activity on Reddit affects the conditional variance of changes in price differentials. The estimate of ϕ is of major interest, pointing to the effects of different kinds of manipulative event activities on the volatility of the changes in pricing differentials. The evidence sheds light on how those activities impact the variation dynamics of pricing inefficiencies.

5 Empirical Results

5.1 Effects of r/wallstreetbets Attention on Stock Returns

Figure 6 provides initial insight into the cumulative abnormal returns estimated for the total number of both the targeted sample and the control sample for

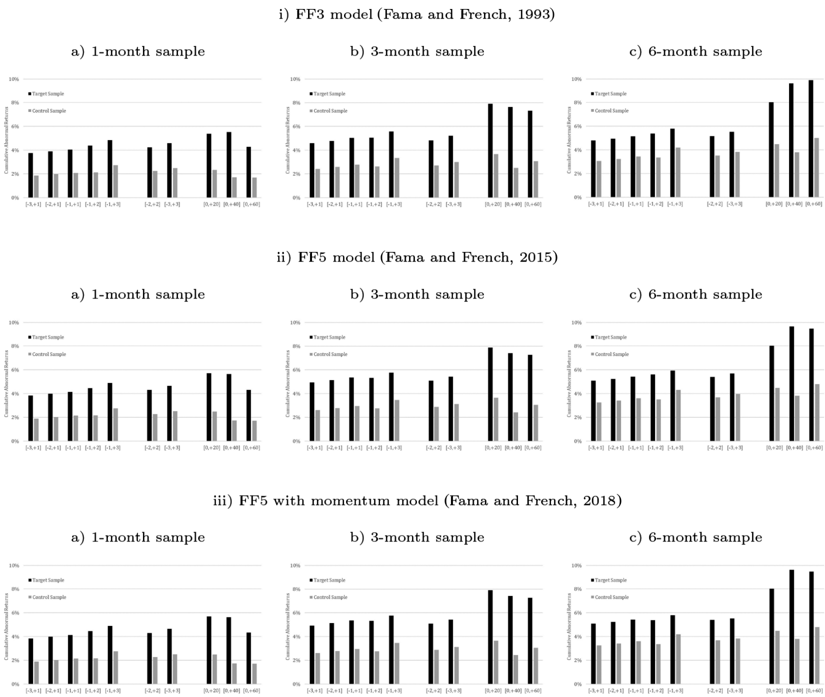


Figure 6: Cumulative Abnormal Excess Returns.

Note: The above figure shows the cumulative abnormal returns throughout the stated event windows for both firms that have been subjects of investor attention in the r/wallstreetbets forum in comparison to a sample of similar companies that are not. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018.

each model. Additional results outside of the presented samples and windows of analysis are available from the authors on request. Across each set of presented results, regardless of either the sample size or window analyzed, the target sample results are significantly above those of the control sample. There is a clear upward trend in the days following the algorithmically identified dates on which the r/wallstreetbets forum instigates focused investor attention on the target sample of companies. For the control sample, there does not appear to be a repeated trend across the windows, samples and models used. When we consider the effects presented for the periods 1, 3, and 6 months after each event, for the 1- and 3-month samples, there is evidence that CAR effects diminish over time; however, similar evidence of diminishing effects appears to be marginal in the 6-month sample. Such evidence suggests the existence of persistence of abnormal returns in the short term after manipulative targeting. In particular, evidence from this preliminary analysis suggests the existence of sustained abnormal returns of approximately +114 basis points in the 5 day window surrounding the event (2 days before through 2 days thereafter) in comparison to +16 basis points for the control group.

These preliminary results provide a useful precursor to those presented in Table 3. In a more thorough analysis, we present the CAR estimation results for the target companies in the top panel and those of the control sample in the lower panel. Specifically, using a difference-in-difference regression, we test the null hypothesis of whether daily cumulative abnormal excess returns are significantly different from zero for a variety of window sizes, models and sample sizes, conditioning on several previously stated risk factors. Consistent with the results provided in the preliminary analysis, the evidence suggests that significant effects are generated for the r/wallstreetbets target sample of companies. However, when we control for company-specific and external risk factors, the results are far more pronounced than those initially presented. For the period consisting of just 1 day before through 1 day after the algorithmically identified date, the results range from +404 through +542 basis points of cumulative abnormal returns generated. The scale of these results is repeated throughout the variety of windows and models presented, indicative of significant and exceptionally pronounced effects. Further, we must note that when we focus on the 1-, 3-, and 6-month periods thereafter, cumulative abnormal returns remain pronounced and significant at the 1% level, ranging between +538 and +804 basis points 1 month later, between +553 and +966 basis points 3 months later, and finally, between +428 and +748 basis points 6 months later. The lower panel of Table 3 presents a difference-in-difference analysis that provides very clear evidence that the selected control sample presents no evidence of significant effects generated by the attention-generating events identified in the r/wallstreetbets forums.

We now focus on the cross-sectional determinants of the market reactions reported in Table 4, which are examined through regressions based on the factors influencing a firm's CARs within the short-term period surrounding the identified event dates. In this context, the null hypothesis is that the CARs can be explained by firm-specific characteristics rather than by the attention from r/wallstreetbets investors. In columns (2) to (4), we test whether the stock market reaction is related to firms' attributes conditional on market risk, size, and value. Consistent with our findings above, we find that a positive market reaction is associated with r/wallstreetbets investor attention generated on the dates detected by the structured search algorithm for a target firm regardless of its economic fundamentals.

In columns (5) through (7), we show the results when we include momentum in the set of risk factors used to calculate abnormal returns. The largest effects (significant at the 1% level) are generated by the Fama and French (1993), Fama and French (2017), and Fama and French (2018) models for the research and development (R&D) dummy variable. This result indicates that companies with R&D capacity are the most likely to generate substantial CARs,

Table 3: Significance of Cumulative Abnormal Returns after r/wallstreetbets Investor Attention.

		Event windows for firms subjected to r/wallstreetbets attention									
Model	Windows	$[-3, +1]$	$[-2, +1]$	$[-1, +1]$	$[-2, +2]$	$[-3, +3]$	$[-1, +2]$	$[-1, +3]$	$[0, +20]$	$[0, +40]$	$[0, +60]$
FF3	1 month	0.0375***	0.0390***	0.0404***	0.0423***	0.0458***	0.0439***	0.0484***	0.0538***	0.0553***	0.0428***
		(0.0045)	(0.0046)	(0.0046)	(0.0043)	(0.0040)	(0.0044)	(0.0041)	(0.0072)	(0.0023)	(0.0021)
	3 months	0.0459***	0.0479***	0.0504***	0.0481***	0.0521***	0.0506***	0.0557***	0.0791***	0.0764***	0.0732***
		(0.0054)	(0.0055)	(0.0056)	(0.0052)	(0.0048)	(0.0053)	(0.0049)	(0.0032)	(0.0020)	(0.0017)
	6 months	0.0480***	0.0495***	0.0515***	0.0517***	0.0552***	0.0538***	0.0579***	0.0804***	0.0961***	0.0788***
		(0.0062)	(0.0064)	(0.0064)	(0.0060)	(0.0056)	(0.0061)	(0.0058)	(0.0040)	(0.0022)	(0.0038)
FF5	1 month	0.0384***	0.0399***	0.0414***	0.0431***	0.0465***	0.0447***	0.0490***	0.0571***	0.0564***	0.0432***
		(0.0045)	(0.0047)	(0.0047)	(0.0044)	(0.0040)	(0.0044)	(0.0041)	(0.0048)	(0.0023)	(0.0021)
	3 months	0.0494***	0.0513***	0.0536***	0.0509***	0.0543***	0.0533***	0.0576***	0.0789***	0.0741***	0.0727***
		(0.0053)	(0.0055)	(0.0055)	(0.0052)	(0.0048)	(0.0052)	(0.0049)	(0.0032)	(0.0022)	(0.0017)
	6 months	0.0509***	0.0523***	0.0542***	0.0541***	0.0570***	0.0560***	0.0595***	0.0804***	0.0966***	0.0747***
		(0.0062)	(0.0064)	(0.0064)	(0.0060)	(0.0056)	(0.0061)	(0.0057)	(0.0040)	(0.0023)	(0.0041)
FF5 w/M	1 month	0.0382***	0.0398***	0.0413***	0.0430***	0.0464***	0.0446***	0.0490***	0.0570***	0.0563***	0.0433***
		(0.0045)	(0.0047)	(0.0047)	(0.0044)	(0.0040)	(0.0044)	(0.0041)	(0.0049)	(0.0023)	(0.0021)
	3 months	0.0493***	0.0512***	0.0536***	0.0509***	0.0543***	0.0532***	0.0576***	0.0790***	0.0743***	0.0727***
		(0.0053)	(0.0055)	(0.0055)	(0.0052)	(0.0048)	(0.0052)	(0.0049)	(0.0032)	(0.0021)	(0.0017)
	6 months	0.0510***	0.0523***	0.0542***	0.0541***	0.0552***	0.0538***	0.0579***	0.0804***	0.0961***	0.0748***
		(0.0062)	(0.0064)	(0.0064)	(0.0060)	(0.0056)	(0.0061)	(0.0058)	(0.0040)	(0.0022)	(0.0042)

