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

Very preliminary draft. Do not distribute.

March 18, 2023

ABSTRACT

We examine information production surrounding earnings announcements on leading investment social networks. In aggregate, information production on social networks displays excessive positive bias about future outcomes on earnings announcements. This bias does not predict fundamentals on earnings announcements. It leads to buying pressure before earnings announcements, thus, distorting prices from fundamentals before earnings announcements with negative news. For rarer cases of extreme pessimism, we find selling price pressure before positive and negative earnings news. Our findings cast doubt on the wisdom of the crowd phenomenon from social networks in forecasting future fundamentals and having a beneficial role in price efficiency.

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

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]

Users of such financial social networks deliberately choose to consume information that aligns with their priors, which could result in echo chambers, according to Cookson, Engelberg, and Mullins (2022), and apparent deviations from rationality such as optimism, according to Caplin and Leahy (2019). Eventually it could create bubbles, if such "wishful thinking" investors engage in information-based trade (Caplin and Leahy (2019)).

In this paper, we examine the implications of the wisdom-of-the-crowd effect on market efficiency around earnings announcements from three leading investor social networks (StockTwits, WallStreetBets, and Seeking Alpha). 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 indeed, a wisdom-of-the-crowd effect exists and investors trade in line with social media

[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.

sentiment, then stock prices leading up to earnings announcements should reflect future fundamentals such as earnings announcement surprises.

We find the information produced and shared on such financial networks and consumed by investors is positively skewed. Specifically, positive sentiments account for more than 80% of post activity for nearly 70% of the earnings announcement sample. Such optimism across all social media platforms before earnings announcements does not predict firm fundamentals, such as earnings surprises, or stock returns on earnings announcements. In other words, we find no evidence of the wisdom-of-the-crowd effect. Instead, our findings suggest the existence of misinformation and positive bias in the content shared by retail investors on social media platforms.

Such excessive optimism has important implications for price efficiency leading to earnings announcements. Our findings suggest that stocks with an abnormally high number of posts on social networks are associated with higher retail trading activity in equity and options markets and, more importantly, with higher 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 to the day of the 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, with increases of up to 2% from five days before to the day of the earnings announcement. 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-related 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 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.

Prior research suggests that retail investors are attracted to news events eliciting them to buy rather than to sell because selling involves the investor owning the stock. This tendency may explain the existance 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.

We then provide a simple theoretical framework that rationalizes the empirical findings of the paper. The model is based on the concept of wishful thinking from Caplin and Leahy (2019) where agents who get utility from their beliefs and therefore interpret information optimistically. The model illustrates how a wishful thinking investor can distort beliefs about upcoming earnings announcements, thus leading to excessive positivism on social media platforms.

Overall, this paper contributes to understanding social media's role in financial markets through retail trading. While social media platforms, like Stocktwits, Wallstreetbets, and Seeking Alpha, 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 events, as evidenced by our findings on the effects of social media sentiment on stock prices before earnings announcements.

Our paper similar to Cookson and Niessner (2020), Cookson, Engelberg, and Mullins (2022),Hu, Jones, Zhang, and Zhang (2021), Bradley, Hanousek Jr, Jame, and Xiao (2021), and Jiao, Veiga, and Walther (2020), analyzes the relation between information in social media and future stock returns. In Cookson, Engelberg, and Mullins (2022), the authors find that retail attention and sentiment of social media information in Stocktwits predict negative future returns. Our findings complement their results. We document that the excess optimism found on such platforms can misinform investors and lead to negative returns, i.e., in the case of excessive positivism before negative earnings announcements. 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. Like other social media platforms, we show that the tone is positively skewed. We also find that tones on Wallstreetbets posts relate to retail order flow but fail to predict firm fundamentals.

Cookson and Niessner (2020) and Cookson, Engelberg, and Mullins (2022) examine the impact of StockTwits on financial markets. As we show in this paper, StockTwits is the most dominant social media platform for explaining price impact before earnings announcements. In Cookson, Engelberg, and Mullins (2022), the authors find that retail attention and sentiment of social media information predict negative future returns. Our findings complement their results. We document that the excess optimism found on such platforms can misinform investors and lead to negative returns, i.e., in the case of excessive positivism before negative earnings announcements. Chen, De, Hu, and Hwang (2014), and Dim (2020) find that the content of Seeking Alpha posts, months priors earnings announcements, is more informative about earnings surprises. Farrell, Green, Jame, and Markov (2022) find Seeking Alpha posts can immediately impact retail trading minutes following its release. More recently, Cookson, Lu, Mullins, and Niessner (2022) highlighted the differences between the 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. 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. Moreover, we find no evidence that Seeking Alpha posts days before earnings are informative about fundamentals.

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. Like other social media platforms, we show that the tone is positively skewed. We also find that tones on Wallstreetbets posts relate to retail order flow but fail to predict firm 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 (2021) showed that retail investors are inclined to trade high-attention stocks, like, stocks in the news, stocks experiencing high abnormal trading volume, stocks with extreme 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. Furthermore, our paper explains why in aggregate retail investors earn poor returns, a fact that has been demonstrated in previous literature (Barber and Odean, 2000, Barber, Lee, Liu, and Odean, 2009, Xiaoyan and Zhang, 2022). Our paper's findings thus warn investors that consuming information from social media can be hazardous to their wealth.

