When Crowds Aren't Wise: Biased Social Networks and its Price Impact

Edna Lopez Avila, Charles Martineau, and Jordi Mondria*

May 6, 2023

ABSTRACT

Information production from investment social networks around earnings announcements matters to price efficiency. Social networks' content is excessively optimistic and associated with buying pressure before announcements. Such pressure deviates prices away from fundamentals before negative news and towards fundamentals before positive news. In rare cases of extreme pessimism, we find selling price pressure before positive and negative earnings news. Surges in retail trading and investors' beliefs susceptible to manipulation amplify these effects. Our results suggest that social networks induce optimistic trading, consistent with a model of wishful thinking.

JEL Classification: E50, G12, G14.
Keywords: attention, earnings announcements, price efficiency, price impact, retail trading, social networks, Stocktwits, Seeking Alpha, WallStreetBets, wishful thinking

*We thank Pat Akey, Alex Corhay, Marina Niessner, Yoshio Nozawa, Terry Odean, Chay Ornthanalai, Mike Simutin, and Xiaofei Zhao for valuable feedback. We acknowledge the financial support from the Canadian Securities Institute, the Social Sciences and Humanities Research Council (SSHRC), and TD MDAL. Jay Cao at TD MDAL and Kevin Li provided excellent research assistance. Lopez Avila: Rotman School of Management, University of Toronto, 105 St-George, Toronto ON, Canada, M5S 3E6 (edna.lopez.avila@rotman.utoronto.ca), Martineau: Rotman School of Management and UTSC Management, University of Toronto, 105 St-George, Toronto ON, Canada, M5S 3E6 (charles.martineau@rotman.utoronto.ca),Mondria: University of Toronto, 150 St. George Street, Toronto ON, Canada, M5S 3G7 (jordi.mondria@utoronto.ca).

1. Introduction

More than a third of new investors use social media to research investment advice (CNBC, 2021). If such advice, though noisy, is independently produced by different contributors, it can benefit investors, as averaging independent judgments of others generally improves accuracy. This is known as the wisdom-of-the-crowd effect. However, information sharing on such networks can amplify noise if the information being shared is not independently produced, resulting in information herding and undermining the wisdom-of-the-crowd effect (Kahneman, Sibony, and Sunstein, 2022). Social media platforms do not guarantee independent information aggregation without external influence from other users' advice. They disseminate information using engagement algorithms influenced by popularity bias.[1]

Cookson, Engelberg, and Mullins (2023) find that financial social networks can serve as a platform for users to consume information that reinforces their pre-existing beliefs, resulting to echo chambers, which can undermine wisdom-of-the-crowd effects. Moreover, when social media information is easy to process, it can induce subjective belief trading, resulting in irrational trading behavior such as optimism and wishful thinking (Caplin and Leahy, 2019).

This paper examines how information from social media (StockTwits, WallStreetBets, and Seeking Alpha) impacts aggregate prices and price efficiency around earnings announcements. Earnings announcements provide an ideal setting for our analysis because conventional sources of information, such as media and analysts' reports, are limited in the days leading up to these events. In contrast, investor social networks experience a surge in information production before earnings announcements. If investors trade in line with social media's "wisdom," then stock prices leading up to earnings announcements should reflect

[1] Algorithms used by such networks are designed to enage users with personalized and relevant information, which could eventually lead to confirmation bias, echo chambers, and ultimately to misinformation spreading. Lorenz, Rauhut, Schweitzer, and Helbing (2011) show that social influence can produce herding behavior and negative side effects for the mechanism underlying the wisdom of the crowds. Nikolov, Oliveira, Flammini, and Menczer (2015) find that algorithmic filters have biases affecting access to information on social media platforms.

future fundamentals such as earnings announcement surprises. However, if the information is generated in echo chambers resulting in subjective beliefs trading as predicted by wishful thinking models, social networks can be detrimental to price efficiency.

We first show that positive sentiment outlooks account for more than 80% of social networks' post activity days before earnings announcements for nearly 70% of the announcement sample. Social media displays even more optimism than sell-side financial analysts.[2] Such excess positivism, if transmitted to trading behaviors of investors, will have repercussions on aggregate prices and price efficiency.

We indeed find that stocks with an abnormally high number of posts on social networks before earnings announcements are associated with higher retail trading activity in equity and options markets. More importantly, consistent with aggregate sentiment being excessively positive, posts activity is associated with greater buying pressure. While such buying pressure can be beneficial for price efficiency before earnings announcements, it can also be detrimental. For stocks with an abnormally high number of posts on social media, we find greater price run-ups of 1% from five days before earnings announcement. These price run-ups occur regardless of whether the announcement has a positive or negative earnings surprise, suggesting that prices become more efficient before positive news as they converge to fundamentals, but more inefficient before negative news as they drift away from fundamentals. Smaller market capitalization stocks experience even larger price run-ups before announcements, with increases of up to 2%. The association between price runs and abnormal social media coverage is robust to controlling for upcoming earnings surprises, abnormal newswire coverage, newswire sentiment, and analyst-recommendation news.

Our findings are mostly driven by the information shared on StockTwits. This is because StockTwits covers a wider cross-section of stocks and the total number of posts far exceeds the ones produced on WallStreetBets and Seeking Alpha. Even though the information on

[2]It has been well documented that sell-side analysts generally provide favorable research reports (e.g., Francis and Philbrick, 1993, Michaely and Womack, 1999, Jackson, 2005).

Seeking Alpha is passed through editors, and created by non-anonymous users who are often educated and experienced, we find no relationship between the content of their posts days leading to earnings announcements and stock fundamentals. Just like analysts publishing recommendations, the majority of Seeking Alpha's posts are created several days after earnings announcements. Another important aspect of StockTwits is that more than 30% of posts focus on small cap stocks (bottom NYSE breakpoint quintile).

Prior research shows that retail investors are attracted to news events eliciting them to buy rather than to sell because selling involves the investor owning the stock (Barber and Odean, 2008). This tendency may explain the existence of positive sentiment among users on social media platforms. However, in rare cases of extreme negativism on Stocktwits, we find evidence of downward price pressure before earnings announcements, thus distorting price efficiency before positive earnings news and improving price efficiency before negative news.

In attempt to shed light on a plausible causal relationship between social media content and price efficiency, we use rounds of stimulus checks during the COVID-19 pandemic as exogenous shocks to retail trading. Greenwood, Laarits, and Wurgler (2022) find that rounds of stimulus checks during the COVID-19 pandemic led to a spur in retail trading. For stocks with high abnormal social media attention, we find an exacerbation in upward price pressure before earnings announcements following the issuance of stimulus checks. The effect is more pronounced for small stocks, where cumulative returns before earnings announcements increased by more than 5% following rounds of stimulus checks.

Previous studies (Chen, De, Hu, and Hwang, 2014, Bartov, Faurel, and Mohanram, 2018, Dim, 2020) find that the content of social media posts can predict fundamental factors such as earnings surprises. Our findings cast doubt that such "wisdom," transmits to investors trading decisions as social media sentiment is associated with price run ups days prior to earnings announcements, whether firms beat or miss earnings targets. We find that social media posts in the days leading to earnings announcements does indeed predict earnings surprises; but only for large stocks.[3] This finding is explained by a sample composition effect where more than 70% of earnings surprises are positive and thus correlates with the excess positivism displayed on social media.

We then present a simple theoretical framework that ratioanlize why social media can induce investors to trade optimistically. Our model is based on "wishful thinking" of Caplin and Leahy (2019), which posits that individuals derive utility from their beliefs and thus tend to interpret information optimistically. The model predicts that investors will display positive (negative) optimism when seeking to buy (sell) stocks. It is well-known that retail investors are more inclined to buy than sell (Barber and Odean, 2008) and, consistent with our findings, we expect investors to display more positive optimism. Furthermore, our model predicts that investors' beliefs can be more easily influenced by subjective factors and thus behave more like wishful-thinking investors when social media content is more easily interpretable, e.g., when the sentiment signal is less "noisy." Using post activity on StockTwits without sentiment (commonly attributed to "bots" activity) as a proxy for cross-sectional variation in noise, we find price run-ups before earnings announcements only for announcements with low noise activity.

Overall, paper contributes to the growing literature examining the impact of social media to financial markets (Chen, De, Hu, and Hwang, 2014, Jiao, Veiga, and Walther, 2020, Cookson and Niessner, 2020, Bradley, Hanousek Jr, Jame, and Xiao, 2021, Hu, Jones, Zhang, and Zhang, 2021, Cookson, Engelberg, and Mullins, 2023, Cookson, Niessner, and Schiller, 2022). While social media platforms have the potential to improve information sharing, the influence of algorithms on content creates a risk of confirmation bias and exposure to misinformation for their users. This can ultimately impact financial efficiency around news

[3]Our analysis is not a replication of previous studies examining the link between social media content and earnings surprises, as we have a different platforms, focus specifically on the information content on the days leading to earnings announcements, and we conduct our analysis in a period during significant growth in retail trading and social media content.

events, as evidenced by our findings on the effects of social media sentiment on stock prices before earnings announcements.

Cookson, Lu, Mullins, and Niessner (2022) highlights the differences between attention and sentiment on the information produced in Stocktwits and Seeking Alpha. They attribute these differences to users' sophistication and the character limit of posts on both platforms. Farrell, Green, Jame, and Markov (2022) find Seeking Alpha posts can immediately impact retail trading minutes following its release. We find evidence that Seeking Alpha posts days priors to earnings announcements relate to higher retail trading but of much smaller magnitude than WallStreetBets and StockTwits and with little price impact.

Bradley, Hanousek Jr, Jame, and Xiao (2021) and Hu, Jones, Zhang, and Zhang (2021) document the impact of the recent rise of WallStreetBets on financial markets. Hu, Jones, Zhang, and Zhang (2021) find that a more positive tone and higher WallStreetBets connectedness predicts higher returns, greater and more positive retail order flow, and lower shorting flows the next day. We show that investors' tone on WallStreetBets is positively skewed and relate to retail order flow but fail to predict firm earnings fundamentals.

Finally, our paper relates to the literature on retail investors' performance, attention-induced trading, and returns. Barber and Odean (2008), and Barber, Huang, Odean, and Schwarz (2022) find that retail investors are inclined to trade high-attention stocks, i.e., stocks in the news, stocks experiencing high abnormal trading volume and one-day returns, and stocks displayed in a "top mover" list in Robinhood app. In line with this result, our paper shows that unsophisticated investors are inclined to trade stocks with high coverage on social media platforms. In line with previous work (Barber and Odean, 2000, Barber, Lee, Liu, and Odean, 2009, Xiaoyan and Zhang, 2022), our paper provides additional evidence as to why, in aggregate, retail investors earn poor returns. Our paper's findings thus warn investors that consuming information from social media can be hazardous to their wealth.

2. Sample Description

The time coverage of this study spans two periods: (1) January 1, 2018, to December 31, 2021, when incorporating information from the three social media platforms of our analysis, and (2) from January 1, 2013, to December 31, 2021, when focusing solely on StockTwits. We select stocks with share codes 10 or 11 from the Center for Research in Security Prices (CRSP) and retrieved their corresponding ticker symbols, daily returns, prices, outstanding shares, and market capitalization. To ensure we have the necessary information for merging with social media posts, we only include stocks with available tickers in CRSP. We also use stock-related news from Ravenpack to control the analysis of information production on social media. We also retrieve retail trading data from TAQ following the approach of Eaton, Green, Roseman, and Wu (2021)

Analyst forecasts and earnings announcements are from Thomson Reuters I/B/E/S. We consider earnings announcements in IBES that meet the following requirements: the earnings date is reported in Compustat, the price of the stock of five days before the announcement is available in CRSP, and the stock price is available on Compustat as of the end of the quarter. We compute the surprise earnings announcement $Surprise_{i,\tau}$ as the difference between the firm earnings per share of quarterly earnings announcement and the consensus analysts forecast, divided by the prices of the stock five days before the earnings announcement day from (I/B/E/S) and Compustat for stock i and earnings announcement τ. The consensus analysts' forecasts consider the median of all analysts' estimates issued over the 90 days before the earnings announcement date. Finally, we winsorize the earnings surprise variable at the 1st and 99th percentile. As in Gregoire and Martineau (2022), we further retrieve analyst recommendations for our sample of stocks from Ravenpack.

2.1. Social media information production and sentiment

Our analysis is based on data obtained from the three leading social media platforms: WallStreetBets, Seeking Alpha, and StockTwits. StockTwits is a social media platform similar to Twitter, where users can post messages or "tweets" about a stock adding a $ Cashtag followed by the stock ticker symbol to express their opinion about it (Exs: $AMZN, $GOOG, $SNAP). Additionally, users on this platform are enabled to tag their posts as either "Bullish" or "Bearish". We obtain all posts from the social media platform via RapidAPI.

For Seeking Alpha, the users are required to refer to a company by its first name and include its stock ticker whenever they mention it in an article that has a longer format than a tweet. This platform offers four distinct options to include a sentiment feature for each article, ranging from "Very Bullish," "Bullish," "Neutral," "Bearish," to "Very Bearish." Like StockTwits, posts on Seeking Alpha are stock-specific and have an explicit sentiment assigned by their respective authors. For the purpose of our analysis, we generalized the classification of these posts by categorizing all posts labeled as "Very Bullish" and "Bullish" as bullish, and those tagged as "Bearish" or "Very Bearish" as bearish. We obtain all the opinion articles of Seeking Alpha through RapidAPI.