Note: This table reports the results of the significance test of cumulative abnormal returns realized within a variety of event windows. Daily abnormal returns represent the return realized by an investor in excess of sources of systematic risks. The table reports the results for different risk factor models and for different estimation periods (from 1 to 6 months surrounding the algorithmically identified manipulation event). The top panel presents companies identified as target companies, while the middle panel presents the results of a similar analysis conducted on a sample of similar companies during the same event windows. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015 and Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Continued.

		Event windows of Control Sample									
Model	Windows	[−3, +1]	[−2, +1]	[−1, +1]	[−2, +2]	[−3, +3]	[−1, +2]	[−1, +3]	[0, +20]	[0, +40]	[0, +60]
FF3	1 month	0.0185	0.0197	0.0209	0.0224	0.0247	0.0212	0.0272	0.0233	0.0169	0.0168
		(0.0217)	(0.0227)	(0.0234)	(0.0224)	(0.0211)	(0.0207)	(0.0225)	(0.0306)	(0.0068)	(0.0082)
	3 months	0.0242	0.0259	0.0278	0.0272	0.0300	0.0262	0.0334	0.0366	0.0250	0.0307
		(0.0207)	(0.0218)	(0.0224)	(0.0216)	(0.0204)	(0.0200)	(0.0217)	(0.0109)	(0.0048)	(0.0052)
	6 months	0.0306	0.0322	0.0343	0.0352	0.0383	0.0336	0.0420	0.0449	0.0379	0.0500
		(0.0295)	(0.0310)	(0.0320)	(0.0307)	(0.0291)	(0.0284)	(0.0311)	(0.0166)	(0.0066)	(0.0143)
FF5	1 month	0.0190	0.0202	0.0214	0.0228	0.0251	0.0217	0.0276	0.0248	0.0173	0.0169
		(0.0220)	(0.0231)	(0.0238)	(0.0226)	(0.0213)	(0.0209)	(0.0227)	(0.0204)	(0.0068)	(0.0082)
	3 months	0.0261	0.0277	0.0296	0.0288	0.0313	0.0276	0.0346	0.0366	0.0242	0.0305
		(0.0205)	(0.0215)	(0.0222)	(0.0213)	(0.0201)	(0.0197)	(0.0214)	(0.0107)	(0.0053)	(0.0052)
	6 months	0.0324	0.0341	0.0361	0.0368	0.0396	0.0350	0.0431	0.0449	0.0381	0.0479
		(0.0294)	(0.0309)	(0.0319)	(0.0305)	(0.0289)	(0.0283)	(0.0309)	(0.0165)	(0.0067)	(0.0155)
FF5 w/M	1 month	0.0189	0.0201	0.0213	0.0227	0.0250	0.0216	0.0275	0.0247	0.0172	0.0170
		(0.0220)	(0.0231)	(0.0238)	(0.0226)	(0.0213)	(0.0210)	(0.0227)	(0.0208)	(0.0068)	(0.0082)
	3 months	0.0261	0.0277	0.0296	0.0287	0.0313	0.0276	0.0346	0.0366	0.0243	0.0305
		(0.0205)	(0.0216)	(0.0222)	(0.0214)	(0.0201)	(0.0197)	(0.0214)	(0.0108)	(0.0050)	(0.0052)
	6 months	0.0325	0.0341	0.0361	0.0368	0.0383	0.0336	0.0420	0.0449	0.0379	0.0480
		(0.0294)	(0.0310)	(0.0319)	(0.0305)	(0.0291)	(0.0284)	(0.0311)	(0.0166)	(0.0066)	(0.0158)

Table 4: Determinants of Cumulative Abnormal Returns after r/wallstreetbets Forum Attention.

Model	Window	Target	Firm size	Leverage	Cash Hold.	R&D	Dividend	Oper. Prof.	Tobin's Q	Obs	R^2
FF3	$[-1,+1]$	0.0421***	−0.0010	0.0071	−0.0107	0.0467***	0.0175	0.0672	−0.0029	1,357	0.1119
		(0.0048)	(0.0037)	(0.1004)	(0.0641)	(0.0068)	(0.0180)	(0.0540)	(0.0996)		
	$[-1,+2]$	0.0453***	−0.0009	0.0035	−0.0158	0.0498***	0.0178	0.0626	−0.0021	1,357	0.1381
		(0.0045)	(0.0036)	(0.0990)	(0.0632)	(0.0066)	(0.0178)	(0.0532)	(0.0982)		
	$[-1,+3]$	0.0499***	−0.0004	−0.0121	−0.0064	0.0449***	0.0179	0.0573	−0.0136	1,357	0.1762
		(0.0042)	(0.0036)	(0.0968)	(0.0617)	(0.0062)	(0.0174)	(0.0521)	(0.0960)		
FF5	$[-1,+1]$	0.0544***	−0.0014	−0.0210	−0.0510	0.0584***	0.0072	−0.0040	−0.0681	1,357	0.1398
		(0.0056)	(0.0043)	(0.1176)	(0.0751)	(0.0200)	(0.0211)	(0.0632)	(0.1167)		
	$[-1,+2]$	0.0552***	−0.0016	−0.0187	−0.0523	0.0578***	0.0074	−0.0089	−0.0605	1,357	0.1490
		(0.0053)	(0.0043)	(0.1170)	(0.0747)	(0.0198)	(0.0210)	(0.0629)	(0.1159)		
	$[-1,+3]$	0.0594***	−0.0010	−0.0358	−0.0404	0.0535***	0.0075	−0.0150	−0.0728	1,357	0.1828
		(0.0050)	(0.0042)	(0.1147)	(0.0731)	(0.0194)	(0.0205)	(0.0616)	(0.1138)		
FF5 w/Mom.	$[-1,+1]$	0.0530***	0.0008	−0.0588	−0.0434	0.0571***	−0.0173	0.0317	−0.1222	1,357	0.1040
		(0.0066)	(0.0050)	(0.1375)	(0.0877)	(0.0234)	(0.0246)	(0.0738)	(0.1364)		
	$[-1,+2]$	0.0573***	0.0008	−0.0599	−0.0472	0.0431***	−0.0170	0.0263	−0.1176	1,357	0.1189
		(0.0063)	(0.0050)	(0.1363)	(0.0870)	(0.0231)	(0.0244)	(0.0732)	(0.1352)		
	$[-1,+3]$	0.0612***	0.0014	−0.0761	−0.0338	0.0532***	−0.0169	0.0201	−0.1291	1,357	0.1429
		(0.0059)	(0.0049)	(0.1345)	(0.0858)	(0.0228)	(0.0241)	(0.0723)	(0.1334)		