2. Data Description

2.1. Stock and earnings announcements sample

The time coverage of this study spans two periods: (1) January 1, 2018, to December 31, 2021, when incorporating information from the three leading social media platforms we consider, and (2) from January 1, 2013, to December 31, 2021, when focusing solely on StockTwits. We select stocks with share code 10 or 11 from the Center for Research in Security Prices (CRSP). We retrieve daily stock returns, prices, outstanding shares, and ticker symbols. Because ticker symbols are required for merging with social media posts, we select only stocks with available tickers in CRSP.

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 $SUE_{i,\tau}$ of stock i on earnings announcement τ, as the difference of 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. 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 for the period of 2018 to 2021 from RavenPack.

2.2. Social media platforms

We merge the data of three of the most important social media platforms used by retail investors: Stocktwits, WallstreetBets, and Seeking Alpha.

For Stocktwits, we download all posts or "tweets" from the social media platform through

RapidAPI. Users of this platform can add a StockTwits $ Cashtag (Exs: $AMZN, $GOOG, $SPY) to the tickers they are writing posts. In addition, users can optionally assign a "Bullish" or a "Bearish" sentiment to each post. Therefore, posts on Stocktwits are stock-specific and have an exact sentiment assigned by their authors.

For Seeking Alpha, we obtain all the opinion articles of all stocks considered in this analysis through RapidAPI. In this platform, the authors must refer to a company by its first name and include its stock ticker when they mention it in an article or post. They also have the option to add a sentiment feature to each article like "Very Bullish," "Bullish," "Neutral," "Bearish," or "Very Bearish." Therefore, similar to Stocktwits, the posts on Seeking Alpha are also stock-specific and have an explicit sentiment assigned by their authors. For our analysis, we consider posts with positive sentiment to those tagged as "Very Bullish" and "Bullish," and we will consider posts with negative sentiment to those tagged as "Bearish" or "Very Bearish."

For WallStreetBets (WSB), we download all posts using the Reddit API. Different from Stocktwits and Seeking Alpha, the posts of WallStreetBets are not tagged either by the stock or by their sentiment. In other words, we do not have a direct variable that identifies the stock the post is related to or the sentiment associated with it. To solve this issue, we first scrap all the tickers from CRSP with codes 10 and 11 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 text. Next, we proceed to calculate the sentiment of each post in Wallstreetbets. Unfortunately, the usage of tools like Loughran and McDonald's dictionaries is not adequate for social media language. As mentioned in Bradley, Hanousek Jr, Jame, and Xiao (2021), the language of WSB users is full of sarcasm, jokes, bad words, slang, and emojis. Therefore, we calculate the sentiment of every post using Machine Learning and Natural Language Processing techniques. Specifically, we train a Supervised

Learning Method used for classification, called: Support Vector Machine, using a subsample of 100,000 Stocktwits stock-related tweets and their associated sentiments (50,000 tweets with Bullish sentiment and 50,000 tweets with Bearish sentiment). To train WSB posts, we use Stocktwits tweets because they use similar language, slang, and emojis to communicate their sentiment towards a stock. 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 WSB expressions described in Appendix B of Bradley, Hanousek Jr, Jame, and Xiao (2021).

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 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 WSB post. For our sample of WallStreetBets posts, we consider posts of all categories except the ones posted by Moderators (tagged as "MOD").

In total, we gathered 89,998,526 posts from Stocktwtis, 457,889 posts from Wallstreetbets, and 65,438 posts from Seeking Alpha. Only 44,557,725 posts on Stocktwits are tagged with a sentiment view; with 88% of posts are bullish, and 12% are bearish. For Seeking Alpha only 41,494 posts are tagged with a sentiment view, with 85% of the posts are bullish, and 15% are bearish. Similarly, 73% of posts on Wallstreetbets have a bullish sentiment, and 27% have a bearish sentiment.

To observe how the participation in financial social networks has changed over time, we plot in Figure 1 the aggregated monthly number of posts for each platform, from January 1, 2018, to December 31, 2021.[2] For all of them, social media activity increased strongly from the last quarter of 2019, in line with the surge of retail trading due to the retail brokerage initiating zero trading costs and COVID. This figure shows that the information produced in Stocktwtis exceeds the number of posts on Wallstreetbets and Seeking Alpha. Stocktwits also has a wider breadth of coverage, with 4,192 different stocks in their posts, versus 3,717 in Wallstreetbets and 2,958 in Seeking Alpha. Therefore the results of our analysis will be highly driven by information produced in Stocktwits.

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 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 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 WSB have no limit of characters for their opinion articles and posts. Contributors on SA 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

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

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. Moreover, they show that a portfolio constructed using messages from professional users exhibits positive cumulative abnormal returns, whereas the novice portfolio exhibits negative abnormal returns. Different from opinion articles of Seeking Alpha, posts on WSB and Stocktwits tend to be considerably less in-depth. According to Bradley, Hanousek Jr, Jame, and Xiao (2021), anecdotal evidence suggests that WSB also places a larger emphasis on highly speculative trading strategies.