Unlike StockTwits and Seeking Alpha, WallStreetBets does not offer the option to tag posts by stock or sentiment. In this sense, there is no direct way to identify which stock a post refers to or the sentiment expressed toward it. To address this issue, we scrap all the tickers considered in our analysis and all the company names related to the ticker. For example, for Bank of America, we search for "Bank of America," "BAC," and "BofA." We consider all the posts where the symbol or word of the stock was mentioned at least once, either in the title or in the post's body text. Next, we proceed to calculate the sentiment of each post on WallStreetBets using Machine Learning and Natural Language Processing techniques. Unfortunately, the usage of tools like Loughran and McDonald's dictionaries is not adequate for the language used on a social media platform such as WallStreetBets.

As mentioned in Bradley, Hanousek Jr, Jame, and Xiao (2021), the users' language of this platform is full of sarcasm, jokes, bad words, slang, and emojis. Therefore, to classify the sentiment of WallStreetBets posts as bullish or bearish, we employ a Supervised Learning Method known as Support Vector Machine described in Section A of the Appendix.

In total, we gathered more than 102 million, 480 thousand, and 107 posts for StockTwits, WallStreetBets, and Seeking Alpha, respectively. Only 46,163,488 posts and 68,167 posts of StockTwits and Seeking Alpha, respectively, are tagged with a sentiment view. The evolution of the information production on these financial social networks over time is plotted in Figure 1, aggregating the monthly number of posts for each platform, from January 1, 2018, to December 31, 2021.[4] From the last quarter of 2019 onwards, social media activity on all platforms increased significantly, in line with the surge of retail trading facilitated by retail brokerages offering zero trading costs and the impact of stimulus checks during the COVID-19 pandemic in the US. The information production of StockTwits clearly exceeds the one from WallStreetBets and Seeking Alpha. Moreover, StockTwits covers a more extensive range of stocks, with 4,192 different stocks mentioned in their posts compared to 3,717 on WallStreetBets and 2,958 on Seeking Alpha. Consequently, the information produced on StockTwits will be a significant driver of the results of our analysis.

To understand the difference in information production, breadth of coverage across platforms, and subsequently the results of this analysis, it is important to understand the characteristics and differences between the three social networks. The posts from Seeking Alpha come from opinion articles that must conform to Seeking Alpha's standards of rigor and clarity. To be eligible for publication, each opinion article passes through editors with credentials including MBAs, Masters in Economics and Commerce, CFA charters, and a post-secondary degree in business journalism from Bloomberg, CNN, TheStreet.com, and MSN Money, among others. In addition, the author of each opinion article receives a payment

[4] For StockTwits and Seeking Alpha, we select all posts even if there is no sentiment labeling for a particular post.

calculated based on how many subscribers read the article. To be a subscriber and have access to all stock-related opinion articles, a regular fee must be covered. On the other hand, StockTwits and WallStreetBets are free platforms with open access to all comments posted on their social platforms. Neither of the platforms has an editorial board, and their users are not economically compensated for posting. Before May 2019, StockTwits comments had a limited length of 140 characters before increasing the limit to 1,000 characters. In contrast, Seeking Alpha and WallStreetBets have no limit of characters for their opinion articles and posts. Contributors on Seeking Alpha should not be surprised by "Decline" responses for articles that cover nanocap stocks trading below a $25 million market cap or 50c share price. However, this is not the case for StockTwits and WallStreetBets. On StockTwits, users can automatically receive all tweets posted on the platform on their feed. But they can customize their feed only by receiving tweets from stocks or users they follow. In addition, StockTwits users can disclose their experience level as a novice, intermediate, and professional. Cookson and Niessner (2020) describes that 20% of StockTwits users classify themselves as professionals, 52% as intermediate, 28% as novices. Different from opinion articles in Seeking Alpha, posts on WallStreetBets and StockTwits tend to be considerably less in-depth. According to Bradley, Hanousek Jr, Jame, and Xiao (2021), anecdotal evidence suggests that WallStreetBets also places a larger emphasis on highly speculative trading strategies.

2.2. Measuring abnormal sentiment and post activity

For each social media platform p, we aggregate sentiment at the firm-day level as follows:

$$Sent^p_{i,t} = \frac{\#Bull^p_{i,t} - \#Bear^p_{i,t}}{\#Bull^p_{i,t} + \#Bear^p_{i,t}} \tag{1}$$

Where $\#Bull^p_{i,t}$ and $\#Bear^p_{i,t}$ correspond to the number of bullish and bearish posts, respectively, for stock i on day t for platform p. $Sent^p_{i,t}$ ranges from -1 in the case where all posts are bearish and to +1 in the case where all posts are bullish.

We further compute a measure of information production based on the abnormal number of posts in the cross-section of stocks as follows:

$$Abn\ post_{i,t}^{p} = A_{i,t}^{p} - \sum_{n=1}^{N} \frac{A_{n,t}^{p}}{N}, \tag{2}$$

where $A_{i,t}^{p} = \frac{M_{i,t}^{p}}{M_{t}^{p}}$ is the proportion of posts of stock i, $i = 1, .., N$, on platform p on day t, and $\sum_{n=1}^{N} \frac{A_{n,t}^{p}}{N}$ is the average proportion of posts of all N stocks on platform p on day t. $Abn\ post_{i,t}^{p}$ measures the abnormal proportion of posts produced for a stock relative to other stocks on a given platform.

Given our analysis focuses on the information produced on social media around earnings announcements, we calculate our base measure of abnormal information production five days before earnings announcement τ as follows:

$$Abn\ post_{i,\tau}^{p} = \sum_{t=-5}^{T=-1} \frac{Abn\ post_{i,\tau,t}^{p}}{5} \tag{3}$$

Similarly, we calculate our base measure of sentiment on social media information as the average sentiment of posts five days before earnings announcements as follows:

$$Sent_{i,\tau}^{p} = \sum_{t=-5}^{T=-1} \frac{Sent_{i,\tau,t}^{p}}{5} \tag{4}$$

3. Empirical Results

3.1. Social media information production around earnings announcements

We start by examining how information on social media is produced around earnings announcements and compare it to information from analyst-related news, such as recommendations, and newswire coverage. Information production by users on social networks is less costly compared to traditional outlets. Therefore, we expect that social networks play an important role in information dissemination before announcements.

For each stock, we compute the number of abnormal posts for a stock ten days around its earnings announcements. Similar to Fisher, Martineau, and Sheng (2022), we take the difference between the number of posts on the day t minus the average number of posts from $t = -30$ to $t = -11$.[5] We plot in Figure 2 the average abnormal number of posts for StockTwits, WallStreetBets, and Seeking Alpha in Panel A. We plot also the abnormal number of analysts' recommendations and the abnormal number of news articles from RavenPack in Panel B. For social networks, we find an increase in information production five days before earnings announcements, followed by a gradual decrease over the next five days. Consistent with Figure 1, we find that the number of abnormal posts is greater for StockTwits than for the other platforms. Interestingly, post activity on Seeking Alpha generally occurs after earnings announcements. Similarly, Panel B shows that analyst recommendations occur mainly after the earnings announcement; in line with the findings of Ivković and Jegadeesh (2004), and Gregoire and Martineau (2022).

Figure 2 shows that social media networks play an important role in providing information about stocks before earnings announcements, addressing a gap in coverage that exists in traditional news sources. While traditional news typically reports on earnings results after they have been released, social media platforms allow investors to track real-time conversations and sentiment around a particular stock in the lead-up to the announcement.

We also find that social media platforms cover a greater number of stocks before earnings announcements compared to traditional news sources. Table 1 provides a detailed breakdown of the coverage of stock earnings announcements for each platform. We consider a platform to cover a stock earnings announcement if at least one post related to that stock appears on the platform within five days before the announcement date. Specifically, from January 2018 to December 2021, StockTwits covers the highest number of earnings announcements, precisely 37,756. WallStreetBets follows it with 5,569, and Seeking Alpha with 2,908 earnings

[5]Stocks with no activity is assigned a value of zero.

announcements covered. As a comparison, analysts' recommendations from Ravenpack cover 3,534 stock earnings announcements.

Panel A of Table 1 provides an overview of the coverage of stock earnings announcements by each platform based on market capitalization. Notably, earnings announcements for small firms are covered to a greater extent on social media platforms than by analysts. Specifically, the percentage of small firms covered in StockTwits is 31%, while the percentage of analysts is only 9%. Otherwise, large firms' percentage covered by analysts is more significant, and it rises to 41% of the total, compared to StockTwits, where it is only 15%. However, in number, the earnings covered by StockTwits (5,710) are more significant than for analysts (1,462). In addition, Panel B shows that the coverage of earnings announcements with positive surprise is greater than those with negative surprise earnings. Though for the three social platforms, the coverage of positive earnings announcements is similar and between 65% and 73%, the same as for analysts.

Figure 3 provides insights into the evolution of social media, newswire, and analyst coverage over time for the different platforms. The figure plots the fraction of stock-earnings announcements sample with at least one post five days before announcements. StockTwits covers approximately 90% of stock-earnings announcement observations, followed by newswire coverage with 20%. Seeking Alpha and analysts' recommendation coverage is about 10%.

Overall, social media has become increasingly important for investors seeking information on stocks before earnings announcements. This is particularly true for small-cap stocks that may not receive coverage in traditional analysts' reports. The timing of information production varies across platforms, and our findings suggest that StockTwits may play a particularly significant role in market efficiency leading up to earnings announcements.

3.2. Positively biased information

Does the information on social media before earnings announcements provide useful information on stock fundamentals? In this section, we turn our attention to the content of social media posts to assess their informativeness. As a first analysis, we examine the distribution of sentiment for every platform, by looking at the proportion of positive sentiment in relation to the total number of posts. Specifically, we calculate the daily proportion of bullish posts in relation to the total number of bullish and bearish posts, for stock i on the day t on platform p, as follows:

$$PosRatio_{i,t}^{p} = \frac{\#Bull_{i,t}^{p}}{\#Bull_{i,t}^{p} + \#Bear_{i,t}^{p}}$$

We further compute the average positive ratio $PosRatio_{i,\tau}^{p}$, of all stock-related posts shared five days before the earnings announcement τ of stock i, on platform p. Figure 4, Panel A, shows the fraction of stock-earnings announcement observations with an average $PosRatio_{i,\tau} \leq 20\%$, between 21 and 39%, 40 and 59%, 60 and 79%, and $\geq 80\%$. Remarkably, we find that more than 70% of the earnings announcements covered by StockTwits have an average positive ratio $\geq 80\%$. For Seeking Alpha, the fraction of earnings announcements with an average positive ratio $\geq 80\%$ is approximately 65%. To evaluate the significance of our findings, we compare them to the fraction of positive posts from analyst recommendations, which serves as a benchmark. Around 60% of analyst recommendations have an aggregate positive ratio higher than 80%, largely below what is found on social media platforms. Overall, we find that the information generated and disseminated on financial networks exhibits a pronounced positive skew. Surprisingly, this skew is even more significant than sell-side financial analysts. Sell-side analysts are also commonly associated with positive bias about future firm performance (e.g., Francis and Philbrick, 1993, Michaely and Womack, 1999, Jackson, 2005).

While the latter evidence already casts doubts about the informativeness of the content shared on social media, we continue to scrutinize this assumption by exploring whether this information relates to stock earnings firms' fundamentals. Specifically, we investigate whether the information shared prior to earnings announcements, as well as its corresponding sentiment, can predict both earnings surprises and future returns of the stock at the announcement date. For that, we regress the earnings surprise $SUE_{i,\tau}$ of stock i on earnings announcement τ on the average of posts' sentiment $Sent^{p}_{i,\tau}$, and the average of abnormal information $AbnPost^{p}_{i,\tau}$ over the five days preceding announcement τ and the interaction of both variables, according to equation 4 and 3, respectively. The regression specification is as follows:

$$Surprise_{i,\tau} = \beta_1 Sent^{p}_{i,\tau} + \beta_2 AbnPost^{p}_{i,\tau} + \beta_3 Sent^{p}_{i,\tau} \times AbnPost^{p}_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

for every platform p = StockTwits, WallStreetBets, Seeking Alpha, and Analysts. The results of these regressions are shown in Table 2 for the full sample, and for large and small stocks in Panels A to C, respectively. Across all panels, we find little evidence of earnings surprise predictability from social media posts' sentiment. Only for large stocks (Panel B), we find suggestive evidence of predictability. *Sent* and *Sent* $\times$ *AbnPost* are positive and statistically significant at the 5 and 10% level, respectively, and imply that earnings surprises predictability from StockTwits sentiment increases in the abnormal post coverage.[6] In contrast, for small stocks, Panel C reports a negative relationship between earnings surprises and StockTwits' sentiments. The negative relationship strengthens in the number of abnormal posts.

Taken together, our findings suggest that post activity in the days preceding earnings does not predict fundamentals on the day of the announcement. This is not surprising given the excess positive optimism observed in our previous results.

[6] The predictability for large stocks is explained by a sample composition effect where more than 70% of earnings surprises are positive and thus correlate with the excess positivism displayed on social media. We find that sentiment fails to predict negative earnings surprises.