Note: The table reports the results for cumulative abnormal returns calculated over several event windows and conditional on different risk factors. We run the following regression analyses $\mathrm{CAR}_t = \alpha + \beta_0 \mathrm{C}_i + \beta' z_i + \epsilon_i$, where CAR_t represents the cumulative abnormal returns for the target firm i over a given event window, z_i represents a vector of firm characteristics calculated over the year prior to the event, and C_t is a dummy variable that takes value one for target firms and zero otherwise. Robust standard errors are reported in parentheses. The panels present the results of models based on the work of Fama and French, 1993, Fama and French, 2015 and Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

particularly the technology development companies most actively sought by r/wallstreetbets users, mirroring the stock preferences of similar messaging board users during the dotcom crisis of the late 1990s and early 2000s (Das and Chen, 2007; Sabherwal *et al.*, 2008). Finally, columns (8) and (10) confirm that the negative effect of data breaches cannot be explained by firm characteristics when abnormal returns are calculated conditional on a Fama and French five-factor specification.

5.2 Are there Effects on Trading Activity and Market Liquidity?

We next separate the causal effects of r/wallstreetbets investor attention, estimating Equation (1) by conditioning returns on a set of risk factors inclusive of the set z_{it}. The results of this analysis are presented in Table 5 with regard to excess returns and Table 6 with regard to market liquidity in the form of trading volumes. The estimates for each associated γ' are presented and specifically analyze the response of y_{it} to investor attention on a certain date for the firm targeted by manipulative forces, relative to the outcomes of a sample of firms with a similar size and geographical location. With both models, we can observe the presence of both significant and pronounced effects upon abnormal returns and liquidity on the day that the effects of investor attention are realized. The effects of abnormal returns across each model not only persist but also increase substantially on each of the days thereafter. Similar results appear when we focus on liquidity effects; however, while the results remain substantially elevated 2 days after the events, liquidity levels do not continue to increase in the same way that abnormal returns do. This result presents evidence of a brief and particularly acute shock to corporate trading conditions for each company exposed to manipulative investor attention in online forums.

In Table 7, we analyze the differential abnormal returns of the sample of target companies and the control sample. However, in this analysis, we separate the observations based on the year that they occur. It is interesting to note that for both abnormal returns and trading volumes, significant effects are observable throughout 2019 and February 2021, indicating that the interactions have been increasing substantially over time. There is evidence to suggest that for the target sample, liquidity effects were observable in 2018, but no significant effects were observed in terms of abnormal returns. Across both analyses, no evidence of a differential significantly greater than zero is evident on the day before the identified manipulation date. The scale of the differentials suggests that the strength of their influence has increased quite dramatically in the period from 2019 through 2021, particularly in terms of market liquidity on the dates on which such behaviors are identified to have occurred. Such results indicate that while attempts at manipulation occurred throughout the period from 2016 through 2018, their effects did not appear to generate significant

Table 5: Excess Returns after r/wallstreetbets Forum Attention.

Model	Window	$D_{\tau+2}.C_i$	$D_{\tau+1}.C_i$	$D_{\tau}.C_i$	$D_{\tau-1}.C_i$	$D_{\tau-2}.C_i$	R^2
FF3 (Fama and French, 1993)	$[-1, +1]$		-0.0146	0.0568^{***}	0.1261^{***}		0.1149
			(0.0327)	(0.0206)	(0.0265)		
	$[-2, +2]$	-0.0404	-0.0465	0.0598^{***}	0.0768^{***}	0.1636^{***}	0.1568
		(0.0286)	(0.0426)	(0.0201)	(0.0214)	(0.0258)	
FF5 (Fama and French, 2017)	$[-1, +1]$		-0.0164	0.0537^{***}	0.1258^{***}		0.1160
			(0.0329)	(0.0208)	(0.0266)		
	$[-2, +2]$	-0.0425	-0.0515	0.0556^{***}	0.0785^{***}	0.1647^{***}	0.1581
		(0.0289)	(0.0432)	(0.0204)	(0.0215)	(0.0259)	
FF5 w/M (Fama and French, 2018)	$[-1, +1]$		-0.0110	0.0508^{***}	0.1202^{***}		0.1339
			(0.0327)	(0.0205)	(0.0264)		
	$[-2, +2]$	-0.0462	-0.0505	0.0576^{***}	0.0754^{***}	0.1583^{***}	0.1730
		(0.0287)	(0.0428)	(0.0402)	(0.0213)	(0.0258)	

Note: To understand the causal effect of investor attention in the /wallstreetbets forum on excess returns, we estimate a set of panel regressions of the form $y_{it} = \alpha + \gamma' D_{i\tau} \mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it}$ where $t = \tau - T_N, \ldots, \tau + T_N$, with τ identifying the event date, y_{it} representing the variable of interest for firm i at time t, $D_{i\tau}$ being a (k+1)-dimensional vector of dummy variables that takes value one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} being a set of control variables, for example, mimicking risk factor portfolios, and μ_i representing the identified firm fixed effects. The null hypothesis that attention from r/wallstreetbets investors influences the outcome of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N. The target company is compared to the control companies through the interaction with $D_{i\tau}$. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Trading Volume effects after r/wallstreetbets Forum Attention.

Model	Window	$D_{\tau+2}.C_i$	$D_{\tau+1}.C_i$	$D_{\tau}.C_i$	$D_{\tau-1}.C_i$	$D_{\tau-2}.C_i$	R^2
Trading Vol.	$[-1,+1]$		0.2608	0.9023***	2.9980***		0.0784
			(0.3296)	(0.0351)	(0.0002)		
	$[-2,+2]$	0.0684	0.2327	0.8757***	2.4081***	0.8078***	0.0901
		(0.1125)	(0.3370)	(0.0348)	(0.0013)	(0.1374)	

Note: To understand the causal effect of investor attention in the r/wallstreetbets forum on market liquidity, we estimate a set of panel regressions of the form $y_{it} = \alpha + \gamma' D_{i\tau}\mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it}$ where $t = \tau - T_N, \ldots, \tau + T_N$, with τ identifying the event date, y_{it} representing the variable of interest for firm i at time t, $D_{i\tau}$ being a (k+1)-dimensional vector of dummy variables that takes a value of one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} being a set of control variables, for example, mimicking risk factor portfolios, and μ_i representing the identified firm fixed effects. The null hypothesis that attention from r/wallstreetbets investors influences the outcome of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N. The target company is compared to the control companies through the interaction with $D_{i\tau}$. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Evidence of a Change in Effects over Time.