2.2.1. Earnings announcements covered in Social Media Platforms

From January 2018 to December 2021, Stocktwits covers the greatest number of earnings announcements, precisely 37,756. Wallstreetbets follow it with 5,569, and Seeking Alpha with 2,908 earnings announcements covered. As a comparison, analysts' recommendations from Ravenpack cover 3,534 stock earnings announcements. A platform covers a stock-earning announcement if at least one post is related to that stock five days before the announcement. In detail, the coverage of stock earnings announcements of every platform by market capitalization is shown on Panel A of Table 1. 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 of Table 1 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. Overall, this is evidence of how important social media is for investors looking for information related to small-cap stocks and can't be found in analysts' reports before earnings announcements.

3. Empirical Results

3.1. Social media information production around earnings announcements

We examine how information on social media is produced around earnings announcements. We compare this information production to analyst-related news (i.e., 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 a measure of abnormal posts ten days around earnings announcements. Similar to Fisher, Martineau, and Sheng (2022), we define the number of abnormal posts for a stock as the difference between the number of posts on the day t minus the average number of posts from $t = -30$ to $= -11$.[3] We plot in Figure 2 the average abnormal number of posts for StockTwits, WallstreetBets, and Seeking Alpha in Panels A to C, respectively. We plot also the abnormal number of analysts' recommendations and the abnormal number of news articles from RavenPack in Panels D and E, respectively. Figure 3 shows the fraction of earnings announcements in our sample with at least one social media post, analyst recommendation, and newswire.

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

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

announcements. Similarly, Panel D shows that analyst recommendations occur mainly after the earnings announcement; consistent with the findings of Ivković and Jegadeesh (2004), and Gregoire and Martineau (2022).

Figure 3 shows that information production on StockTwits covers a wide cross-section of stocks compared to WallStreetBets. About 70% of our stock-earnings announcement observations in our sample have at least one post and increase to 80% the day before the earnings announcements. The fraction of observations with at least one post on WallStreetBets and Seeking Alpha is less than 10%. For newswires, a little less than 20% of firms get news coverage the day before earnings announcements. We conclude that the timing of information production varies across platforms and that if social media plays a role in market efficiency leading to earnings announcements, StockTwits is expected to play an important role.

If users that share content on social networks are generally retail investors (Cookson, Lu, Mullins, and Niessner, 2022), such investors are particularly prone to trading attention-grabbing stocks Barber and Odean (2008). We next examine factors that are expected to influence the amount of shared information prior to earnings announcements. To examine the abnormal number of posts in the cross-section of stocks, we need to take into account the growth in the number of posts over time. We define the following measure of abnormal posts for each social network platform:

$$Abn\ post_{i,j,t} = P_{i,j,t} - \sum_{n=1}^{N} \frac{P_{n,j,t}}{N} \tag{1}$$

where $P_{i,j,t}$ is the proportion of posts of stock i, $i = 1, .., N$, for stock i on platform j on day t. $\sum_{n=1}^{N} \frac{P_{n,j,t}}{n}$ is the average of the proportion of posts of all n stocks at day t.

$$P_{i,j,t} = \frac{\text{\#posts of stock i, for platform j, on day t}}{\text{\#posts on day t}} \tag{2}$$

$P_{i,t} - \sum_{j=1}^{n} \frac{P_{j,t}}{n}$ measures the abnormal number of posts of stock i as a proportion of all posts compared to all other stocks' post at same day t. We next run the following regression

$$\begin{aligned} Abn\ post^{j}_{[-5,-1],i} = & \beta_1 Abn\ ret_{[-30,-6]i} + \beta_2 |Abn\ ret|_{[-30,-6],i} + \beta_3 Surprise_{[q-1],i} + \\ & \beta_4 LnMCAP_i + \beta_5 Analysts_i + \beta_6 News\ count_{[-30,-6]i} + \beta_7 LnVol_{[-30,-6],i} + \\ & \beta_8 Retail\ LnVol_{[-30,-6],i} + \beta_9 Abn\ post^{j}_{[-30,-6],i} + \alpha_i + \alpha_t + \varepsilon_i \end{aligned}$$

where *Abn post* corresponds to the average daily abnormal number of posts defined in Equation (1) five to one day before earnings announcement for stock-earnings announcement i for social network platform $j = \{\text{ST, WSB, SA}\}$. *Abn ret* and $|Abn\ ret|$ correspond to the abnormal return over 30 days until 5 days and its absolute value before the earnings announcement, respectively. *Surprise* is the earnings announcement surprise of the previous quarter. $LnMCAP$ is the natural logarithm of market capitalization. *Analysts* corresponds to the total number of analysts issuing a forecast for earnings announcement i. *News count* is the total number of news articles from RavenPack from 30 to 5 days before the earnings announcement. $LnVol$ and *Retail* $LnVol$ is the average natural logarithm of the total daily volume and retail volume from 30 to 5 days before the announcement. $Abn\ post^{j}_{[-30,-5],i}$ is the average daily abnormal posts from 30 to 5 days before announcements for social platform j. α_i and α_t correspond to the stock and year-quarter fixed effects, respectively. All independent variables except the log-transformed variables are standardized.