3.3. Retail trading

Cookson, Lu, Mullins, and Niessner (2022) find that users who consume information on social networks are generally retail investors, and as noted by Barber and Odean (2008) such investors are often inclined to trade attention-grabbing stocks. It is reasonable to expect that individual investors' attention is drawn to stocks with high information production on social networks, which they may use as a source of investment advice for trading, even if the content of this information is biased or uninformative. To examine this premise, in this section, we investigate the correlation of abnormal information production and retail trading variables both in the equity and in the options market in the five days before the earnings announcement. Concretely, we regress the retail trading volume and retail order imbalance for equity and option volume on the main variable of abnormal information production of the three platforms, as follows:

$$\Delta Retail\ vlm_{i,\tau} = \sum_p \beta_p AbnPost^p_{i,\tau} + \delta_1 |Surprise|_{i,\tau} + \delta_2 |News\ sent|_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \tag{5}$$

$$Retail\ OI_{i,\tau} = \sum_p \beta_p AbnPost^p_{i,\tau} + \gamma_1 Surprise_{i,\tau} + \gamma_2 News\ sent_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \tag{6}$$

where $p = \{$Stocktwits, WalStreetbets, Seeking Alpha$\}$, corresponding to the three social media platforms. $\Delta Retail\ vlm$ is the change in retail volume from $t = [-60, -6]$ to $t = [-5, -1]$ for stock i and earning announcement τ. $Retail\ OI$ is the average daily retail volume order imbalance for $t = [-5, -1]$. We explain the computation of both measures in Section B of the Appendix. $AbnPost^p$ is the average of abnormal information over the five days preceding announcement τ of stock i for social media platform p. $Surprise$ ($|Surprise|$) is the earnings announcement (absolute) surprise for announcement τ. $News\ sent$ ($|News\ sent|$) is the daily average (absolute average) news sentiment from RavenPack five days before announcement τ.

Consistent with the motivating intuition, Panel A of Table 3 shows how abnormal posts on Stocktwtits, WallStreetBets, and Seeking Alpha positively relate to retail trading in equity and option markets after controlling for absolute earning surprises and absolute news sentiment. Notably, the relationship is stronger for StockTwits, in detail four times greater than WallStreetBets and ten times greater than Seeking Alpha. The effect of StockTwits on retail option volume is also greater than the other platforms. Importantly, Panel B provides evidence of a positive relationship between abnormal posts and positive retail stock order imbalance, highlighting that stocks with high information production on social media experience greater buying pressure from retail investors, particularly on StockTwits.

These results provide valuable insights into the relationship between social media information and the behavior of retail investors in the stock market. Our findings confirm the results of previous research by Barber and Odean (2008) that investors tend to be net buyers of attention-grabbing stocks. However, we extend these results by demonstrating that this effect is particularly true for stocks that experience high abnormal information production on social media before earnings announcements.

3.4. Price impact and price efficiency

After establishing that social media information production increases retail trading before earnings announcements, particularly for buy trades, we now investigate whether this has an impact on stock returns and price efficiency. As a first step, we define the stock's buy-and-hold abnormal returns (BHAR) from day t to day T as follows:

$$BHAR[t,T]_{i,\tau} = \prod_{t}^{T}(1 + R_{i,t}) - \prod_{t}^{T}(1 + R_{m,t}), \tag{7}$$

where $R_{i,t}$ is the daily stock return of the stock, $R_{m,t}$ is the return of the value-weighted CRSP returns.

Figure 5 shows the average of buy-and-hold abnormal returns five days before to five days after earnings announcements $BHAR[-5,5]_{i,\tau}$ for stocks with high and low abnormal posts and their corresponding 95% confidence intervals for positive and negative earnings surprises in Panels A and B, respectively. We define high (low) abnormal posts for earnings announcements of stocks that have an average of abnormal information production greater than (less or equal to) zero in the window $t = [-5, -1]$. Panel A reveals for earnings announcements with a positive surprise, stocks with a higher volume of social media information production experience greater price run-ups, of approximately 1%, five to one day before announcements. Conversely, stocks with low abnormal posts experience limited price increases prior to the announcement. This suggests that social media-induced retail trading is pushing stock prices toward their fundamentals. However, upon examining earnings with negative surprises (Panel B), we observe a similar positive cumulative return increase of 1% for stocks with high abnormal information production on social media, in contrast to stocks with low abnormal posts that demonstrate no change in cumulative returns before announcements. In line with our previous results, this evidence shows that positively biased information content from social media is highly correlated with buying retail trading activity and leads to positive price run-ups before earnings announcements.

The prior work of Barber, Huang, Odean, and Schwarz (2022) shows that price distortions are more pronounced among small-cap stocks. We repeat our analysis but for small and large stocks. As in Martineau (2022), we define small stocks as stocks belonging to the bottom two NYSE breakpoint quintiles. Figure 6 shows that price run-ups for small stocks with high abnormal posts before earnings announcement is stronger than for large stocks. The price run-up for small stocks with high abnormal posts is about 2%. For large stocks, we find limited evidence of price run-ups. Indeed, if retail traders trade on information from social media, it is expected that their trades will have a larger price impact for small than large stocks.

We next examine the statistical significance and robustness of the latter results after controlling for other factors potentially impacting price run-ups before earnings announcements.

We run the following regression:

$$BHAR[-5,-1]_{i,\tau} = \sum_p \beta_1^p \mathbb{1}_{i,\tau}^p + \sum_p \beta_2^p \mathbb{1}_{i,\tau}^p \times \mathbb{1}_i^{Small} + \beta_3 \mathbb{1}_i^{Small} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}$$

where $BHAR[-5,-1]_{i,\tau}$ is the buy-and-hold return five days before to one day before earnings announcement τ for stock i. $\mathbb{1}^p$ corresponds to indicator variables equal to one if the average abnormal number of posts of stock i five days before announcement τ is positive, zero otherwise, in the social media platform $p = \{\text{ST, WSB, SA}\}$. $\mathbb{1}^{Small}$ is an indicator variable equal to one if the stock-earnings announcement i belongs to the bottom two NYSE market capitalization quintiles. The control variables are $\mathbb{1}^{Ana}$, $\mathbb{1}^{Ana} \times \mathbb{1}^{Small}$, $\mathbb{1}^{News}$, $\mathbb{1}^{News} \times \mathbb{1}^{Small}$, $Surprise$, $News\ sent$, and $Analyst\ sent$. Similar to how we define abnormal information production on social media before earnings (see Equation 3), $\mathbb{1}^{Ana}$ ($\mathbb{1}^{News}$) is an indicator variable equal to one if the number of abnormal analyst recommendations (newswire article) of stock i before earnings announcement τ, is positive, zero otherwise. $Surprise$ is the earnings surprise of earnings announcement τ of stock i. $News\ sent$ is the average news sentiment in RavenPack five to one day before earnings announcements. α_i and α_t correspond to firm and year-quarter fixed effects, respectively. The results are reported in Table 4 for earnings announcements with upcoming positive earnings surprises in columns (1)-(4) and negative surprises in columns (5)-(8).

The first main result that stands out from Table 4 is the importance of StockTwits. Across all model specifications, the loadings for $\mathbb{1}^{ST}$ vary between 0.008 and 0.032 and are at least statistically significant at the 5% confidence level, whereas the loadings for $\mathbb{1}^{WSB}$ are positive but not statistically significant and negative for $\mathbb{1}^{SA}$. This confirms that, in aggregate, StockTwits is the main social media platform that has the largest potential impact on prices and price efficiency because it provides greater coverage in the cross-section of stocks. The loadings for $\mathbb{1}^{ST} \times \mathbb{1}^{Small}$ confirm the findings in Figure 6, that most of the effect is primarily observed in small-cap stocks. Adding the control variables does not influence the effect of abnormal post activity on pre-announcement returns.

Overall, our results confirm that stocks with high information production on social media platforms (especially on StockTwits) before earnings announcements with upcoming positive or negative surprise present price run-ups of 1%. These results are stronger for small stocks. Proving that the information on social media is highly positively skewed, therefore uninformative, and is associated with higher buying pressure from retail investors that distort price efficiency before earnings announcements.

3.5. Does negative sentiment matter?

Our analysis so far has demonstrated that abnormal information production on social media platforms is linked to an increase in cumulative stock returns before earnings announcements, reflecting the positively biased nature of social media content. However, it has not accounted for the smaller proportion of posts with a highly negative sentiment. To address this gap, in this section, we turn our attention to stocks with abnormal production of bearish posts before earnings announcements and examine the role of negative sentiment on stock prices. Given the limited sample size, we expand our data sample period from 2013 to 2021, focusing solely on StockTwits platform, which has demonstrated the greatest potential for impacting stock prices. Figure 7 presents the buy-and-hold abnormal returns from five days before and after earnings announcement for stocks with high positive sentiment ratio (≥ 0.80) and low positive sentiment ratio (≤ 0.20, i.e., negative sentiment) and high and low abnormal posts. The sentiment ratios are the daily average over the window $t = [-5, -1]$. Panels A to C present the results for the full sample, large stocks, and small stocks, respectively.

Figure 7 confirms that indeed negative sentiment matters. In particular, we observe in the second and fourth sub-panels, a downward drift in stock prices before earnings announcements for the rarer cases of extreme pessimism. Before earnings announcements with negative surprises, we find that stocks with high negative sentiment and abnormal information production on social media become more efficient as their prices move toward fundamentals.

In contrast, for earnings announcements with positive surprises, we observe that the prices of stocks with high negative sentiment and abnormal information production on social media become less efficient, as they are pushed away from fundamentals. We further report in Table 5 the specific magnitudes for the drifts before earnings announcements ($BHAR[-5,-1]]$) from Figure 7 and the statistical significance between high and low abnormal posts. The table confirms that the difference in $BHAR[-5,-1]$ between high and low abnormal stocks is statistically significant at the conventional level. Overall, conditioning on extreme pessimism suggests that prices become more efficient prior to earnings announcements with negative surprises but deviate from fundamentals before positive surprise announcements.

4. Additional Results

4.1. The effect of Government Stimulus on Retail Trading and Price Efficiency

Our findings suggest that social media platforms' content relates to retail trading, price impact, and price efficiency before earnings announcements. In an attempt to shed light on a plausible causal relationship between social media content and price efficiency, we use rounds of stimulus checks during the COVID-19 pandemic as exogenous shocks to retail trading. Greenwood, Laarits, and Wurgler (2022) find that rounds of stimulus checks during the COVID-19 pandemic led to a spur in retail trading. Table 6 reports the coefficients of the following regression:

$$\begin{aligned} BHAR[-5,-1]_{i,\tau} = &\beta_1 \mathbb{1}^{ST}_{i,\tau} + \beta_2 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_3 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Stim}_{\tau} + \\ &\beta_4 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Small}_{i} \times \mathbb{1}^{Stim}_{\tau} + \beta_5 \mathbb{1}^{Small}_{i} + \\ &\beta_6 \mathbb{1}^{Stim}_{\tau} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \end{aligned} \tag{8}$$

where $\mathbb{1}^{ST}$ is an indicator variable equal to one if the abnormal information production, of stock i five days before earnings announcement τ on platform StockTwits, is positive, zero otherwise. $\mathbb{1}^{Stim}$ is an indicator variable equal to one for earnings announcement τ occurring during 2020-Q2, and 2021-Q1, and 2021-Q2, zero otherwise. These quarters correspond to the stimulus check arrivals. The control variables are the same as in Table 4.

Columns (1)-(2), (3)-(4), and (5)-(6) reports the results for the full sample, earnings with upcoming positive earnings surprises, and negative surprises, respectively. We find the effect of stimulus to BHAR only for small stocks. The incremental effect of stimulus checks to BHAR before earnings ($\mathbb{1}^{ST} \times \mathbb{1}^{Small} \times \mathbb{1}^{Stim}$) are economically large and statistically significant. We find an increase in pre-earnings BHAR during rounds of stimulus checks of 5.3% and 3.9%, respectively. The effect magnitude is also large for negative surprises (5.1%) but not statistically significant. Overall, these sudden increases in retail trading provide suggestive causal evidence of retail trading impacting price formation in line with the content retail investors consume on social media.

4.2. Are pre-announcement returns a result of a pump and dump scheme?

A recent group of eight individuals accused of running a pump-and-dump scheme on social media was charged by the U.S. Justice Department and the SEC of earning more than $100 million in illicit stock market profits a (Ott, 2022). Fraudulent individuals using social media to manipulate prices is not a recent phenomenon (Wasik, 2013). A natural question is whether price drifts before earnings announcements for stocks that gather much attention on social media are a result of a pump-and-dump scheme. At priori, our findings indicate that the pump-and-dump scheme is not a systematic force explaining price drifts before announcements because we find limited evidence of significant return reversals following earnings announcements (see Figure 5 and 6).

To shed light on the potential role of pump-and-dump scheme, we regress post-announcement returns (BHAR[1,5]) on pre-announcement returns (BHAR[-1,-5]). We would expect a negative relationship between post- and pre-announcement cumulative returns if pump-and-dump scheme drives our price-runs leading to earnings announcements. We report the results in Table 7 for the three social media platforms. Columns (1), (4), and (7) report that BHAR[-1,-5] does not predict BHAR[1,5]. In the remaining model specifications, interacting BHAR[-1,-5] with $\mathbb{1}^{AbnPost}$ and $\mathbb{1}^{Small}$ shows no statistical significance. We conclude that the pump-and-dump scheme plays (if any) a small role in driving price drifts before earnings announcements.