	Abnormal returns					
Window	2016	2017	2018	2019	2020	2021#
$D_{\tau-1}.C_i$	0.0031	−0.0267	−0.0045	0.0264	0.0386	0.0515
	(0.0660)	(0.0188)	(0.0073)	(0.0393)	(0.0600)	(0.0707)
$D_{\tau}.C_i$	0.0556	0.0470	0.0182	0.0588***	0.1619**	0.2179***
	(0.0992)	(0.0354)	(0.0113)	(0.0059)	(0.0744)	(0.0521)
$D_{\tau+1}.C_i$	−0.0268	−0.0145	0.0105	0.0330***	0.1572***	0.2769***
	(0.0854)	(0.0244)	(0.0103)	(0.0045)	(0.0379)	(0.0277)
R^2	0.0828	0.1377	0.1836	0.2487	0.1703	0.2159
	Trading volumes					
Window	2016	2017	2018	2019	2020	2021#
$D_{\tau-1}.C_i$	0.0536	0.0381	0.0654	0.1363	0.0147	0.0301
	(0.0906)	(0.0577)	(0.0900)	(0.0931)	(0.1023)	(0.0946)
$D_{\tau}.C_i$	0.2522	0.1266	0.6190***	0.8733***	4.9232***	4.3002***
	(0.1596)	(0.1241)	(0.0738)	(0.0955)	(0.1296)	(0.0904)
$D_{\tau+1}.C_i$	0.1497	0.0163	0.4510***	0.9997***	0.8208***	0.8789***
	(0.0952)	(0.0616)	(0.1144)	(0.0001)	(0.1028)	(0.0944)
R^2	0.1451	0.1713	0.2588	0.2170	0.3107	0.3287

Note: To understand the causal effect of investor attention in the r/wallstreetbets forum on daily abnormal returns, we estimate, separated by the year in which the algorithmically identified event occurs, a set of panel regressions of the form $y_{it} = \alpha + \gamma' D_{i\tau}\mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it}$ where $t = \tau - T_N, \ldots, \tau + T_N$, with τ identifying the event date, y_{it} representing the variable of interest of firm i at time t, $D_{i\tau}$ being a (k+1)-dimensional vector of dummy variables that takes a value of one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} being a set of control variables, for example, mimicking risk factor portfolios, and μ_i representing the identified firm fixed effects. The null hypothesis that attention from r/wallstreetbets investors influences the outcome of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N. The target company is compared to the control companies through the interaction with $D_{i\tau}$. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. # indicates that data were only available up to February for the year 2021.

influence until the period from 2018 through early 2021. In line with the growth of the user numbers observed in Figure 1, this result would indicate that such influence and side effects on targeted corporations are correlated with the number of registered forum users and the influence that they can generate as a group. In Table 8, we again observe no significantly different results on the day before an identified event; however, when we focus on the scale of investor attention as measured by the Reddit rating of the threads in which manipulative statements and attempts to generate pumps and dumps are made, there is clear evidence to suggest that the larger the subsequent rating, the larger is the scale of the positive influence on abnormal returns.

Table 8: Effect of the Amount of Investor Attention on Excess Returns.

	Reddit rating			
Window	<1K	1.0k<x<5k	5.0k<x<10k	>10k
$D_{\tau-1}.C_i$	0.0042	0.0037	0.0011	0.0273
	(0.0160)	(0.0041)	(0.0005)	(0.0826)
$D_{\tau}.C_i$	0.0040	0.0117**	0.0587***	0.1461***
	(0.0187)	(0.0056)	(0.0072)	(0.0201)
$D_{\tau+1}.C_i$	0.0045	0.0137***	0.0410***	0.1702***
	(0.0096)	(0.0039)	(0.0046)	(0.0267)
R^2	0.1717	0.1386	0.2319	0.2012

Note: To understand the causal effect of investor attention in the r/wallstreetbets forum on daily abnormal returns, we analyze the associated Reddit rating of each comment identified as inciting a pump or dump. (A key feature of Reddit is that users can cast positive or negative votes, called upvotes and downvotes, respectively, for each post and comment on the site. The number of upvotes or downvotes determines posts' visibility on the site, so the most popular content is displayed to the most people. Users can also earn "karma" for their posts and comments, a status that reflects their standing within the community and their contributions to Reddit.) We estimate a set of panel regressions of the form $y_{it} = \alpha + \gamma' D_{i\tau}\mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it}$ where $t = \tau - T_N, \ldots, \tau + T_N$, with τ identifying the event date, y_{it} representing the variable of interest for firm i at time t, $D_{i\tau}$ being a (k+1)-dimensional vector of dummy variables that takes a value of one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} being a set of control variables, for example, mimicking risk factor portfolios, and μ_i representing the identified firm fixed effects. The null hypothesis that attention from r/wallstreetbets investors influences the outcome of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N. The target company is compared to the control companies through the interaction with $D_{i\tau}$. The panels present the results of models based on the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018.***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

5.3 Different Results Based on the Presence of Explicit or Implicit Commentary

We further investigate whether there are differences in responses based on the type of financial market product that has been explicitly targeted. For this analysis, we maintain the structure in a similar manner to that used before; however, we expand the lexicon of terms not only to focus on terminology indicative of proposed actions to "pump" or to "dump" but also to include terms such as "gold" or "silver," which would indicate the targeting of commodity markets. Alternatively, terms such as "Bitcoin" or "crypto" would indicate the targeting of cryptocurrencies. Table 9 reports such results, indicating that associated futures products, cryptocurrencies, commodities and volatility indices (such as the CBOE Volatility Index, or VIX) present positive, though insignificant, estimates, and therefore, no r/WallStreetBets influence can be identified. However, for options and stocks, there is significant evidence of abnormal returns on the day of and day after stated manipulative attempts.

Table 9: Does the Amount of Investor Attention Influence Other Types of Financial Products?

Window	Futures	Crypto	Vol. Indices	Options	Commodities	Stocks
$D_{\tau-1}.C_i$	0.0038	0.0267	0.0232	0.0091	−0.0379	0.0032
	(0.0070)	(0.0156)	(0.0135)	(0.0046)	(0.0083)	(0.0028)
$D_{\tau}.C_i$	0.0021	0.0512	0.0675	0.0376***	0.0423	0.0138**
	(0.0062)	(0.0569)	(0.0748)	(0.0021)	(0.0147)	(0.0037)
$D_{\tau+1}.C_i$	0.0023	0.0405	0.0351	0.0172***	−0.0067	0.0111***
	(0.0044)	(0.0535)	(0.0466)	(0.0052)	(0.0145)	(0.0023)
R^2	0.0470	0.1301	0.0417	0.1044	0.0828	0.1393

Note: To understand the causal effect of investor attention through the r/WallStreetBets forum on daily abnormal returns, we estimate, as separated by the type of financial market product in which the algorithmically identified event is specified, a set of panel regressions of the form $y_{it} = \alpha + \gamma' D_{i\tau}\mathrm{C}_i + \beta' z_{it} + \mu_t + \epsilon_{it}$ where $t = \tau - T_N, \ldots, \tau + T_N$ in which τ identifies the event date, y_{it} represents the variable of interest firm i at time t, $D_{i\tau}$ is a (k+1)-dimensional vector of dummy variables that takes a value of one in the interval $[\tau - k; \tau + k]$ and zero otherwise, z_{it} is a set of control variables, for example, mimicking risk factor portfolios, and μ_i represents the identified fixed firm effects. The null hypothesis that the attention of r/WallStreetBets investors influences the amount of interest is tested based on the regression coefficient γ', which represents the reaction of y_{it} over the event window T_N. The target company is compared to the control companies through its interaction with $D_{i\tau}$. The panels are presented based on the use of methodological structures following the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Does the Type of r/WallStreetBets Announcement Matter?