The results of Table 2 suggest that drivers to information production vary across platforms. Column (1) shows that stock returns, absolute returns, and trade volume positively predict higher information content for StockTwits before announcements. These results are consistent with Barber and Odean (2008), where large returns and high trade volume drive retail investor attention. Controlling for abnormal lag posts, column (2) reports that only returns and absolute returns positively relate to higher abnormal posts. Columns (3)-(6) show that volume predicts posts for WallStreetBets and absolute return predicts posts for Seeking Alpha. The lag posts variables show that abnormal posts activity does not predict abnormal posts across platforms.

3.2. Positively biased information production and fundamentals

We next turn our attention to the content of social media posts to assess their informativeness. We examine the distribution of sentiments for every platform. Using the sentiment of every post, we compute a variable called positive ratio for stock i on the day t on platform p, as follows:

$$PosRatio_{i,t,p} = \frac{N_{i,t,p}^{bullish}}{N_{i,t,p}^{bullish} + N_{i,t,p}^{bullish}}$$

We further compute the average positive ratio of all the posts shared five days before the earnings announcement on each platform. Figure 4, Panel A, shows the fraction of stock-earnings announcements observation with positive sentiment post ratio $\leq 20\%$, between 20 and 40%, 40 and 60%, 60 and 80%, and greater than 80%. Remarkably, we find that more than 70% of the total posts in Stocktwits and WSB before earnings announcement have a positive ratio greater than 80%. For Seeking Alpha, the fraction of positive posts greater than 80% is around 65%. We benchmark these numbers to the fraction of positive posts analyst recommendations. Around 60% of analyst recommendations have an aggregate positive ratio higher than 80%, largely below what is found on social media platforms. Moreover, let's take the whole data sample and not limiting to the five days before earnings announcements but aggregating it at the monthly frequencies. Panel B shows a similar distribution.

The latter evidence already casts doubts about the informativeness of the content shared on social media. However, we further examine this premise by validating if the information shared on social media relates to stock earnings firms' fundamentals. Specifically, if the information produced before announcements and its correspondent sentiment can predict earnings surprise and announcement date stock future returns. For that, we regress the earnings surprise $SUE_{i,t}$ of stock i on earnings announcement t on the sentiment average of posts, our average abnormal measure $AbnPost_{i,t}$ on five days before the announcement, and the interaction of both variables. We later run the same regression for the buy-and-hold abnormal returns from 5 days before the earnings announcement. The specifications of both regressions are as follow:

$$Surprise_{i,t} = \beta_1 Sent_{i,j,t} + \beta_2 \mathbb{1}_{[AbnPost]i,j,t} + \beta_3 Sent_{i,j,t} \times \mathbb{1}_{[AbnPost]i,j,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

$$BHAR[0,1]_{i,t} = \beta_1 Sent_{i,j,t} + \beta_2 \mathbb{1}_{[AbnPost]i,j,t} + \beta_3 Sent_{i,j,t} \times \mathbb{1}_{[AbnPost]i,j,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Surprise$ is the earnings surprise for stocks' earnings announcement i and $BHAR[0,1]$ corresponds to the buy-and-hold abnormal returns on the earnings announcement date and the following day. $Sent$ and $AbnPost$ correspond to the sentiment measure and abnormal information production measure, respectively, for $j =$ StockTwits, WallStreetBets, Seeking Alpha, and Analysts.

The results of these regressions are shown in Table 3. Panel A shows that sentiment and abnormal information production can not predict earnings surprises on any social media platforms. Similarly, Panel B exhibits that sentiment and abnormal information production on WSB and Seeking Alpha do not have any predictive power of future stock returns. For Stocktwits, the coefficients are negative and statistically significant at the 10% level. Overall, these results confirm that the information produced and shared on social media platforms is not related to fundamentals and, thus, not informative. Indeed, it is not valuable for investment decision-making and could negatively impact investors' wealth.

3.3. Social media information production and retail trading around earnings announcements

Because investors appeal to these platforms as a source of investment advice, even if the content is biased and uninformative, we now investigate its influence on investors' trading. In this sense, we analyze the correlation of abnormal information production with retail trading variables both in the equity and in the options market in the five days before the earnings announcement. We calculate retail traders' total trading volume on the equity market, following the novel approach of 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 number of buys and sell trades (and in volume) in the equity market as:

$$\textit{Retail } OI_{i,t} = \frac{Buy_{i,t} - Sell_{i,t}}{Buy_{i,t} + Sell_{i,t}} \tag{3}$$

We further retrieve retail trading in options markets from Nasdaq. Our Nasdaq dataset covers all electronic trades that occur on the Nasdaq Options Market (NOM) or Nasdaq PHLX (PHLX). It provides the daily number of opening buys, opening sells, closing buys, and closing sells of call and put options initiated by non-professional customers. Opening buys, closing buys of call options, and opening sells, and closing sells of put options are trades that establish as volume coming from a long position in a stock. Opening buys, closing buys of put options, and opening sells and closing sells of call options are trades that establish as volume coming from a short position in a stock. We define *Retail long option volume* as the sum of opening and closing buys for call options, and opening and closing sells for put options. We next define *Retail short option volume* as the sum of opening and closing buys for put options and opening and closing sells for call options. We then compute order imbalance measures of retail trading in the equity and options market as follows:

$$\textit{Retail option } OI_{i,t} = \frac{\textit{Retail long option volume}_{i,t} - \textit{Retail short option volume}_{i,t}}{\textit{Retail long option volume}_{i,t} - \textit{Retail short option volume}_{i,t}} \tag{4}$$