5. Wishful Thinking and Earnings Announcements

The objective of this section is to provide a simple theoretical framework that can rationalize why investors consume information that is optimistically biased and trade on such information. The model is based on Caplin and Leahy (2019). Let's consider a wishful thinking investor who is considering buying $q > 0$ shares of an asset with price p before the release of the company's earnings announcement. For simplicity we will abstract about how q and p are determined and we will take them as given. After the release of the earnings announcement the asset payoff $\tilde{v}$ can take two values: a high value $v_H = p + v$ after a positive surprise or a low value $v_L = p - v$ after a negative surprise, where $v > 0$ and $v_H > v_L$. There is an objective probability for each value. With probability $\bar{\pi}_H$ there is a positive surprise and a high realization of the asset v_H and with probability $\bar{\pi}_L$ there is a negative surprise and a low realization of the asset v_L. An alternative interpretation of the objective probabilities is that these probabilities represent the consensus or mainstream opinion in case there are agents with heterogeneous information.

The key assumption of the model is that we allow the wishful thinking investor to have subjective beliefs about the probability realization of $\tilde{v}$. Let's denote π_H the subjective probability that there is a positive surprise v_H and π_L the subjective probability that there is a negative surprise v_L. These subjective beliefs may differ from objective beliefs. But deviating from objective beliefs is costly. We represent the cost of deviating from objective beliefs by the Kullback-Leibler distance:

$$\frac{1}{\theta}\pi_H \ln \frac{\pi_H}{\bar{\pi}_H} + \frac{1}{\theta}\pi_L \ln \frac{\pi_L}{\bar{\pi}_L}.$$

The parameter θ represents the ease with which the agent can manipulate their beliefs. The larger is θ the greater the amount of evidence the agent would need before they reject their chosen beliefs in favor of the objective ones. In other words, the larger θ the more likely the investor is to opt for subjective beliefs. The lower the θ, the more costly is to deviate from the objective beliefs.

The investor's expected utility of holding the asset and manipulating beliefs is then given by:

$$EU(\pi_H, \pi_L) = q(\pi_H v_H + \pi_L v_L - p) - \frac{1}{\theta}\pi_H \ln \frac{\pi_H}{\bar{\pi}_H} - \frac{1}{\theta}\pi_L \ln \frac{\pi_L}{\bar{\pi}_L}. \tag{9}$$

The investor understands the preferences and that the beliefs differ from the objective beliefs. The wishful thinking investor will choose subjective beliefs π_H and π_L by maximizing expected utility in (9) taking into account that $\pi_H + \pi_L = 1$. The optimization problem leads the investor to choose the following subjective beliefs:[7]

$$\pi_H = \frac{\bar{\pi}_H \exp(\theta q v_H)}{\bar{\pi}_H \exp(\theta q v_H) + \bar{\pi}_L \exp(\theta q v_L)}. \tag{10}$$

The investor chooses to distort beliefs towards states with positive surprises v_H so that $\pi_H > \bar{\pi}_H$ for $\bar{\pi}_H \in (0,1)$. The investor exhibits wishful thinking behavior by being over-optimistic about the high utility states. In other words, the wishful thinking investor obtains utility from the anticipation about future events. At the extremes, when the objective probability is either zero or one, then subjective probabilities are equal to the objectives probabilities and the investor is rational. A wishful thinking investor will not get any utility for dreaming about impossible events. As the cost of manipulating beliefs decreases (θ increases), the beliefs become even more distorted towards positive surprises. The same effect

[7]See the Appendix for derivations

appears the more shares q the investor is considering to buy, as q increases the subjective probability π_H deviates more from the objective probability $\bar{\pi}_H$ and thus more positive optimistic biased are investors.

We can observe how a wishful thinking investor distorts beliefs in a numerical example in Figure 8a. In this figure, we set the following parameters: $v_H = 3$, $v_L = 1$, $\theta = .5$ and $q = 1$. The solid line represents the beliefs of a wishful thinking investor given by (10). The dashed line represents the beliefs of a rational investor that uses the objective beliefs $\pi_H^{Rational} = \bar{\pi}_H$. The figure shows that the wishful thinking investor distorts beliefs towards positive surprises. Even when the probability of a positive surprise is less likely than a negative surprise $\bar{\pi}_H < 0.5$, the wishful thinking investor may distort beliefs so that $\pi_H > 0.5$. In words, even when the consensus is that there will be a negative surprise, the wishful thinking investor may think that a positive surprise is more likely (see for example when $\bar{\pi}_H = 0.4$, then $\pi_H > 0.5$). As the consensus probabilities get closer to the extremes, when events are almost certain, then wishful thinking investors resemble rational investors. In Figure 8b, we observe how beliefs get distorted as the number of shares q increases. As the stakes increase, there is an increase in the distortion of beliefs.

The wishful thinking investor will choose to purchase q units of the asset at price p when the expected utility in equation (9) with subjective beliefs given by (10) is positive $EU(\pi_H, \pi_L) \geq 0$, which happens when:

$$\bar{\pi}_H \geq \frac{\exp(\theta q p) - \exp(\theta q v_L)}{\exp(\theta q v_H) - \exp(\theta q v_L)} = \frac{1}{1 + \exp(\theta q v)} = \bar{\pi}_H^{cutoff}.$$

Thus, a wishful thinking investor will choose to purchase the q shares of an asset at price p when $\bar{\pi}_H \geq \bar{\pi}_H^{cutoff}$. Instead a rational investor with $\pi_H^{Rational} = \bar{\pi}_H$ would choose to purchase the q shares of an asset at price p when $\bar{\pi}_H \geq 0.5$. We can see that a wishful thinking investor would make the same choices as a rational investor only when it is infinitely costly to distort beliefs ($\theta = 0$). For any $\theta > 0$, the wishful thinking investor will have a lower cutoff to purchase the asset than a rational investor such that $\bar{\pi}_H^{cutoff} < 0.5$.

This simple wishful thinking model predicts that retail investors will display positive (negative) optimism when searching to buy (sell). Since it is well known that retail investors are more inclined to buy than sell (Barber and Odean, 2008), investors will display positive optimism. This is what our main finding about aggregate positive price pressure before earnings announcements conveys.

5.1. Empirical implications

Our model suggests that investors are more likely to engage in wishful thinking when it is easier to depart from objective beliefs, which occurs when the cost of deviating from those beliefs is low. This situation arises when information on a particular stock is well-covered by social media and easier for investors to interpret. For example, when a user scrolls down posts on social platforms, the user receives a more precise signal of investor sentiment when there are more posts with a "bullish" or "bearish" tag. More than 50% posts on StockTwits are not tagged with a sentiment.

We use post activity on StockTwits without sentiment (commonly attributed to "bots" activity) as a proxy for cross-sectional variation in noisy signals. For each stock-earnings announcement, we compute the fraction of unsigned posts without sentiment five days before earnings announcements as $1-(\#Bull+\#Bear)/\#Posts$, where $\#Posts$ is the total number of posts.

Figure 9 plots the BHAR five days before and after earnings announcements for stocks with low (bottom quintile) and high (top quintile) fractions of unsigned posts and for positive and negative earnings surprises. We further split the analysis for stocks with high abnormal post coverage in Panel A and low abnormal post coverage in Panel B. We select only small stocks because, as our research shows, social media's effect on price impact is more significant for small stocks. The figure shows that only stocks with a low fraction of unsigned posts and a high number of abnormal posts (Panel A) exhibit positive cumulative returns before earnings announcements. Stocks with a high fraction of unsigned posts have pre-announcement cumulative returns close to zero. Overall, these findings align with our wishful thinking model prediction that when it is easier (less costly) for investors to process information, they are more likely to trade according to their subjective beliefs.

6. Conclusion

We examine information production surrounding earnings announcements on leading investment social networks. In aggregate, information production on social media displays excessive positively skewed optimism about future outcomes on earnings announcements. Such biased optimism does not predict fundamentals on earnings announcements and leads to price run-ups before earnings announcements, thus, distorting prices from fundamentals before negative earnings announcements. We attribute our findings to individual investors being net buyers of attention-grabbing stocks (Barber and Odean, 2008), obtaining utility from their beliefs and interpreting information optimistically, i.e., wishful thinking (Caplin and Leahy, 2019).

Some users on such social media platforms might be sophisticated at forecasting fundamentals. However, in aggregate, our findings cast doubt on the wisdom of the crowd phenomenon from social media platforms in forecasting future fundamentals and having a beneficial role in price efficiency and investment-making decisions for retail traders.

References

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention-induced trading and returns: Evidence from robinhood users, *The Journal of Finance* 77, 3141–3190.

Barber, Brad M, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just how much do individual investors lose by trading?, *The Review of Financial Studies* 22, 609–632.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *The Journal of Finance* 55, 773–806.

Barber, Brad M, and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *The Review of Financial Studies* 21, 785–818.

Bartov, Eli, Lucile Faurel, and Partha S. Mohanram, 2018, Can twitter help predict firm-level earnings and stock returns?, *The Accounting Review* 93, 25–57.

Bradley, Daniel, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao, 2021, Place your bets? The market consequences of investment research on reddit's wallstreetbets, *Working paper*.

Caplin, Andrew, and John V Leahy, 2019, Wishful thinking, *NBER Working Paper*.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of crowds: The value of stock opinions transmitted through social media, *The Review of Financial Studies* 27, 1367–1403.

CNBC, 2021, Invest in you next gen survey finds behaviors of new investors are vastly different from experienced investors, https://www.cnbc.com/2021/08/23/invest-in-you-next-gen-survey-finds-behaviors-of-new-investors-are-vastly-different-from-experienced-investors.html.

Cookson, J. Anthony, Joseph E. Engelberg, and William Mullins, 2023, Echo chambers, *The Review of Financial Studies* 36, 450–500.

Cookson, J Anthony, Runjing Lu, William Mullins, and Marina Niessner, 2022, The social signal, *Working paper*.

Cookson, J. Anthony, and Marina Niessner, 2020, Why don't we agree? Evidence from a social network of investors, *The Journal of Finance* 75, 173–228.

———, and Christoph Schiller, 2022, Can social media inform corporate decisions? Evidence from merger withdrawals, *Working paper*.

Dim, Chukwuma, 2020, Should retail investors listen to social media analysts? Evidence from text-implied beliefs, *Working paper*.

Eaton, Gregory W, T Clifton Green, Brian Roseman, and Yanbin Wu, 2021, Zero-commission individual investors, high frequency traders, and stock market quality, *High Frequency Traders, and Stock Market Quality (January 2021)*.

Farrell, Michael, T Clifton Green, Russell Jame, and Stanimir Markov, 2022, The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145, 616–641.

Fisher, Adlai, Charles Martineau, and Jinfei Sheng, 2022, Macroeconomic attention and announcement risk premia, *The Review of Financial Studies* 35, 5057–5093.

Francis, Jennifer, and Donna Philbrick, 1993, Analysts' decisions as products of a multi-task environment, *Journal of Accounting Research* 31, 216–230.

Greenwood, Robin, Toomas Laarits, and Jeffrey Wurgler, 2022, Stock market stimulus, *Review of Financial Studies, forthcoming*.

Gregoire, Vincent, and Charles Martineau, 2022, How is earnings news transmitted to stock prices?, *Journal of Accounting Research* 60, 261–297.

Hu, Danqi, Charles M Jones, Valerie Zhang, and Xiaoyan Zhang, 2021, The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery, *Working paper*.

Ivković, Zoran, and Narasimhan Jegadeesh, 2004, The timing and value of forecast and recommendation revisions, *Journal of Financial Economics* 73, 433–463.

Jackson, Andrew R., 2005, Trade generation, reputation, and sell-side analysts, *The Journal of Finance* 60, 673–717.

Jiao, Peiran, Andre Veiga, and Ansgar Walther, 2020, Social media, news media and the stock market, *Journal of Economic Behavior & Organization* 176, 63–90.

Kahneman, Daniel, Olivier Sibony, and CR Sunstein, 2022, *Noise* (HarperCollins UK).

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing, 2011, How social influence can undermine the wisdom of crowd effect, *Proceedings of the national academy of sciences* 108, 9020–9025.

Martineau, Charles, 2022, Rest in peace post-earnings announcement drift, *Critical Finance Review* 11, 613–646.

Michaely, Roni, and Kent L. Womack, 1999, Conflict of interest and the credibility of underwriter analyst recommendations, *The Review of Financial Studies* 12, 653–686.

Nikolov, Dimitar, Diego FM Oliveira, Alessandro Flammini, and Filippo Menczer, 2015, Measuring online social bubbles, *PeerJ Computer Science* 1, e38.

Ott, Matt, 2022, Social media influencers are charged with feeding followers 'a steady diet of misinformation' in a pump and dump stock scheme that netted $100 million, https://fortune.com/2022/12/14/social-media-influencers-pump-and-dump-stock-scheme/.

Wasik, John F., 2013, 'pump and dump' schemes resurface in social media, https://www.forbes.com/sites/johnwasik/2013/06/28/pump-and-dump-schemes-resurface-in-social-media/?sh=7d817d206474.