	Explicit		Implicit	
Window	Pump	Dump	Pos. Sent.	Neg. Sent.
$D_{\tau-1}.C_i$	−0.0015	−0.0265	0.0088	−0.0171
	(0.0462)	(0.0413)	(0.0068)	(0.0065)
$D_{\tau}.C_i$	0.0721***	−0.0525***	0.0208**	−0.0124
	(0.0156)	(0.0063)	(0.0082)	(0.0062)
$D_{\tau+1}.C_i$	0.1375***	−0.0102***	0.0183***	0.0012
	(0.0316)	(0.0015)	(0.0046)	(0.0065)
R^2	0.1614	0.0710	0.2037	0.0559

Note: In the above table, we separate the previous methodologies based on whether each is determined to be an explicit or implicit statement to "pump" or "dump" each algorithmically identified event. Further, we investigate whether each event varies in terms of whether the product used was the spot stock price or the option. The panels are presented based on the use of methodological structures following the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Further, in Table 10, we investigate whether there are significant differences based on whether the r/WallStreetBets comments generate significant effects depending on whether they are deemed explicit (i.e., directly using manipulative terms) or implicit (identified based on frequency of lexicon usage). We immediately identify that those events that propose the use of techniques to "pump" financial market products have significantly positive and substantial effects on both the day of and day after such statements. While for explicit dumps, significant negative returns are identified on the day of and day after such statements, and the scale of such returns are not the same as those for "pumps" and are found to diminish quite rapidly. For implicit manipulative statements, only those found to incorporate broad positive sentiments have significant positive effects, while negative sentiment is found to generate no significant effects whatsoever.

5.4 Does Attention from r/wallstreetbets Investors Generate Deviations from Analysts' Expectations?

Considering that the identified manipulative events generate significant effects on both abnormal returns and liquidity, the next stage of our analysis investigates whether such events cause errors in analyst recommendations and share price estimations. To complete this analysis, we utilize an EGARCH(1,1) model that is robust across a variety of pre-estimation procedures and estimate the level of volatility and long-term error that can be attributed to the month after each algorithmically identified r/wallstreetbets manipulation event. For brevity, only an overview of these EGARCH(1,1) results is presented in this research. A full list of estimated coefficients is available from the authors upon request.

The results of this analysis are presented in Table 11, with a visual representation of the calculated coefficients in several box plots in Figure 7. Considering the significant results in isolation, positive results here indicate overpricing of each analyzed corporate event by professional sectoral analysts, while negative results denote underpricing. Put simply, the figure indicates when the stated I/B/E/S outcome deviated significantly from the actual share price response to each dummy event. Significantly positive results are approximately twice as prevalent as significantly negative results (179 and 82 events, respectively). However, when we separate these results further by the type of event that occurred, several significant results appear for the type of derivative advertised in each manipulative statement, the response of Reddit users as indicated by the respective score of each post, and finally, changes in misestimation over time.

In Table 11, while we observe substantive differentials in the summary statistics for each EGARCH(1,1) approach, the proportion of significant outcomes presents interesting results. For manipulative events relating to positive sentiment and pump events, the proportion of significantly positive outcomes is more than 81%. When we consider negative sentiment and dump events, the number of significant events is more than 64%. This indicates that when manipulative events occur, approximately half generate significant deviations from analyst expectations, indicating that industry experts do not account for such manipulative behavior in their estimation methodologies. Over time, such evidence suggests that the mean returns generated in periods after /wallstreetbets attention have increased. For companies that have experienced repeated focus of attention, there is evidence to suggest that such effects have amplified over time, in line with the growth of the number of forum users, and potential traders of such information that has been made available. Further, such results indicate a significant deviation from fundamental expectations and values. Concerning the type of financial product used in each identified manipulative event, significant results are more pronounced when stock options are involved in betting/commenting activities than when stocks themselves are used. Mispricing is more likely to occur, therefore, when the manipulation involves a derivative. If we consider both reach and effects over time, the results are even more pronounced.

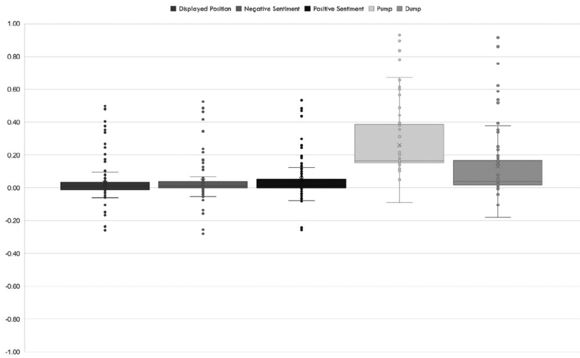
Such results are further supported in the box plots presented in Figure 7. When we consider estimated mispricing, significant average deviations are found for all explicitly and implicitly manipulative events relating to pumps and dumps. Further, similarly elevated results are identified for manipulative events that generated more substantiative reach in terms of the size of the respective audience and Reddit rating. Such results further support those

Table 11: Test for Evidence of r/wallstreetbets-induced Financial Market Volatility.

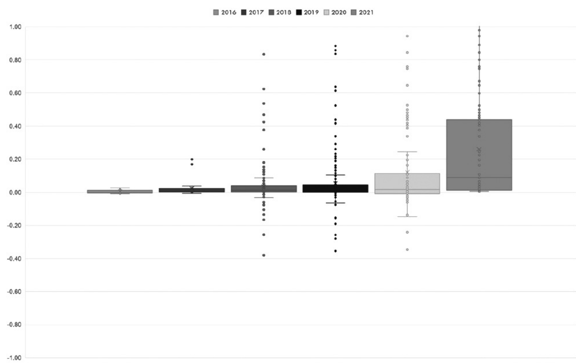
Group	Pos.	Sig. Pos.	%	Neg.	Sig. Neg.	%	Obs.	Mean	Std.	Skew.	Kurt.	Max.	Min.
Type of algorithmically identified event analyzed													
Negative sentiment	98	31	31.6%	261	162	62.1%	359	0.012	0.126	1.387	4.909	0.526	−0.280
Positive sentiment	197	130	66.0%	145	16	11.0%	342	0.055	0.123	2.405	5.502	0.533	−0.076
Pump	351	287	81.8%	160	6	3.8%	511	0.180	0.188	1.844	3.755	0.931	−0.089
Dump	34	6	17.6%	111	72	64.9%	145	0.131	0.204	2.279	5.466	0.917	−0.180
Year in which the manipulative event took place													
2016	10	6	60.0%	1	0	0.0%	11	0.004	0.012	1.080	0.126	0.027	−0.009
2017	17	9	52.9%	10	7	70.0%	27	0.020	0.045	3.410	11.371	0.199	−0.009
2018	126	71	56.3%	78	41	52.6%	204	0.045	0.152	1.724	7.402	0.834	−0.381
2019	160	91	56.9%	127	60	47.2%	287	0.053	0.172	2.807	10.377	0.884	−0.356
2020	371	210	56.6%	227	111	48.9%	598	0.128	0.250	1.613	1.876	0.943	−0.346
2021	162	87	53.7%	61	37	60.7%	223	0.124	0.246	1.650	2.085	0.958	−0.349

Note: The above estimates present the summary statistics of the EGARCH(1,1) approach used to analyze the effects of the identified manipulative events on the market volatility of the companies identified as targets of the r/wallstreetbets forums. Further specific coefficients of the the EGARCH(1,1) results, along with a range of pre-estimation and post-estimation results, are, for brevity, omitted from this final version and are available from the authors upon request. The EGARCH(1,1) model is specified as: $\log(h_t) = \omega + \alpha \left| \frac{e_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \frac{e_{t-1}}{\sqrt{h_{t-1}}} + \beta \log(h_{t-1}) + \phi D_t$, where $\log(\cdot)$ is the natural logarithm and α examines the effect of recent shocks, while β measures the persistence. To ensure h_t is stationary, β should be less than one. γ captures the asymmetric response of the conditional variance to recent negative shocks in comparison with the response to recent positive shocks. D_t is the commenting/betting activity dummy. ϕ examines whether one specific commenting/betting activity on Reddit affects the conditional variance of changes in price differentials. The estimate of ϕ is of major interest regarding the effects of different kinds of commenting/betting activities on the volatility of the changes in pricing differentials. The evidence sheds light on how those activities impact the variation dynamics of pricing inefficiencies.

a) By event type



b) By year



c) By Reddit rating

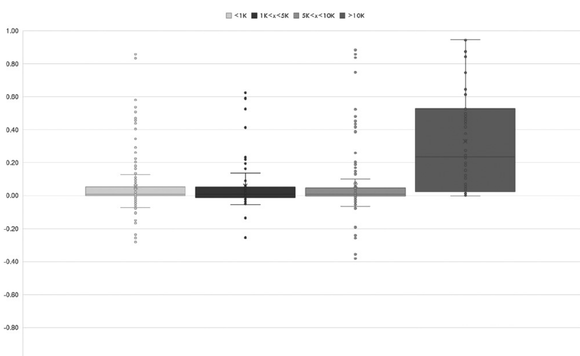


Figure 7: EGARCH-calculated Volatility Due to Identified Manipulative Events Related to r/wallstreetbets.