To examine the implication of positively-skewed optimism on social media on retail trading and retail trading order imbalance, we regress the retail trading volume and retail order imbalance on our main variables of abnormal information production of the three platforms, as follows:

$$\Delta \textit{Retail trading} = \sum_j \beta_j AbnPost^j_{i,t} + \delta_1 |Surprise|_{i,t} + \delta_2 |News\ sent|_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}$$

$$\textit{Retail } OI = \sum_j \beta_j AbnPost^j_{i,t} + \gamma_1 Surprise_{i,t} + \gamma_2 News\ sent_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where j = {ST, WSB, SA}, corresponding to the three social media platforms. The dependent variables $\Delta Retail\ trading$ is the change in retail trading (number of trades), retail volume, and retail option trading volume from $t = [-60, -6]$ to $t = [-5, -1]$ for stock i on time t. $Retail\ OI$ is the average daily retail trading (volume) order imbalance and retail option order imbalance for $t = [-5, -1]$. $AbnPost_{i,t}^{ST}$, $AbnPost_{i,t}^{WSB}$, and $AbnPost_{i,t}^{SA}$ is the daily average level of abnormal posts on StockTwits, WallStreetBets, and Seeking Alpha, respectively. $Surprise$ ($|Surprise|$) is the earnings announcement (absolute) surprise. $News\ sent$ ($|News\ sent|$) is the daily average (absolute average) news sentiment from RavenPack for $t = [-5, -1]$.

Panel A reports that 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. The relationship between abnormal posts from StockTwits and retail equity trading is four times greater than WallStreetBets and ten times greater than Seeking Alpha. Similarly, the impact of StockTwits on retail option volume is greater than the other platforms. Panel B reports that abnormal posts are associated with positive order imbalance, i.e., stocks with high information production on social media experience greater buying pressure.

Overall, these results confirm the idea of Barber and Odean (2008) that investors are net buyers of attention-grabbing stocks, we show that is for stocks that experienced high abnormal information on social media. Also, our findings are in line with Cookson, Lu, Mullins, and Niessner (2022), who showed a connection between the direction of retail trading and the sentiment of social media information.

3.4. Social media and price impact

After highlighting that information produced on social media platforms increases retail trading before earnings announcements, specifically for buy trades, we now examine its impact on stock returns and price efficiency. As first step, we compute the stock's buy-and-hold abnormal returns five days before to five days after earnings announcements as follows:

$$BHAR_{[-5,5],i,t} = \prod_{k=-5}^{5}(1 + R_{i,k}) - \prod_{k=-5}^{5}(1 + R_{m,k}), \quad (5)$$

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

Figure 5 shows the average $BHAR$ 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 and low abnormal posts for stocks as stocks with daily average abnormal posts greater than zero in the window $t = [-5, -1]$. Panel A shows greater price run-ups for stocks of approximately 1% with high abnormal posts before earnings announcements. At first, this could suggest that social media-induced retail trading is pushing prices toward fundamentals. However, Panel B shows that before earnings announcements with negative surprises, we find positive price run-ups of also 1%. This suggests a distortion in price efficiency. 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. Previous work from Barber, Huang, Odean, and Schwarz (2021) shows that price distortions are more significant for small-cap stocks. Figure 7 shows that, indeed, 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 is about 2%. For large stocks, we find limited evidence of price run-ups. 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.

To ensure statistical significance on these results and robust to other factors potentially impacting price run-ups before earnings announcements, we run the following regression

$$BHAR_{[-5,-1]} = \sum_j \beta_j \mathbb{1}_{[j],i,t} + \sum_j \beta_j \mathbb{1}_{[j],i,t} \times \mathbb{1}_{[Small]i,t} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}$$

where $\mathbb{1}_{[j]}$ corresponds to indicator variables equal to one if the social media platform $j =$ {ST, WSB, SA} abnormal number of posts at time t 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. The control variables are $\mathbb{1}_{Ana}$, $\mathbb{1}_{Ana} \times \mathbb{1}_{Small}$, $\mathbb{1}_{News}$, $\mathbb{1}_{News} \times \mathbb{1}_{Small}$, $Surprise$, $News\ sent_{[-5,-1]}$, and $Analyst\ sent_{[-5,-1]}$. Similar to how we define abnormal, posts on social media (see Equation 1), $\mathbb{1}_{Ana}$ ($\mathbb{1}_{News}$) is an indicator variable equal to one if the number of abnormal analyst recommendations (newswire article) is positive, zero otherwise. $Surprise$ is the earnings surprise of earnings announcement i. $News\ sent$ is the average news sentiment in RavenPack five to one day before the earnings announcement. α_i and α_t correspond to firm 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).