Xiaoyan, Charles M Jones Donghui Shi, and Zhang Xinran Zhang, 2022, Understanding retail investors: Evidence from china, *Working paper*.

Figure 1. Social Media Information Production

This figure shows the monthly number of stock-specific posts on StockTwits, WallStreetBets, and Seeking Alpha, in Panels A to C, respectively. The sample period is from January 1, 2018, to December 31, 2021.

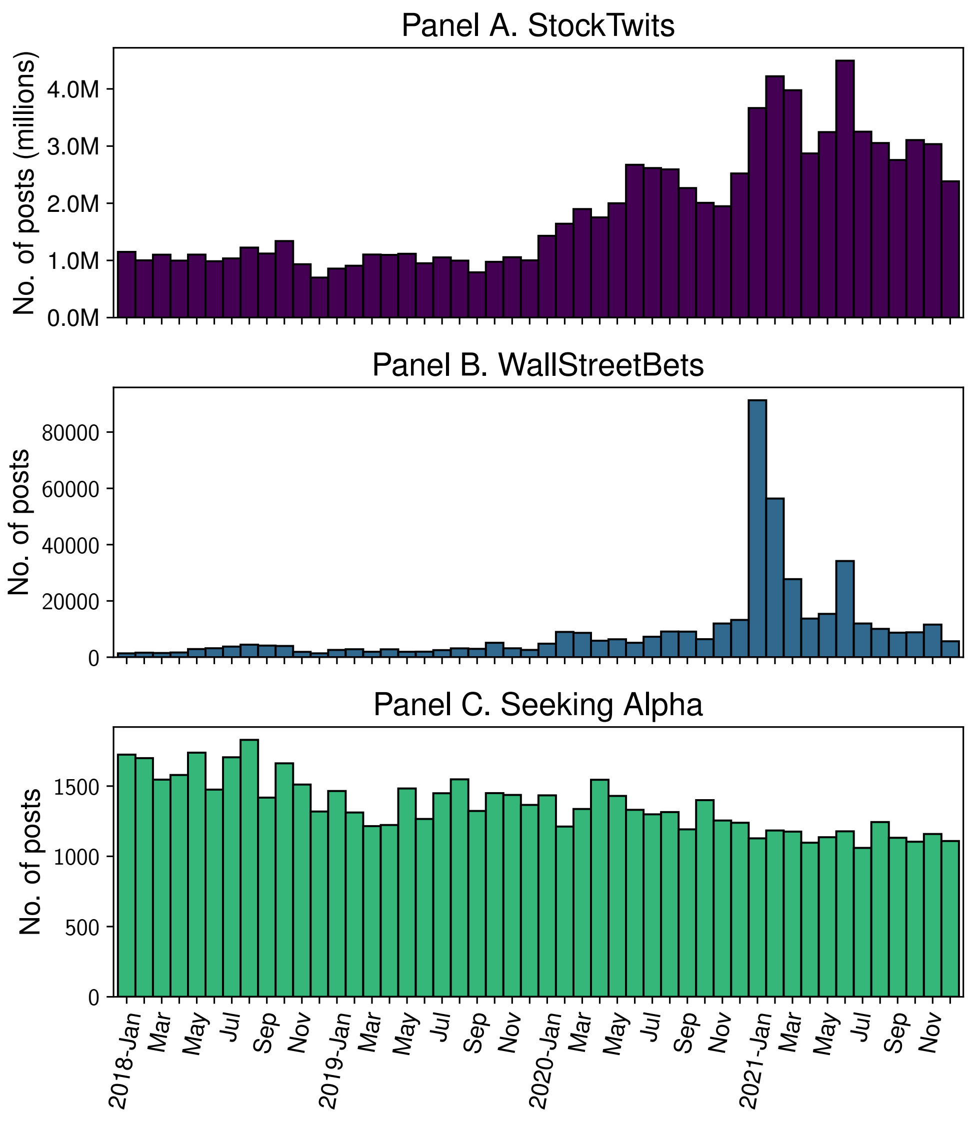
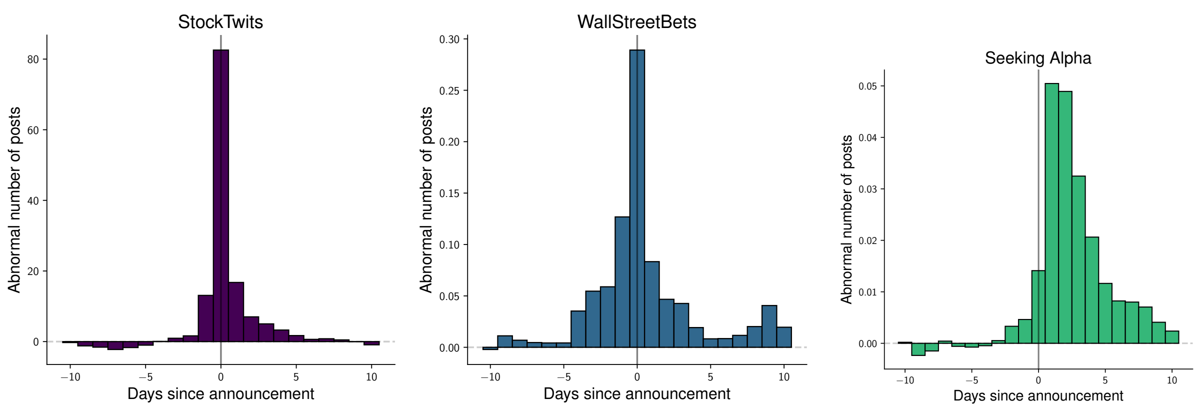


Figure 2. Social Media, Analyst, and Newswire Information Production Around Earnings Announcements

This figure shows the mean of the abnormal number of posts for StockTwits, WallStreetBets, and Seeking Alpha in Panel A. Panel B shows the abnormal number of analyst recommendations and newswire articles in Ravenpack. We define the number of abnormal posts for a stock as the difference between the number of posts on day t minus the average number of posts from $t = -30$ to $= -11$. The sample period is from January 1, 2018, to December 31, 2021.

Panel A: Social media platforms



Panel B: Traditional news

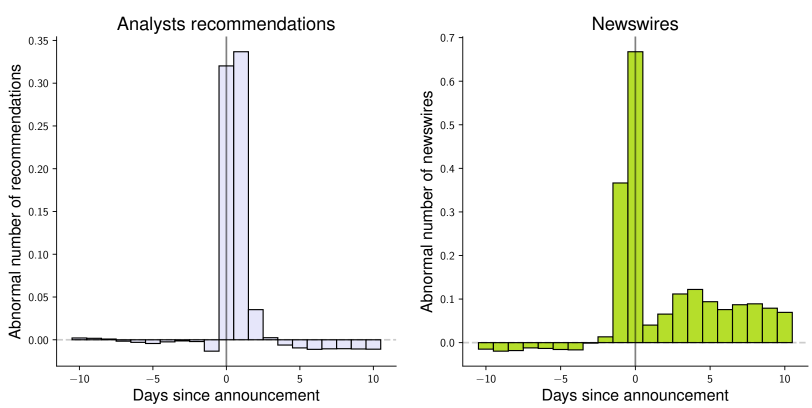


Figure 3. Sample Fraction of Stock-Earnings Announcement Coverage by News Source

This figure shows the fraction of stock-earnings announcement observations, for every quarter, with at least one post/news five days before earnings announcements for StockTwits, WallStreetBets, Seeking Alpha, analysts' recommendations, and newswire articles in Ravenpack. The sample period is from January 1, 2018, to December 31, 2021.

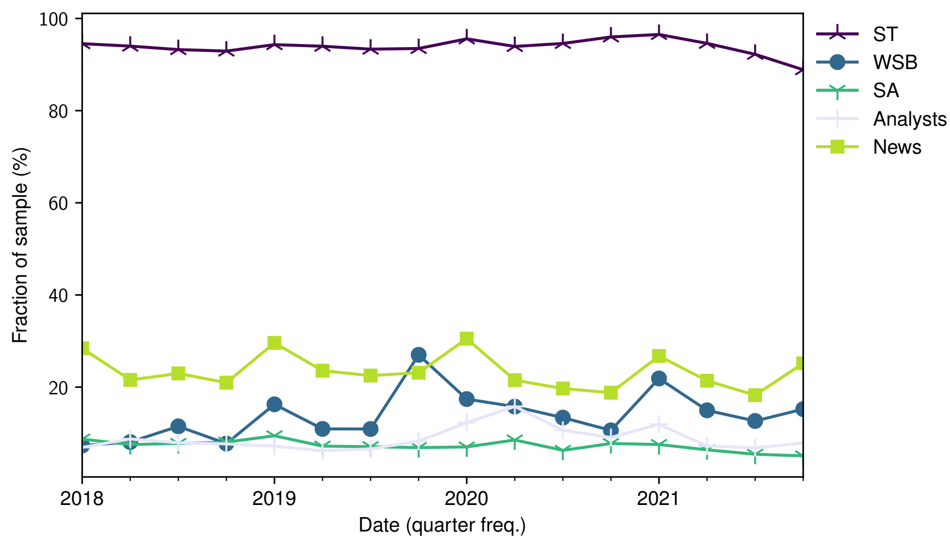


Figure 4. Positive Skewness in Sentiment

This figure shows the fraction of stock-earnings announcements observation with positive sentiment posts ratio $\leq 20\%$, between 21 and 39%, 40 and 59%, 60 and 79%, and $\geq 80\%$ five days before earnings announcements. The ratio is computed as the fraction of positive sentiment posts divided by the sum of positive and negative sentiment posts. The sample period is from January 1, 2018, to December 31, 2021.

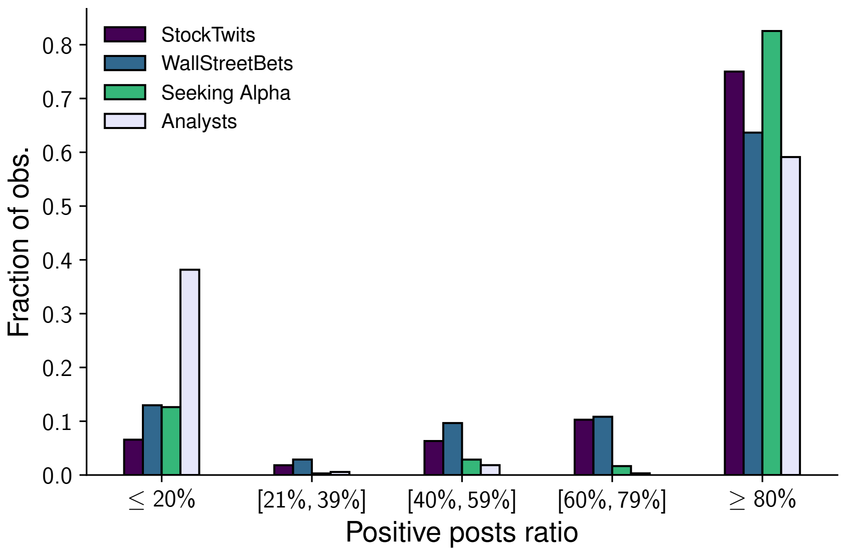
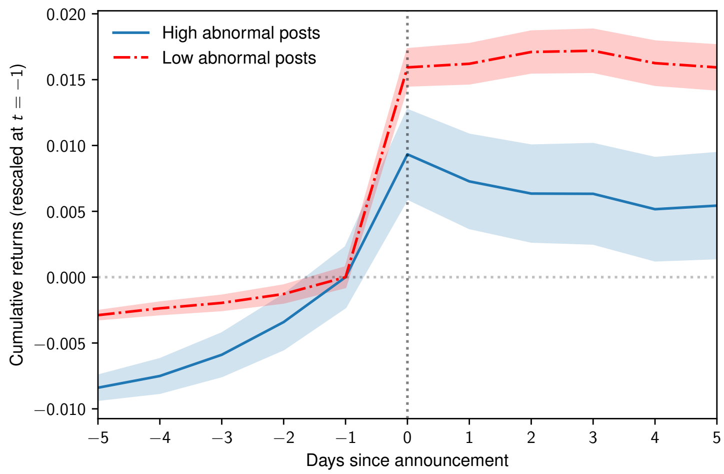


Figure 5. BHAR Around Earnings Announcement

This figure shows the buy-and-hold abnormal returns (BHAR) five days before and after earnings announcements for stocks with high and low social media abnormal information production. BHAR for positive and negative earnings surprises are plotted in Panels A and B, respectively. High (low) abnormal posts are defined asstocks that have an average of abnormal information production greater (less or equal to) than zero in the window $t = [-5, -1]$, for either StockTwits, WallStreetBets, or Seeking Alpha. The plots are rescaled such that lines cut the y-axis at $t = -1$. The shaded area corresponds to the 95% confidence interval. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. Positive surprise