Note: Further calculations and related forum data available from the authors upon request.

presented earlier in terms of abnormal results and liquidity, indicating not only that r/wallstreetbets can generate substantial market effects but also that these effects are not in line with the fundamental estimates of industry experts.

5.5 Robustness Testing Procedures

When considering the substantive issues surrounding the use of message boards during the dot-com crisis (Das and Chen, 2007; Sabherwal *et al.*, 2008; Hanke and Hauser, 2008; Leung and Ton, 2015), much evidence has suggested that technology stocks, primarily those surrounding Internet startups, are among the key targets for such tactical manipulation. To add methodological robustness to our selected frameworks, we conduct two additional robustness checks through which we can present further evidence of methodological validation. First, we focus on the segregation of results to analyze whether the identified effects of investor attention sourced within the r/WallStreetBets forum focus on one particular sector when compared to another. To conduct such an analysis, we utilize the TRBC Economic Sectors, which is the Primary Thomson Reuters Business Classification (TRBC) Economic Sector Description, where companies are classified with increasing granularity by economic sector, business sector, industry group, industry and activity. We therefore consider the responses of each methodological structure to identify whether the above results are driven by a particular sectoral obsession, as was the case during the dot-com crisis. This robustness check is repeated both for returns and market liquidity. The second set of robustness checks focuses on the repetition of the above analyses using a placebo set of the data, frequency and timing of analysis to specifically test whether the estimated responses to our selected events are significantly different from zero. The results are estimated via the inclusion of firm fixed effects and the same variety of included risk factors that were used in the above analyses.

The sectoral effects that r/WallStreetBets investor attention has on abnormal returns are presented in Table 12. Across each of the selected methodological structures (Fama and French, 1993; Fama and French, 2017; Fama and French, 2018), there are quite uniform results, in which there are no substantial outliers present. Across all TRBC sectors analyzed, no significant estimates for $D_{\tau-1}.C_i$ are found, indicating that there are no significant abnormal returns identified for any of the algorithmically identified events on the previous day. This result further supports the earlier identified estimates that there is no evidence of pre-event abnormalities or investors attempting to invest en masse prior to the forthcoming accused manipulative attempt. On the day of the event, estimated through the results for $D_{\tau}.C_i$, across the three methodological structures, both financial and technological stocks present significantly different returns from zero, particularly the latter, for example,

Table 12: Industry-specific Results on Excess Stock Returns.

	F&F3 (Fama and French, 1993)							
Window	Basic Mat.	Cons. Cyc.	Cons. Noncyc.	Energy	Financial	Healthcare	Industrial	Technology
$D_{\tau-1}.C_i$	0.0377	−0.0060	0.0152	0.0013	0.0106	−0.0117	0.0055	0.0212
	(0.2665)	(0.0053)	(0.0156)	(0.0049)	(0.0168)	(0.0102)	(0.0050)	(0.0340)
$D_{\tau}.C_i$	0.2361	0.0851	0.0678	0.0020	0.0537***	0.0093	0.0160	0.2816***
	(0.2087)	(0.0763)	(0.2010)	(0.0053)	(0.0134)	(0.0101)	(0.0156)	(0.0354)
$D_{\tau+1}.C_i$	0.2094	0.1519***	0.1499*	0.0011	0.0400***	0.0141**	0.0100	0.2611***
	(0.2087)	(0.0567)	(0.0789)	(0.0027)	(0.0084)	(0.0045)	(0.0160)	(0.0337)
R^2	0.0717	0.2299	0.0684	0.0290	0.1913	0.0431	0.1566	0.2492
	F&F5 (Fama and French, 2017)							
Window	Basic Mat.	Cons. Cyc.	Cons. Noncyc.	Energy	Financial	Healthcare	Industrial	Technology
$D_{\tau-1}.C_i$	0.0309	−0.0372	0.0153	0.0017	0.0034	−0.0057	0.0022	0.0180
	(0.0806)	(0.0448)	(0.0235)	(0.0091)	(0.0065)	(0.0113)	(0.0059)	(0.0477)
$D_{\tau}.C_i$	0.2311	0.1091*	0.0490	0.0083	0.0445***	0.0313	0.0315	0.2749***
	(0.2672)	(0.0520)	(0.0304)	(0.0981)	(0.0030)	(0.0111)	(0.0086)	(0.3173)
$D_{\tau+1}.C_i$	0.2095	0.1512***	0.1625***	0.0055	0.0595***	0.0135***	0.0258	0.2627***
	(0.2092)	(0.0484)	(0.0119)	(0.0052)	(0.0091)	(0.0051)	(0.0191)	(0.2052)
R^2	0.0701	0.2190	0.0632	0.2981	0.1422	0.0577	0.1499	0.2601
	F&F5 w/Mom. (Fama and French, 2018)							
Window	Basic Mat.	Cons. Cyc.	Cons. Noncyc.	Energy	Financial	Healthcare	Industrial	Technology
$D_{\tau-1}.C_i$	0.0440	−0.0299	0.0154	0.0020	0.0120	−0.0100	0.0030	0.0089
	(0.0565)	(0.0538)	(0.0392)	(0.0084)	(0.0105)	(0.0142)	(0.0088)	(0.0055)
$D_{\tau}.C_i$	0.2339	0.0778	0.0695	0.0054	0.0484***	0.0057	0.0136	0.2568***
	(0.1872)	(0.0783)	(0.0508)	(0.0091)	(0.0110)	(0.0140)	(0.0277)	(0.2084)
$D_{\tau+1}.C_i$	0.2041	0.1622***	0.1008***	0.0064	0.0422***	0.0191***	0.0189	0.2636***
	(0.1466)	(0.0581)	(0.0199)	(0.0048)	(0.0093)	(0.0064)	(0.0285)	(0.2160)
R^2	0.0711	0.2257	0.0589	0.0280	0.1672	0.0429	0.1469	0.2816

Note: In the above table, we separate the previous methodologies based on the sector in which each stock is traded. The sectors are defined as TRBC Economic Sectors, which is the Primary Thomson Reuters Business Classification Economic Sector Description, where companies are classified with increasing granularity by economic sector, business sector, industry group, industry, and activity. The panels are presented based on the use of methodological structures following the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 13: Industry-specific Results on Liquidity.