The first main result that stands out from Table 5 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 potential largest impact on prices and price efficiency because it provides greater coverage in the cross-section of stocks. When interacting with $\mathbb{1}_{Small}$, it confirms the finding in Figure 5 that most of the effect of social media occurs for small stocks. Finally, the overall results remain the same when including the control variables.

In aggregate, these results confirm that stocks with high information production on social media platforms (especially on Stocktwits) before earnings announcements present price run-ups of 1%, indistinctively if they have a positive or negative surprise. These results are stronger for stocks with low market capitalization. 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?

Abnormal stock information production generated by content on social media platforms is related to an increase in the cumulative stock returns before the earnings announcement, aligned with the fact that content in social media is positively biased. We next focus on the smaller proportion of posts that are highly negative in tone. Because the sample size is small, we examine the role of negative sentiment for StockTwits only starting from 2013 to 2021. Figure 7 presents the buy-and-hold abnormal returns from 5 days before to 5 days after earnings announcement for stocks with high positive sentiment ratio (≥ 0.80) and low positive sentiment ratio (< 0.20) 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 sentiment matters. In the rarer cases of extreme pessimism, the second and fourth columns show downward price drifts in earnings announcements. When negative sentiment is high, prices before earnings announcements with negative surprises become more efficient as prices move toward fundamentals. In contrast, prices become less efficient before positive surprises as prices are pushed away from fundamentals. Overall, we find that even in extreme pessimism, social media information fails to make prices more efficient. We also report in Table 6 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.

3.6. 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 result from the pump and dump scheme. At Priori, our findings indicate that the pump and dump scheme is not a systematic force explaining price drifts because we find limited evidence of significant return reversals following earnings announcements (see Figure 5 and 6). We further examine whether pre-announcement returns (BHAR[-1,-5]) negatively predict post-announcement returns (BHAR[1,5]). We separately report the regression results in Table 7 for 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 $\mathbb{1}_{[Small]}$ show no statistical significance. We conclude that the pump and dump scheme plays (if any) a small role in price drifts before earnings announcements.

3.7. Retail trades that follow social media advice

The results in Table 3 show that sentiment and abnormal information production of StockTwits, WallStreetBets, and Seeking Alpha fail to predict earnings surprises and announcement date returns. Previous studies find that retail trading can predict future returns. For example, Cookson, Lu, Mullins, and Niessner (2022) find that sentiment-induced retail trading imbalance predicts positive next-day returns. Similar to Cookson, Lu, Mullins, and Niessner (2022), we decompose the retail trading order imbalance into four components: order imbalance that is related to information production, sentiment, interaction of information production and sentiment, and the residual component that is orthogonal to the previous three components. We compute this decomposition by regressing the retail trading order imbalance on our information production variable, sentiment, and the interaction of information production and sentiment from 2013 to 2021 using daily StockTwits data only.

We define the "Order imbalance induced" as the sum of these three components. "Order Imbalance Residual" is orthogonal to "Order Imbalance Induced". We proceed after to forecast earnings surprises and announcement day returns as in section 3.2, as follow:

$$Surprise_{i,t} = \beta_1 OIInduced_{i,t} + \beta_2 OIResidual_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Surprise$ is the earnings surprise for stocks' earnings announcement i. $OIInduced$ is the average order imbalance induced by the information production and sentiment of Stocktwits, from five days before the stock earnings announcement. $OIResidual$ is the average order imbalance not related and independent from the information of Stocktwits, from five days before the stock earnings announcement.

The results of these regressions, shown in Table 8, suggest that the average retail order imbalance five days before announcements predicts earning surprises (column (1), Panel A). But columns (3) and (4) report that the power of forecasting stock fundamentals comes from the order imbalance that is not induced and independent from the information produced Stocktwits (i.e., OIResidual). Our findings suggests that retail trading followed by social media is not informative about stock fundamentals, but retail trading unrelated to StockTwits is informative about fundamentals.

4. A Simple Model of Wishful Thinking to 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 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 (or "posts" in the context of social media) 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

$$EU(\pi_H, \pi_L) = q(\pi_H v_H + \pi_L v_L - p) - \frac{1}{\theta}\pi_H \ln \frac{\pi_H}{\bar{\pi}_H} - \frac{1}{\theta}\pi_L \ln \frac{\pi_L}{\bar{\pi}_L}. \tag{6}$$

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 (6) taking into account that $\pi_H + \pi_L = 1$. The optimization problem leads the investor to choose the following subjective beliefs:[4]

$$\pi_H = \frac{\bar{\pi}_H \exp(\theta q v_H)}{\bar{\pi}_H \exp(\theta q v_H) + \bar{\pi}_L \exp(\theta q v_L)}. \tag{7}$$

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 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 9a. 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

[4]See the Appendix for derivations

(7). 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 9b, 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 (6) with subjective beliefs given by (7) 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 ($\theta \to \infty$). For any finite θ, 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.

A key implication of the model is that investors are more likely to be predisposed to wishful thinking when the ease of deviating from objective beliefs is greater. We show this will be the case as stocks with more social media attention lead to more positive optimism. We calculate the cumulative distribution of the positive sentiment ratios before earnings announcements for stocks with high and low abnormal posts five days before earnings announcements. Figure 8 presents the cumulative distribution and shows that cumulative distribution exhibits a more significant mass for high positive sentiment ratios, closer to 1 for "high abnormal posts" than "low abnormal posts." In short, social media content with higher information production exhibits greater positive optimism and thus makes it easier for investors' beliefs to deviate from objective beliefs.