Panel B. Negative surprise

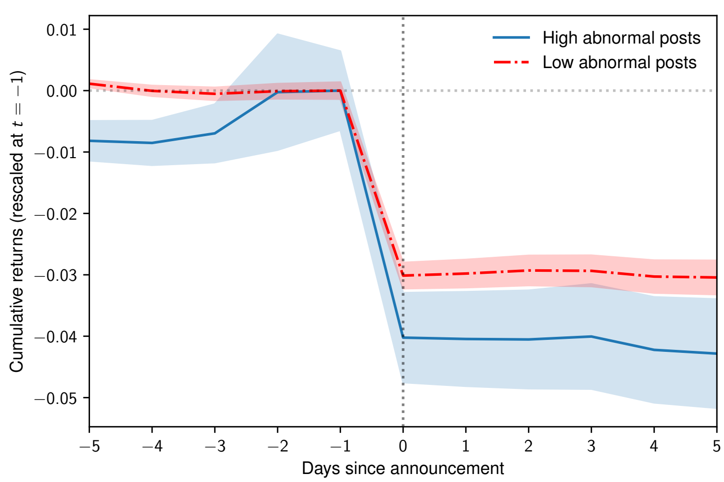
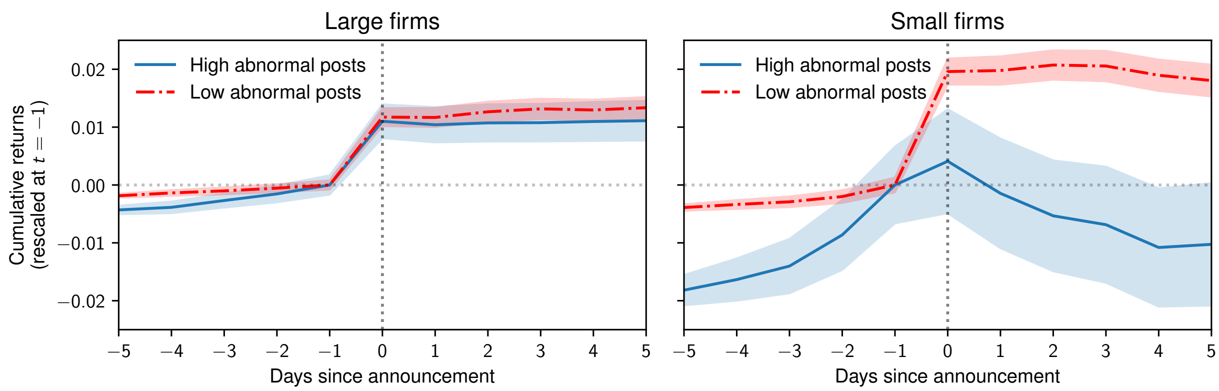


Figure 6. BHAR Around Earnings Announcement for Large and Small Firms

This figure shows the buy-and-hold abnormal returns (BHAR) five days before and after earnings announcements for stocks with high and low social media information production and for large and small firms. BHAR for positive and negative earnings surprises are plotted in Panels A and B, respectively. Large (small) firms are defined as firms with market capitalization belonging to the top three (bottom two) NYSE market capitalization quintiles. High (low) abnormal posts are defined asstocks that have an average of abnormal information production greater (less or equal to) than zero in the window $t = [-5, -1]$, for either StockTwits, WallStreetBets, or Seeking Alpha. The plots are rescaled such that lines cut the y-axis at $t = -1$. The shaded area corresponds to the 95% confidence interval. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. Positive surprises



Panel B. Negative surprises

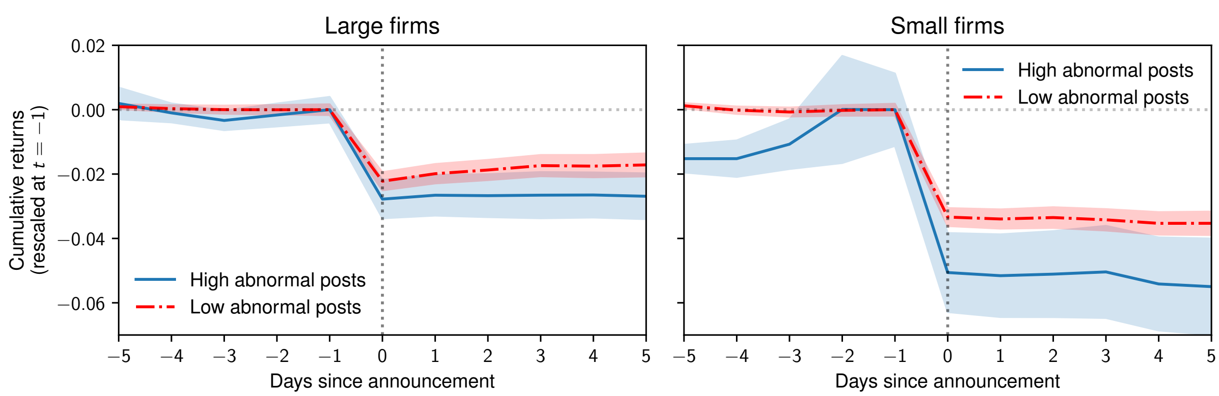
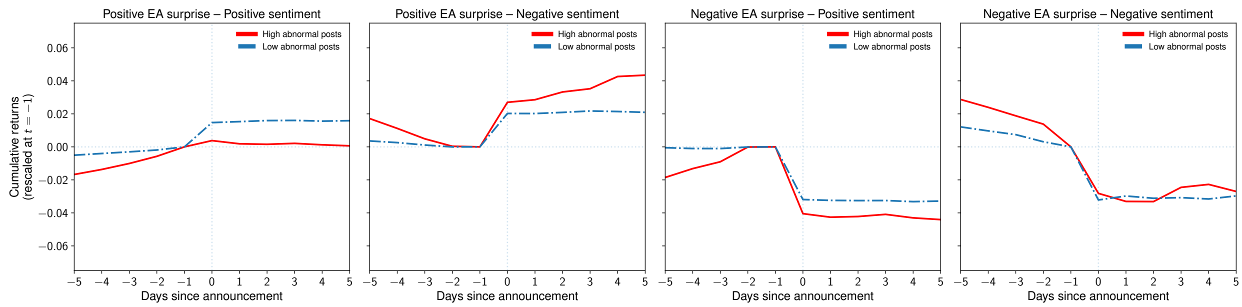


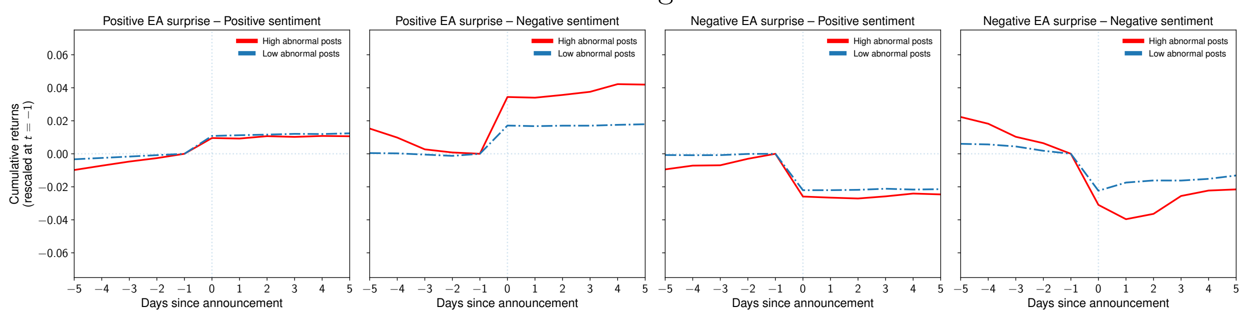
Figure 7. BHAR Around Earnings Announcement for High and Low Sentiment Stocks

This figure shows the buy-and-hold abnormal returns (BHAR) five days before and after earnings announcements for stocks with a high positive sentiment ratio (≥ 0.80) and low positive sentiment ratio (< 0.20, i.e., negative sentiment) for positive and negative earnings surprises. The BHAR are presented for the full sample, large, and small firms in Panels A to C, respectively. The sample is restricted to stocks with abnormal posts on StockTwits greater than zero. Large (small) firms are defined as firms with market capitalization belonging to the top three (bottom two) NYSE market capitalization quintiles. The plots are rescaled such that lines cut the y-axis at $t = -1$. The sample period is from January 1, 2013, to December 31, 2021.

Panel A. Full sample



Panel B. Large firms



Panel C. Small firms

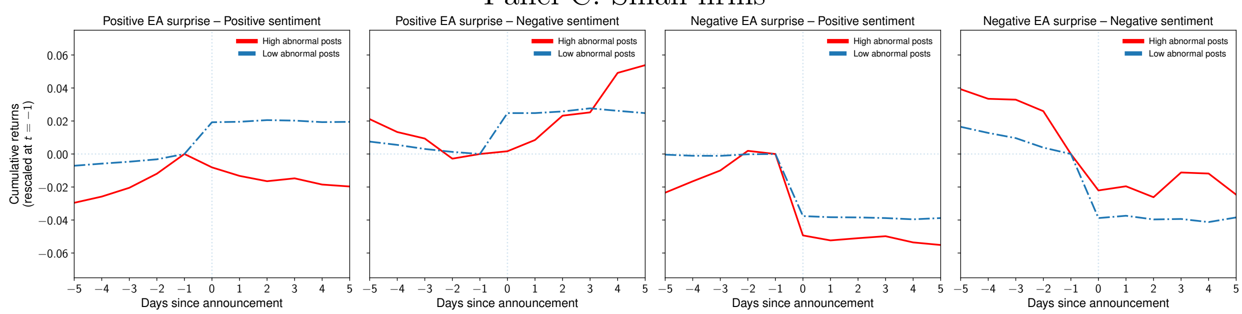
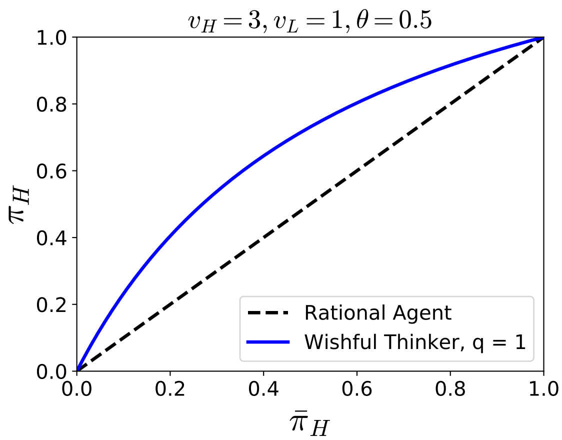
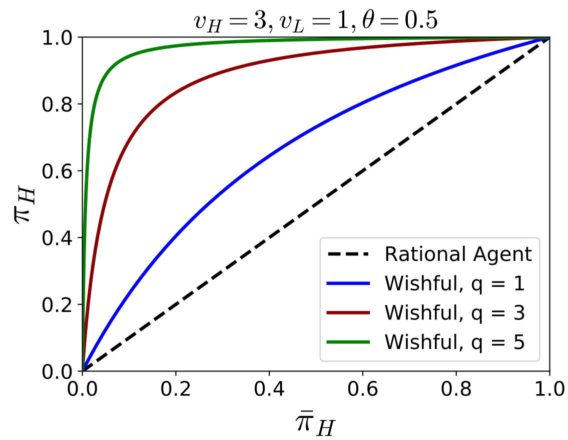


Figure 8. Plots of π_H for Wishful Thinking vs Rational Investors

Dashed black line represents a rational agent. Solid lines represent wishful thinking investors for different quantities q.



(a)

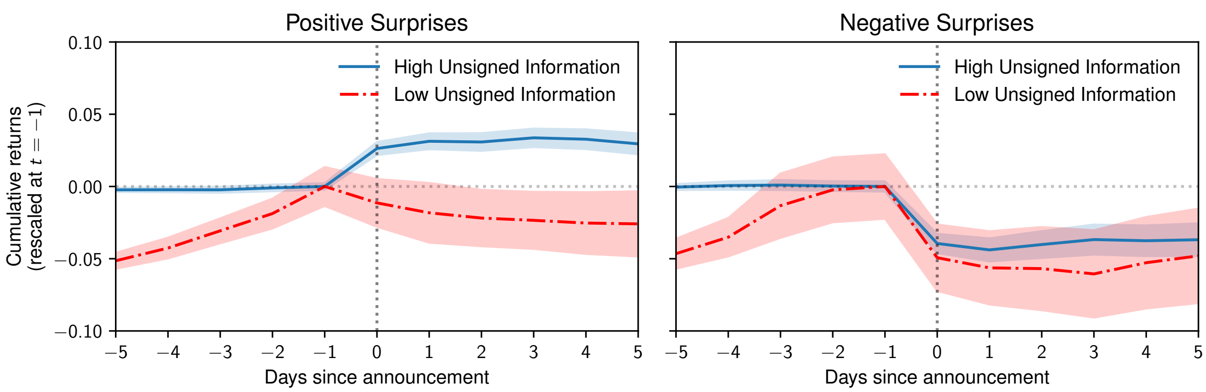


(b)

Figure 9. BHAR Around Earnings Announcement for Small Firms with Unsigned Posts

This figure shows the buy-and-hold abnormal returns (BHAR) five days before and after earnings announcements for small firms only (bottom two NYSE breakpoint quintiles). Panels A and B show the BHAR for stocks with high and low abnormal posts on StockTwits, respectively. The solid (dashed) line corresponds to the top quintile (bottom quintile) fraction of unsigned posts five days before announcements. The fraction of unsigned posts is computed as $1 - (\#Bull + \#Bear)/\#Posts$, where $\#Posts$ is the total number of posts. The plots are rescaled such that lines cut the y-axis at $t = -1$. The shaded area corresponds to the 95% confidence interval. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. High abnormal posts