Window	Basic Mat.	Cons. Cyc.	Cons. Noncyc.	Financial	Healthcare	Industrial	Technology
$D_{\tau-1}.C_i$	0.0270	0.0260	0.0325	0.0046	−0.0721	0.1094	0.0059
	(0.0310)	(0.0144)	(0.0151)	(0.0218)	(0.0661)	(0.0978)	(0.0051)
$D_{\tau}.C_i$	0.0833	0.0456***	0.0577***	0.0782***	0.0564	0.0655***	0.1602***
	(0.0690)	(0.0067)	(0.0104)	(0.0279)	(0.0686)	(0.0104)	(0.0059)
$D_{\tau+1}.C_i$	0.1071	0.0033	0.0031	0.0307*	0.0091***	0.0573***	0.1412***
	(0.0346)	(0.0083)	(0.0164)	(0.0183)	(0.0013)	(0.0019)	(0.0051)
R^2	0.1093	0.1361	0.1016	0.0571	0.1126	0.1301	0.1530

Note: In the above table, we separate the previous methodologies based on the sector in which each stock is traded. The sectors are defined as TRBC Economic Sectors, which is the Primary Thomson Reuters Business Classification Economic Sector Description, where companies are classified with increasing granularity by economic sector, business sector, industry group, industry and activity. The panels are presented based on the use of methodological structures following the work of Fama and French, 1993; Fama and French, 2015; Fama and French, 2018. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

when considering the substantial number of attempts to manipulate the price of Tesla and GameStop. All results with regard to technology stocks are found to be 5.6% abnormally different and significant at the 1% level. There are also significantly abnormal returns presented for consumer cyclical stocks under the Fama and French (2017) framework. On the day after the identified event, as measured by the variable $D_{\tau+1} \cdot C_i$, significant results are identified across all three frameworks for the consumer cyclical, consumer noncyclical, financial, healthcare, and technology sectors. Such results indicate that on the day on which investor attention is channeled through commentary on the r/WallStreetBets forum, both financial and technological stocks appear to be immediately susceptible. These results are further supported through the repeated selection of manipulative proponents to initiate conversation and coordinated responses against stocks and options relating to companies such as Apple, Micron Technologies, Microsoft, Snapchat, Tesla, Virgin Galactic Holdings, Uber, and WeWork. These companies account for approximately 27% of the entire sample of companies that were algorithmically determined to be exposed to potential manipulative techniques. Further, in Table 13, for the same companies that we identify, the same trends in terms of liquidity indicate that both abnormal returns and trading volume in the same capacity directly are influenced by these events. Such results not only provide further robustness to those results previously identified but also present evidence that a broad range of companies and sectors have been exposed to such significant external attention and attempted crowd-driven manipulative techniques.

We finally examine the robustness of our presented results when compared to the incorporation and addition of undisclosed features that influence the returns and liquidity of our algorithmically identified events, beyond that of the identified effects outside of the influence of investor attention sourced within the focus of the r/WallStreetBets forum. We re-estimate our CAR analysis within the context of a difference-in-difference testing procedure.[9] To complete such a task, while using the same company returns and trading volume, we proceed to re-estimate our selected methodology in the sample period, which is taken to be 1 month (20 days) after each algorithmically identified event, and re-estimate our results to calculate γ' using the same methodological specification as before.

The results of this placebo test are presented in Table 14, where excess returns are presented in the upper panel, and liquidity as measured through trading volume is presented in the lower panel. We can clearly identify that in terms of both market returns and liquidity, none of the responses to our selected events are significantly different from zero, which is particularly

[9] More specifically, we re-estimate the following regression analyses: $\mathrm{CAR}_T = \alpha + \beta_0 \mathrm{C}_i + \beta' z_i + \epsilon_i$ where CAR_t represents the cumulative abnormal returns of the target company, firm i over a given event window, z_i represents a vector of firm characteristics calculated over the year prior to the event.

Table 14: Difference-in-difference Testing Procedures for Excess Returns and Trading Volume.

		Excess returns				
Methodological structure	Window	$D_{\tau-2}.C_i$	$D_{\tau-1}.C_i$	$D_{\tau}.C_i$	$D_{\tau+1}.C_i$	$D_{\tau+2}.C_i$
F&F3 (Fama and French, 1993)	$[-1,+1]$	0.0000	0.0006	−0.0023	−0.0058	0.0000
		(0.0000)	(0.0491)	(0.0411)	(0.0497)	(0.0000)
	$[-2,+2]$	0.0024	0.0024	0.0059	0.0042	−0.0093
		(0.0551)	(0.0899)	(0.0441)	(0.0426)	(0.0477)
F&F5 (Fama and French, 2017)	$[-1,+1]$	0.0001	−0.0021	0.0035	0.0072	0.0000
		(0.0021)	(0.0678)	(0.0360)	(0.0428)	(0.0000)
	$[-2,+2]$	−0.0024	−0.0029	0.0030	0.0042	0.0102
		(0.0481)	(0.0822)	(0.0406)	(0.0280)	(0.0271)
F&F5 w /M (Fama and French, 2018)	$[-1,+1]$	0.0000	0.0001	−0.0047	−0.0065	0.0000
		(0.0000)	(0.0443)	(0.0366)	(0.0527)	(0.0000)
	$[-2,+2]$	0.0048	0.0022	−0.0062	−0.0041	−0.0127
		(0.0500)	(0.0705)	(0.0278)	(0.0401)	(0.0439)

		Trading volume				
Methodological structure	Window	$D_{\tau-2}.C_i$	$D_{\tau-1}.C_i$	$D_{\tau}.C_i$	$D_{\tau+1}.C_i$	$D_{\tau+2}.C_i$
Trading volumes	$[-1,+1]$	0.0000	0.0018	0.0049	0.0188	0.0003
		(0.0000)	(0.1318)	(0.0140)	(0.0001)	(0.0000)
	$[-2,+2]$	0.0006	0.0017	0.0063	0.0133	0.0040
		(0.0477)	(0.1366)	(0.0141)	(0.0588)	(0.0560)

Note: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

interesting when we consider that the results are stable when considering the inclusion of firm fixed effects and the variety of included risk factors. Further, the result that there is no significant difference between the target firms or the set of selected control firms at the revised announcement date provides evidence that the earlier results cannot be determined to have been generated by spurious effects. Specifically, the algorithmically determined dates on which the r/WallStreetBets Reddit forum has been identified, either explicitly or implicitly, as attempting to propagate manipulation techniques are not contaminated by observable factors and external shocks and are found to generate significant effects both in terms of returns and liquidity. Such an outcome is of particular concern in the context of regulation, policy making and those responsible for the maintenance of financial market stability.[10]

6 Conclusions

Throughout the period from 2018 to 2021, the Reddit forum r/wallstreetbets has grown at an exponential rate, with evidence suggesting that companies receiving forum attention have experienced unprecedented levels of options trading volumes in the days surrounding identified manipulative events. Targeted stocks, with prior options trading levels below 10% of stock liquidity, have experienced levels of up to 100% and above. In each presented analysis, regardless of either the sample size or window investigated, the target sample results, based on several testing applications, are significantly above those for the control sample. There is a clear, persistent upward trend in the days following the algorithmically identified dates on which the r/wallstreetbets forum instigates focused investor attention on the target sample of companies. Our analysis suggests the existence of sustained abnormal returns of approximately +114 basis points in the five-day window surrounding such manipulative events. Short-term results range from +414 through +542 basis points. Further, we note that when we focus on the 1-, 3-, and 6-month periods thereafter, cumulative abnormal returns remain pronounced and significant at the 1% level, ranging between +538 and +804 basis points 1 month later, between +553 and +966 basis points 3 months later, and finally, between +428 and +748 basis points 6 months later. Market liquidity effects are found to respond similarly, and there is evidence of a positive correlation with the growth in r/wallstreetbets registered users. Further differential effects are identified depending on whether the identified manipulative behavior is implicitly or explicitly incited in the forum. Companies with R&D capacity are not only the ones most likely to show substantial CARs within the specified models

[10]Further discussion based on these key results are presented in Section C of the Online Appendix.

but also are those most likely to be subjected to such manipulative forces due to their propensity to generate rapid abnormal returns in a very short time frame.