5. Conclusion

We examine information production surrounding earnings announcements on leading investment social networks: WallStreetBets, Stocktwits, and Seeking Alpha. 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) and obtaining utility from their beliefs and interpret 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, 2021, Attention induced trading and returns: Evidence from robinhood users, *Journal of Finance, forthcoming.*

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.

——— , 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.

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, .

Cookson, J Anthony, Joseph Engelberg, and William Mullins, 2022, Echo chambers, forthcoming, *Review of Financial Studies.*

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.

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.

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.

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.

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, .

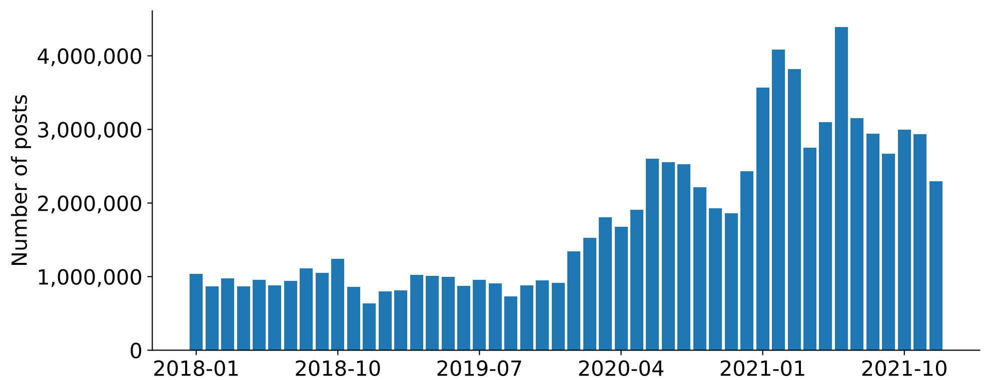
Wasik, John F., 2013, 'pump and dump' schemes resurface in social media, .

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

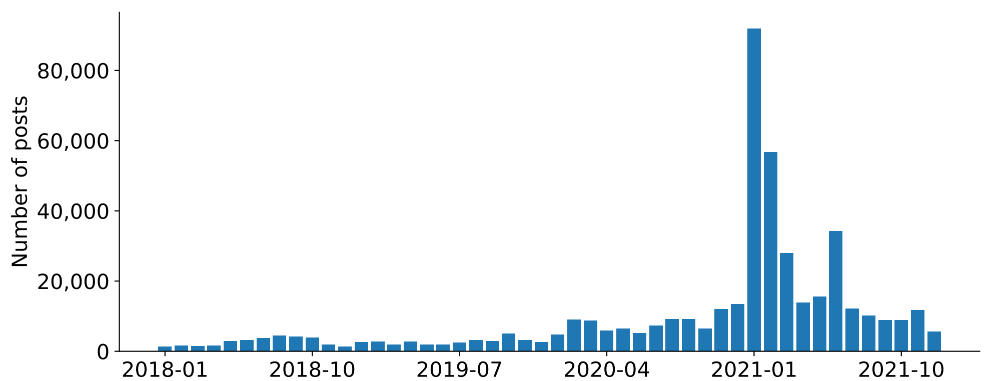
Figure 1. Social Media Information Production

This figure shows the number of stock-specific posts on social media platforms aggregated monthly, from January 1, 2018, to December 31, 2021.

Panel A. Stocktwits



Panel B. Wallstreetbets



Panel C. Seeking Alpha

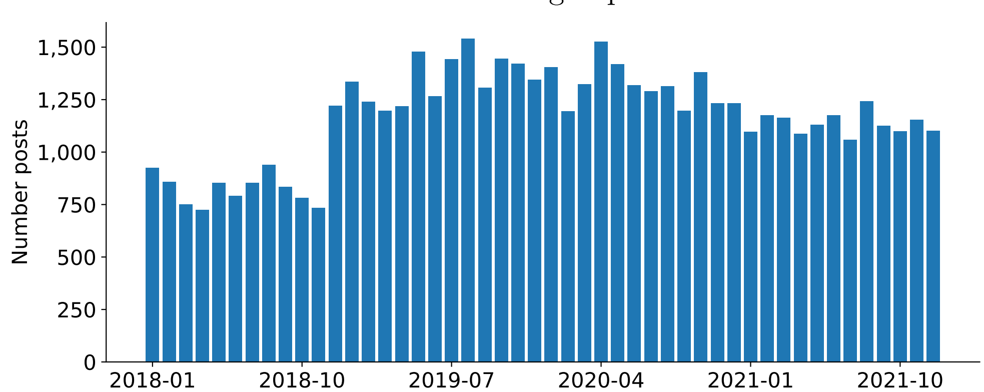
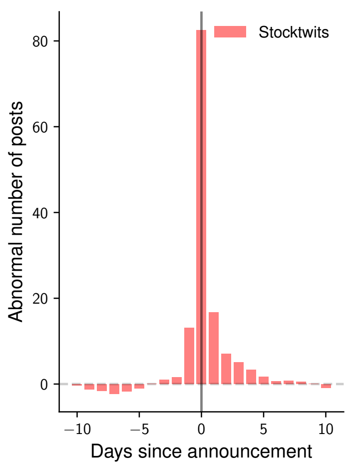