Panel B. Low abnormal posts

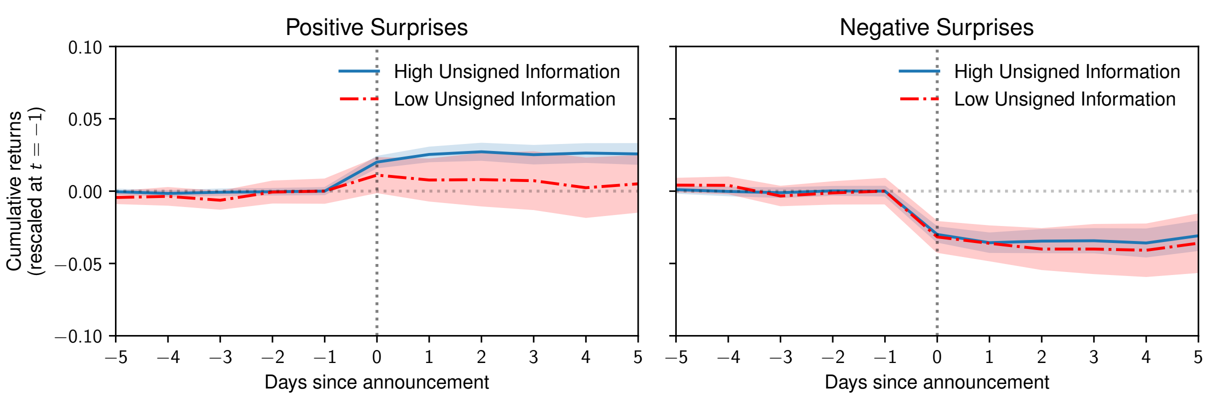


Table 1
Sample of Earnings Announcements by Social Media Platforms and Analysts

This table reports the number of stock-earnings announcement observations and the number of posts by social media platforms by NYSE market capitalization breakpoints quintiles in Panel A and for positive, neutral, and negative surprises in Panel B. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. Summary statistics by NYSE market capitalization breakpoints

	StockTwits				WallStreetBets				Seeking Alpha				Analysts			
	Stock-EA obs.		Posts		Stock-EA obs.		Posts		Stock-EA obs.		Posts		Stock-EA obs.		Rec.	
	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)
1 (small)	11,817	31.3	1,224,865	26.7	757	13.6	1,598	5.6	347	11.9	377	8.8	336	9.5	419	8.1
2	7,873	20.9	540,853	11.8	733	13.2	1,958	6.9	303	10.4	336	7.8	429	12.1	576	11.1
3	6,560	17.4	629,506	13.7	796	14.3	3,046	10.7	322	11.1	367	8.6	591	16.7	780	15.0
4	5,796	15.4	511,526	11.2	1,069	19.2	3,803	13.4	489	16.8	565	13.2	716	20.3	989	19.1
5 (large)	5,710	15.1	1,680,222	36.6	2,214	39.8	18,033	63.4	1,447	49.8	2,637	61.6	1,462	41.4	2,420	46.7
Total	37,756		4,586,972		5,569		28,438		2,908		4,282		3,534		5,184	

Panel B. Summary statistics by earnings surprises

	StockTwits				WallStreetBets				Seeking Alpha				Analysts			
	Stock-EA obs.		Posts		Stock-EA obs.		Posts		Stock-EA obs.		Posts		Stock-EA obs.		Rec.	
	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)	N. obs.	(%)
Positive	24,778	65.6	2,701,572	58.9	4,082	73.3	21,054	74.0	2,093	72.0	3,070	71.7	2,594	73.4	3,861	74.5
Neutral	1,583	4.2	205,984	4.5	176	3.2	689	2.4	106	3.6	172	4.0	107	3.0	131	2.5
Negative	11,395	30.2	1,679,416	36.6	1,311	23.5	6,695	23.5	709	24.4	1,040	24.3	833	23.6	1,192	23.0
Total	37,756		4,586,972		5,569		28,438		2,908		4,282		3,534		5,184	

Table 2
Forecasting Earnings Surprise with Sentiment

This table reports the coefficients of the following regression estimated for each news source:

$$Surprise_{i,\tau} = \beta_1 Sent^p_{i,\tau} + \beta_2 AbnPost^p_{i,\tau} + \beta_3 Sent^p_{i,\tau} \times AbnPost^p_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Surprise$ is the earnings surprise of earnings announcement τ for stock i. $Sent^p$ corresponds to the average sentiment five days before earnings announcement. $AbnPost^p$ is the average abnormal information production five days before earning announcements for various news sources p = {StockTwits, WallStreetBets, Seeking Alpha, and Analysts}. Robust standard errors clustered by stock and year-quarter are presented in parentheses and *, ** , and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. Full-sample

	StockTwits		WallStreetBets		Seeking Alpha		Analyst	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sent	0.024	0.011	0.019	-0.025	0.038	-0.380	0.032	0.039
	(0.022)	(0.032)	(0.047)	(0.068)	(0.171)	(0.430)	(0.049)	(0.064)
AbnPost		0.021		0.012		-0.185		-0.005
		(0.037)		(0.026)		(0.116)		(0.008)
Sent×AbnPost		-0.041		0.037		0.100		-0.003
		(0.058)		(0.028)		(0.118)		(0.011)
N	27,055	27,055	5,569	5,569	1,988	1,988	3,534	3,534
R^2	0.0000	0.0001	0.0001	0.0014	0.0002	0.0107	0.0004	0.0005

Panel B. Large firms

	StockTwits		WallStreetBets		Seeking Alpha		Analyst	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sent	0.015	0.038**	0.021	-0.045	0.012	-0.032	0.041	0.061
	(0.016)	(0.019)	(0.028)	(0.033)	(0.075)	(0.177)	(0.030)	(0.042)
AbnPost		-0.080*		-0.015		-0.099		-0.005
		(0.045)		(0.026)		(0.063)		(0.007)
Sent×AbnPost		0.096*		0.068***		0.013		-0.010
		(0.049)		(0.026)		(0.038)		(0.010)
N	13,931	13,931	4,079	4,079	1,537	1,537	2,769	2,769
R^2	0.0001	0.0018	0.0003	0.0054	0.0001	0.0089	0.0011	0.0016

Panel C. Small firms

	StockTwits		WallStreetBets		Seeking Alpha		Analyst	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sent	0.045	-0.116	-0.069	-0.048	-0.005	-2.541	0.226	0.044
	(0.050)	(0.105)	(0.258)	(0.377)	(1.120)	(3.152)	(0.364)	(0.533)
AbnPost		0.312**		0.070		-0.609		-0.018
		(0.131)		(0.091)		(0.509)		(0.071)
Sent×AbnPost		-0.375**		-0.010		0.392		0.037
		(0.171)		(0.098)		(0.529)		(0.074)
N	13,124	13,124	1,490	1,490	451	451	765	765
R^2	0.0001	0.0009	0.0003	0.0036	0.0000	0.0493	0.0069	0.0120

Table 3
Retail Trading and Social Media Information Production

This table reports the coefficients of the following regression

$$\Delta Retail\ vlm_{i,\tau} = \sum_j \beta_p AbnPost^p_{i,\tau} + \delta_1 |Surprise|_{i,\tau} + \delta_2 |News\ sent|_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel A,}$$

$$Retail\ OI_{i,\tau} = \sum_j \beta_p AbnPost^p_{i,\tau} + \gamma_1 Surprise_{i,\tau} + \gamma_2 News\ sent_{i,\tau} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel B,}$$

$\Delta Retail\ vlm$ corresponds to the change in retail volume, and retail option volume from $t = [-60, -6]$ to $t = [-5, -1]$ of earning announcement τ for stock i, in columns (1)-(4) and columns (5)-(8), respectively. $Retail\ OI$ is the average daily retail volume order imbalance and retail option order imbalance for $t = [-5, -1]$, in columns (1)-(4) and columns (5)-(8), respectively. $AbnPost^p$ is the average abnormal information production five days before announcement τ of stock i for social media platform $p = \{\text{ST, WSB, SA}\}$. $Surprise$ ($|Surprise|$) is the earnings announcement (absolute) surprise. $News\ sent$ ($|News\ sent|$) is the daily average (absolute average) news sentiment from RavenPack five days before announcement. All independent variables except the log-transformed variables are standardized. Robust standard errors clustered by stock and year-quarter are presented in parentheses and *, ** , and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 1, 2018, to December 31, 2021.

Panel A. Retail volume

	Dependent variable:							
	ΔRetail volume				ΔRetail option volume			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnPost^{ST}	0.132***			0.133***	0.078***			0.069***
	(0.011)			(0.011)	(0.009)			(0.007)
AbnPost^{WSB}		0.036***		-0.006		0.040**		0.018
		(0.008)		(0.007)		(0.016)		(0.017)
AbnPost^{SA}			0.011***	0.004*			0.013***	0.009***
			(0.002)	(0.002)			(0.003)	(0.003)
|Surprise|	0.002	0.136	0.127	-0.003	-0.233	-0.151	-0.162	-0.225
	(0.317)	(0.326)	(0.329)	(0.317)	(0.436)	(0.431)	(0.430)	(0.432)
$\|\text{News sent}\|_{[-5,-1]}$	0.267*	0.302*	0.286*	0.264*	0.217*	0.247**	0.229*	0.228**
	(0.159)	(0.168)	(0.169)	(0.159)	(0.115)	(0.121)	(0.123)	(0.116)
N	40,254	40,254	40,254	40,254	40,254	40,254	40,254	40,254
R^2	0.0206	0.0022	0.0006	0.0206	0.0038	0.0013	0.0003	0.0041

Panel B. Retail order imbalance

	Dependent variable:							
	Retail volume OI				Retail option volume OI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnPost^{ST}	0.004***			0.004***	0.002			0.000
	(0.001)			(0.001)	(0.002)			(0.002)
AbnPost^{WSB}		0.001*		0.000		0.003*		0.003*
		(0.001)		(0.001)		(0.002)		(0.002)
AbnPost^{SA}			-0.000	-0.000			0.002	0.002
			(0.001)	(0.001)			(0.001)	(0.001)
Surprise	0.110***	0.110***	0.110***	0.110***	0.042	0.041	0.042	0.042
	(0.040)	(0.040)	(0.040)	(0.040)	(0.059)	(0.059)	(0.059)	(0.059)
$\text{News sent}_{[-5,-1]}$	0.004	0.005	0.005	0.004	-0.024	-0.022	-0.023	-0.022
	(0.017)	(0.017)	(0.017)	(0.017)	(0.038)	(0.037)	(0.038)	(0.037)
N	39,259	39,259	39,259	39,259	40,254	40,254	40,254	40,254
R^2	0.0004	0.0002	0.0002	0.0004	0.0000	0.0001	0.0001	0.0001

Table 4
Pre-Earnings Announcement BHAR and Social Media Information Production

This table reports the coefficients of the following regression:

$$BHAR[-5,-1]_{i,\tau} = \beta_1 \mathbb{1}^{ST}_{i,\tau} + \beta_2 \mathbb{1}^{WSB}_{i,\tau} + \beta_3 \mathbb{1}^{SA}_{i,\tau} + \beta_4 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_5 \mathbb{1}^{WSB}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_6 \mathbb{1}^{SA}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_7 \mathbb{1}^{Small}_{i} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $BHAR[-5,-1]$ is the buy-and-hold abnormal return five days before earnings announcement τ for stock i. $\mathbb{1}^p$ corresponds to indicator variables equal to one if the average abnormal number of posts for stock i five days before announcement τ is positive, zero otherwise, for social media platform $p = \{\text{ST, WSB, SA}\}$. $\mathbb{1}^{Small}$ is an indicator variable equal to one if the stock-earnings announcement i belongs to the bottom two NYSE market capitalization quintiles, zero otherwise. The control variables include $\mathbb{1}^{Ana}$ ($\mathbb{1}^{News}$), which is an indicator variable equal to one if the number of abnormal analyst recommendations (newswire article) of stock i five days before earnings announcement, is positive, zero otherwise. *Surprise* is the earnings surprise. *News sent* is the average news sentiment in RavenPack five days before earnings announcements. α_i and α_t correspond to stock and year-quarter fixed effects, respectively. The results are reported for earnings announcements with upcoming positive earnings surprises in columns (1)-(4) and negative surprises in columns (5)-(8). Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, ** , and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 1, 2018, to December 31, 2021.