While significant effects on abnormal returns and liquidity are identified in the preliminary analysis, we next identify significant mispricing and errors in analysts' recommendations based upon the respective I/B/E/S ratios of each analyzed stock. Evidence suggests the presence of significant mispricing in the period after such manipulative attempts, based upon intentions to pump stock prices, that is, drive prices upwards, generating substantial deviations from professional and industrial estimations, depending on forum reach and timing.

There are substantial contextual differences between this online forum and its previous analogues: the forum's global scale, the immediacy of information transfer in comparison to that in the period surrounding the dotcom collapse, the ease of digital cloaking, the broad availability of leveraged derivative trading accounts, and finally, the presence of algorithm trading and systems that instantly respond to momentum and positional movement. Using leverage and coordinated action, groups of small traders can theoretically generate significant effects, where smaller or illiquid corporations could be quite vulnerable. While the events described in this paper are not unlike those that took place on the messaging boards both before and during the dotcom collapse, financial and technological development over the past two decades has generated an environment in which pronounced, coordinated, manipulative action can destabilize corporate entities. Specific, high-frequency analyses of such trading dynamics and leverage use, alongside regulatory and legal scrutiny, would constitute useful avenues for future research.

From a regulatory perspective, recent r/wallstreetbets activity has generated significant alarm. The January 2021 manipulative attack on GameStop provided evidence of the disruption that this group can generate. Our results present substantial evidence that the group can generate not only abnormal returns but also significant volatility effects along with widespread instability. The group's modus operandi is based on an anti-establishment view or, more recently, an anti-hedge fund strategy, driven by financial anarchists who aim to either beat the system or break it, while still seeking "tendies" in the same manner as their dotcom predecessors. The added dimension of a group equally content with generating disruption and financial destruction and with generating losses correctly generates anxiety among regulators, policymakers and governments alike. The alarm raised by hedge funds in early 2021, when they attempted to respond to the use of Robinhood and preempt the actions of r/wallstreetbets traders, indicates that these "autists" have effected a great change in the financial world.

References

Ackert, L., L. Jiang, H. Lee, and J. Liu. 2016. "Influential investors in online stock forums". *International Review of Financial Analysis*. 45: 39–46.

Akyildirim, E., S. Corbet, D. Cumming, B. Lucey, and A. Sensoy. 2020. "Riding the Wave of Crypto-Exuberance: The Potential Misusage of Corporate Blockchain Announcements". *Technological Forecasting and Social Change*. 159.

Allen, F., X. Gu, and J. Jagtiani. 2021. "A Survey of Fintech Research and Policy Discussion". *Review of Corporate Finance*. 1.

Antweiler, W. and M. Frank. 2004. "Is all that talk just noise? The information content of Internet stock message boards". *Journal of Finance*. 59(3): 1259–1294.

Bhattacharya, U., N. Galpin, R. Ray, and X. Yu. 2009. "The role of the media in the internet IPO bubble". *Journal of Financial and Quantitative Analysis*. 44(3): 657–682.

Cioroianu, I., S. Corbet, and C. Larkin. 2021. "The differential impact of corporate blockchain-development as conditioned by sentiment and financial desperation". *Journal of Corporate Finance*. 66: 101814.

Comerton-Forde, C. and T. J. Putniņš. 2015. "Dark trading and price discovery". *Journal of Financial Economics*. 118(1): 70–92.

Das, S., M.-J. Asís, and P. Tufano. 2005. "EInformation: A clinical study of investor discussion and sentiment". *Financial Management*. 34(3): 103–137.

Das, S. and M. Chen. 2007. "Yahoo! for Amazon: Sentiment extraction from small talk on the Web". *Management Science*. 53(9): 1375–1388.

Dellarocas, C. 2003. "The digitization of word of mouth: Promise and challenges of online feedback mechanisms". *Management Science*. 49(10): 1407–1424.

Dellarocas, C. 2006. "Strategic manipulation of internet opinion forums: Implications for consumers and firms". *Management Science*. 52(10): 1577–1593.

Deng, S., A. P. Sinha, and H. Zhao. 2017. "Adapting sentiment lexicons to domain-specific social media texts". *Decision Support Systems*. 94: 65–76.

Fama, E. F. and K. R. French. 1993. "French, 1993, Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics*. 33(1): 3–56.

Fama, E. F. and K. R. French. 2015. "A five-factor asset pricing model". *Journal of Financial Economics*. 116(1): 1–22.

Fama, E. F. and K. R. French. 2017. "International tests of a five-factor asset pricing model". *Journal of Financial Economics*. 123(3): 441–463.

Fama, E. F. and K. R. French. 2018. "Choosing factors". *Journal of Financial Economics*. 128(2): 234–252.

Hanke, M. and F. Hauser. 2008. "On the effects of stock spam e-mails". *Journal of Financial Markets*. 11(1): 57–83.

Hu, Y., Y. Hou, L. Oxley, and S. Corbet. 2021. "Does blockchain patent-development influence Bitcoin risk?" *Journal of International Financial Markets, Institutions and Money*. 70.

Jian, L. and R. Sami. 2012. "Aggregation and manipulation in prediction markets: Effects of trading mechanism and information distribution". *Management Science*. 58(1): 123–140.

Jiang, G., P. Mahoney, and J. Mei. 2005. "Market manipulation: A comprehensive study of stock pools". *Journal of Financial Economics*. 77(1): 147–170.

Khwaja, A. and A. Mian. 2005. "Unchecked intermediaries: Price manipulation in an emerging stock market". *Journal of Financial Economics*. 78(1): 203–241.

Leung, H. and T. Ton. 2015. "The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks". *Journal of Banking and Finance*. 55: 37–55.

Luca, M. and G. Zervas. 2016. "Fake it till you make it: Reputation, competition, and Yelp review fraud". *Management Science*. 62(12): 3412–3427.

Meegan, A., S. Corbet, C. Larkin, and B. Lucey. 2021. "Does cryptocurrency pricing response to regulatory intervention depend on underlying blockchain architecture?" *Journal of International Financial Markets, Institutions and Money*. 70.

Nelson, D. B. 1991. "Conditional heteroskedasticity in asset returns: A new approach". *Econometrica: Journal of the Econometric Society*: 347–370.

Philippi, S., M. C. Schuhmacher, N. Bastian, *et al.* 2021. "Attracting Investors in Initial Coin Offerings: The Relevance of Specific Technological Capabilities for Fundraising Success". *Review of Corporate Finance*. 1(3-4): 455–485.

Renault, T. 2017. "Intraday online investor sentiment and return patterns in the U.S. stock market". *Journal of Banking and Finance*. 84: 25–40.

Rice, D. R. and C. Zorn. 2021. "Corpus-based dictionaries for sentiment analysis of specialized vocabularies". *Political Science Research and Methods*. 9(1): 20–35.

Sabherwal, S., S. Sarkar, and Y. Zhang. 2008. "Online talk: Does it matter?" *Managerial Finance*. 34(6): 423–436.

Sabherwal, S., S. Sarkar, and Y. Zhang. 2011. "Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news". *Journal of Business Finance and Accounting*. 38(9-10): 1209–1237.

Sprenger, T., P. Sandner, A. Tumasjan, and I. Welpe. 2014. "News or noise? Using twitter to identify and understand company-specific news flow". *Journal of Business Finance and Accounting*. 41(7–8): 791–830.