	Surprise> 0				Surprise< 0			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}^{ST}$	0.020***	0.019***	0.008**	0.008**	0.032***	0.031**	0.019***	0.019***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.012)	(0.012)	(0.007)	(0.007)
$\mathbb{1}^{WSB}$		0.004*	0.003	0.003		0.008	0.000	0.001
		(0.002)	(0.002)	(0.002)		(0.008)	(0.004)	(0.004)
$\mathbb{1}^{SA}$		0.001	-0.003	-0.003		-0.011*	-0.003	-0.002
		(0.003)	(0.003)	(0.003)		(0.005)	(0.004)	(0.004)
$\mathbb{1}^{ST} \times \mathbb{1}^{Small}$			0.027***	0.025***			0.015	0.012
			(0.009)	(0.009)			(0.013)	(0.012)
$\mathbb{1}^{WSB} \times \mathbb{1}^{Small}$			0.004	0.004			0.016	0.014
			(0.004)	(0.004)			(0.016)	(0.015)
$\mathbb{1}^{SA} \times \mathbb{1}^{Small}$			0.021*	0.020*			-0.018	-0.020
			(0.012)	(0.011)			(0.014)	(0.015)
$\mathbb{1}^{Small}$			-0.004*	-0.005**			-0.004	-0.003
			(0.002)	(0.002)			(0.006)	(0.007)
$\mathbb{1}^{Ana}$				0.001				0.003
				(0.002)				(0.004)
$\mathbb{1}^{Ana} \times \mathbb{1}^{Small}$				-0.001				-0.010
				(0.006)				(0.018)
$\mathbb{1}^{News}$				0.847				-0.367
				(0.691)				(0.741)
$\mathbb{1}^{News} \times \mathbb{1}^{Small}$				2.408				13.426
				(4.183)				(9.780)
Surprise				0.389***				-0.171
				(0.089)				(0.114)
News sent				0.096***				0.249***
				(0.019)				(0.057)
N	26,195	26,195	26,195	26,195	12,347	12,347	12,347	12,347
R^2	0.0042	0.0045	0.0079	0.0151	0.0055	0.0064	0.0077	0.0168

Table 5
BHAR Prior to Earnings Announcements
Conditioned on StockTwits' Abnormal Coverage and Sentiment

This table reports the average buy-and-hold abnormal returns five days before earnings announcements for positive (+Surp) and negative (−Surp) earnings surprise and by StockTwits' sentiment. +Sent (−Sent) corresponds to average daily posts with positive ratios greater than 0.80 (0.20) for five days before earnings announcement. The numbers in parentheses correspond to the t-statistic computed using clustered standard errors at the stock and year-quarter. The sample period is from January 1, 2013, to December 31, 2021.

	Full sample				Large stocks				Small stocks			
	+Surp		−Surp		+Surp		−Surp		+Surp		−Surp	
	+Sent	−Sent	+Sent	−Sent	+Sent	−Sent	+Sent	−Sent	+Sent	−Sent	+Sent	−Sent
Abn posts	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
High	0.019	-0.015	0.023	-0.026	0.012	-0.012	0.012	-0.023	0.029	-0.022	0.030	-0.032
Low	0.006	-0.005	0.000	-0.013	0.005	-0.001	0.001	-0.007	0.007	-0.009	-0.001	-0.017
Diff	0.013	-0.010	0.023	-0.013	0.007	-0.011	0.011	-0.016	0.022	-0.014	0.030	-0.015
T-stat	[7.500]	[-4.194]	[4.099]	[-2.774]	[5.864]	[-3.472]	[3.274]	[-2.598]	[5.427]	[-2.397]	[3.670]	[-1.821]

Table 6
Pre-Earnings Announcement BHAR and Social Media Information Production around US Government Stimulus

This table reports the coefficients of the following regression:

$$BHAR[-5,-1]_{i,\tau} = \beta_1 \mathbb{1}^{ST}_{i,\tau} + \beta_2 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_3 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Stim}_{\tau} + \beta_4 \mathbb{1}^{ST}_{i,\tau} \times \mathbb{1}^{Small}_{i} \times \mathbb{1}^{Stim}_{\tau} + \beta_5 \mathbb{1}^{Small}_{i} + \beta_6 \mathbb{1}^{Stim}_{\tau} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $BHAR[-5,-1]$ is the buy-and-hold abnormal return five days before earnings announcement τ for stock i. $\mathbb{1}^{ST}$ is an indicator variables equal to one if the abnormal information production for stock i five days before earnings announcement τ on StockTwits, is positive, zero otherwise. $\mathbb{1}^{Small}$ is an indicator variable equal to one if the stock-earnings announcement i belongs to the bottom two NYSE market capitalization quintiles, zero otherwise. $\mathbb{1}^{Stim}$ is an indicator variable equal to one for earnings announcement τ occurring during 2020-Q2, and 2021-Q1, and 2021-Q2, zero otherwise. These quarters corresponds to the stimulus check arrivals. The control variables are the same as in Table 4. α_i and α_t correspond to firm and year-quarter fixed effects, respectively. The results are reported for the full sample in columns (1)-(2), and for earnings announcements with upcoming positive earnings surprises in columns (3)-(4) and negative surprises in columns (5)-(6). Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, ** , and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 1, 2018, to December 31, 2021.

	Full sample		Surprise>0		Surprise<0	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}^{ST}$	0.009***	0.009***	0.008**	0.008**	0.019***	0.020***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)
$\mathbb{1}^{ST} \times \mathbb{1}^{Small}$	0.024***	0.022***	0.024***	0.021***	0.011	0.006
	(0.006)	(0.006)	(0.007)	(0.007)	(0.013)	(0.013)
$\mathbb{1}^{ST} \times \mathbb{1}^{Stim}$	-0.006	-0.006	0.001	0.001	-0.011	-0.011
	(0.003)	(0.003)	(0.002)	(0.002)	(0.007)	(0.007)
$\mathbb{1}^{ST} \times \mathbb{1}^{Small} \times \mathbb{1}^{Stim}$	0.053**	0.053**	0.039*	0.039*	0.048	0.051
	(0.018)	(0.020)	(0.013)	(0.014)	(0.051)	(0.064)
$\mathbb{1}^{Small}$	-0.006*	-0.006*	-0.004*	-0.005**	-0.008	-0.007
	(0.010)	(0.011)	(0.011)	(0.011)	(0.019)	(0.019)
$\mathbb{1}^{Stim}$	0.024	0.023	0.016	0.018	0.036	0.037
	(0.025)	(0.025)	(0.021)	(0.020)	(0.033)	(0.032)
N	40,253	40,253	26,195	26,195	12,346	12,346
R^2	0.0107	0.0154	0.0088	0.0159	0.0096	0.0189
Yr-Qtr & Firm F.E.	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y

Table 7
Any Evidence of a Systematic Pump-and-Dump Scheme?

This table reports the coefficients of the following regression:

$$\begin{aligned} BHAR[1,5]_{i,t} = & \beta_1 BHAR[-5,-1]_{i,\tau} + \beta_2 BHAR[-5,-1]_{i,\tau} \times \mathbb{1}^{AbnPost}_{i,\tau} + \beta_3 BHAR[-5,-1]_{i,\tau} \times \mathbb{1}^{Small}_{i} + \\ & \beta_4 BHAR[-5,-1]_{i,t} \times \mathbb{1}^{AbnPost}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_5 \mathbb{1}^{AbnPost}_{i,\tau} \times \mathbb{1}^{Small}_{i} + \beta_6 \mathbb{1}^{AbnPost}_{i,\tau} + \\ & \beta_7 \mathbb{1}^{Small}_{i} + \alpha_i + \alpha_t + \varepsilon_{i,t} \end{aligned}$$

BHAR[1,5] and BHAR[-5,-1] correspond to the buy-and-hold abnormal returns around earnings announcement date for $t = [1,5]$ and $t = [-5,-1]$, respectively. $\mathbb{1}^{AbnPost}$ in an indicator variable equal to one if the abnormal information production on one of the social media p is positive, zero otherwise. $\mathbb{1}^{Small}$ is an indicator variable equal to one if the stock-earnings announcement i belongs to the bottom two NYSE market capitalization quintiles, zero otherwise. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, ** , and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 1, 2018, to December 31, 2021.

	StockTwits			WallStreetBets			Seeking Alpha		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
BHAR[-5,-1]	-0.011	-0.016	0.017	-0.011	-0.005	0.020	-0.011	-0.005	-0.007
	(0.011)	(0.021)	(0.034)	(0.011)	(0.013)	(0.032)	(0.011)	(0.013)	(0.034)
BHAR[-5,-1]$\times \mathbb{1}^{AbnPost}$		0.009	-0.020		-0.019	-0.026		-0.027	0.055
		(0.002)	(0.001)		(0.002)	(0.002)		(0.002)	(0.002)
BHAR[-5,-1]$\times \mathbb{1}^{Small}$			-0.041			-0.029			0.002
			(0.004)			(0.003)			(0.003)
BHAR[-5,-1] $\times \mathbb{1}^{AbnPost} \times \mathbb{1}^{Small}$			0.038			0.008			-0.112
		(0.031)	(0.032)		(0.020)	(0.031)		(0.027)	(0.058)
$\mathbb{1}^{AbnPost} \times \mathbb{1}^{Small}$			-0.007*			-0.009***			-0.008
			(0.042)			(0.040)			(0.041)
$\mathbb{1}^{AbnPost}$		-0.001	0.002*		-0.000	0.004*		-0.003	-0.001
			(0.004)			(0.003)			(0.005)
$\mathbb{1}^{Small}$			0.016***			0.016***			0.015***
			(0.027)			(0.038)			(0.082)
N	40,236	40,236	40,236	40,236	40,236	40,236	40,236	40,236	40,236
R^2	0.0001	0.0001	0.0016	0.0001	0.0002	0.0019	0.0001	0.0003	0.0022

Appendix

A. Sentiment classication for WallStreetBets

To train our model, we use Stocktwits posts since authors on both platforms use comparable language to communicate. In addition, both platforms are free, they do not use any editorial board to review the posts, and they allow users greater anonymity than Seeking Alpha. This allows us to consider the expressions described in Appendix B of Bradley, Hanousek Jr, Jame, and Xiao (2021). Specifically, we use a subsample of 100,000 Stocktwits stock-related tweets and their associated sentiments (50,000 bullish posts and 50,000 bearish posts). Similar to Dim (2020), we preprocess all posts of the subsample to reduce the vocabulary and vectorize the text corpus into unigrams and bigrams, eliminating all that appeared less than 1% and normalizing the times they appeared on the text using a Term Frequency - Inverse Document Frequency (tf-idf) algorithm. We later calculate the parameters of the Linear Support Vector Classifier (SVC) and test its accuracy by taking a test set of 30% of our subsample data. The optimal hyperparameter of the SVC model is c=1.7, and achieves an accuracy score of 71% on the test data. With this chosen parameter, we use the trained SVC linear model to determine the sentiment of every WallStreetBets post. We consider posts on WallStreetBets of all categories except the ones posted by Moderators (tagged as "MOD"). All the posts from this platform were downloaded using the Reddit API.

B. Retail order imbalance measures

We calculate retail trading volume on the equity market, following Eaton, Green, Roseman, and Wu (2021). Using the number of equity trades and volume initiated by retail traders from TAQ, the authors propose to flag trade as retail when it is executed at a price improvement. We compute retail order imbalance using the volume derivated from buy and sell trades in the equity market as:

$$Retail\ OI_{i,t} = \frac{Buy_{i,t} - Sell_{i,t}}{Buy_{i,t} + Sell_{i,t}} \tag{11}$$

We further retrieve retail trading in options markets from Nasdaq, that covers all electronic trades on the Nasdaq Options Market (NOM) or Nasdaq PHLX (PHLX). This dataset provides information on option trades made by non-professional customers. For our analysis, we distinguish between trades that establish a long position in a stock and those that come from a short position. Specifically, we categorize opening buys and closing buys of call options, as well as opening sells and closing sells of put options, as trades that establish a long position in a stock. Conversely, we consider opening buys and closing buys of put options, as well as opening sells and closing sells of call options, as trades that indicate a short position in a stock. Therefore, we define *Retail long (short) option volume* as the sum of opening and closing buys for call (put) options, and opening and closing sells for put (call) options. We then compute order imbalance measures of retail trading in the equity and options market as follows:

$$\textit{Retail option } OI_{i,t} = \frac{\textit{Retail long opt vlm}_{i,t} - \textit{Retail short opt vlm}_{i,t}}{\textit{Retail long opt vlm}_{i,t} - \textit{Retail short opt vlm}_{i,t}} \tag{12}$$

C. Model derivation

Derivation of Results The wishful thinking investor will choose subjective beliefs π_H and π_L by maximizing expected utility in (9) taking into account that $\pi_H + \pi_L = 1$. The Lagrangian of the investor is given by

$$\mathcal{L} = q(\pi_H v_H + \pi_L v_L - p) - \frac{1}{\theta}\pi_H \ln\frac{\pi_H}{\bar{\pi}_H} - \frac{1}{\theta}\pi_L \ln\frac{\pi_L}{\bar{\pi}_L} - \mu(\pi_H + \pi_L - 1)$$

where μ is a Lagrange multiplier. The first order condition with respect to π_H is given by

$$qv_H - \frac{1}{\theta}\ln\frac{\pi_H}{\bar{\pi}_H} - \frac{1}{\theta} - \mu = 0.$$

A similar first order condition can be found for π_L. The first order conditions can be rearranged to yield

$$\pi_H = \bar{\pi}_H \exp\left(\theta q v_H - \theta\mu - 1\right) \qquad \text{and} \qquad \pi_L = \bar{\pi}_L \exp\left(\theta q v_L - \theta\mu - 1\right). \tag{13}$$

Plugging (13) into $\pi_H + \pi_L = 1$, we obtain

$$\exp\left(\theta\mu + 1\right) = \bar{\pi}_H \exp\left(\theta q v_H\right) + \bar{\pi}_L \exp\left(\theta q v_L\right)$$

If we plug this expression back into (13), we get (10).