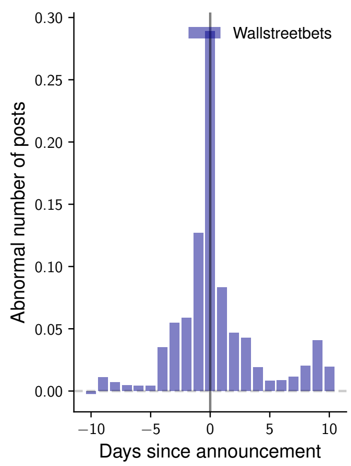
Figure 2. Social Media and Analyst Information Production

This figure shows the mean of the abnormal number of posts and the fraction of the stock-earnings announcement observations with at least one post ten days around earnings announcements for StockTwits, WallStreetBets, and Seeking Alpha in Panels A to C, respectively. Panels D and E show the abnormal number of analyst recommendations and news articles in Ravenpack, respectively. 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.

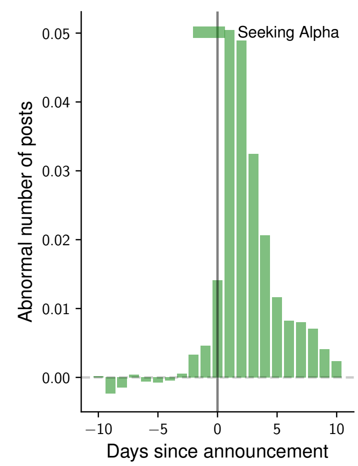
Panel A. Stocktwits



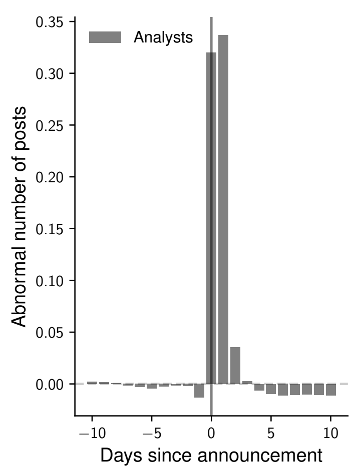
Panel B. Wallstreetbets



Panel C. Seeking Alpha



Panel C. Analysts



Panel D. News

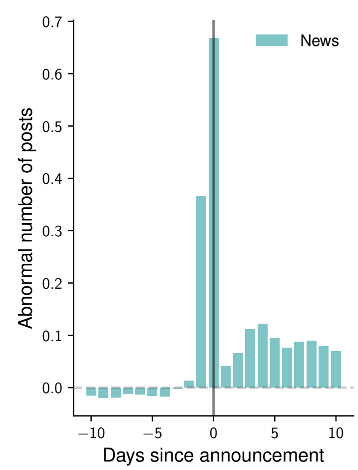
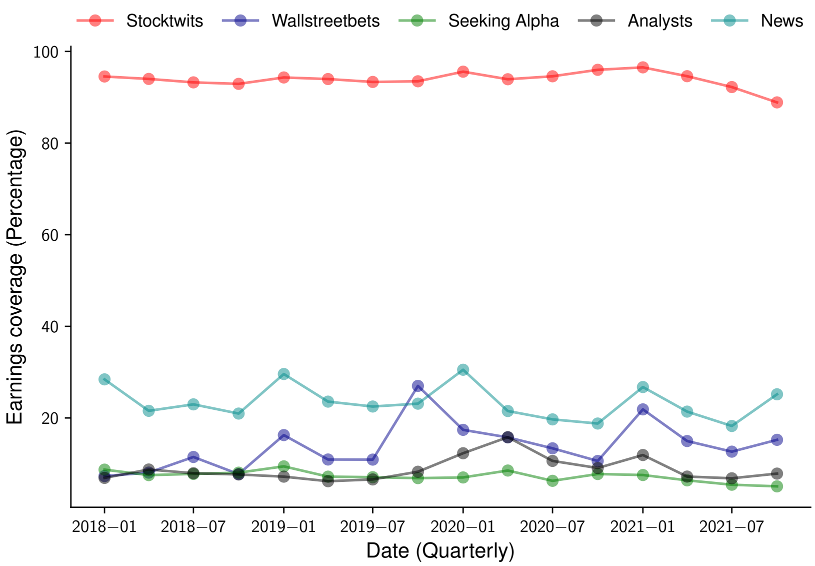


Figure 3. Information Coverage

This figure shows the fraction of stock-earnings announcement observations, for every quarter, with at least one post 5 days around earnings announcements for StockTwits, WallStreetBets, Seeking Alpha, Analysts recommendations and News articles in Ravenpack. The sample period is from January 1, 2018, to December 31, 2021.

Panel A.Percentage of Earnings Aannouncements
mentioned at least once 5 days before, in every quarter



Panel B.Percentage of Earnings Aannouncements
mentioned at least once 30 days before, in every quarter

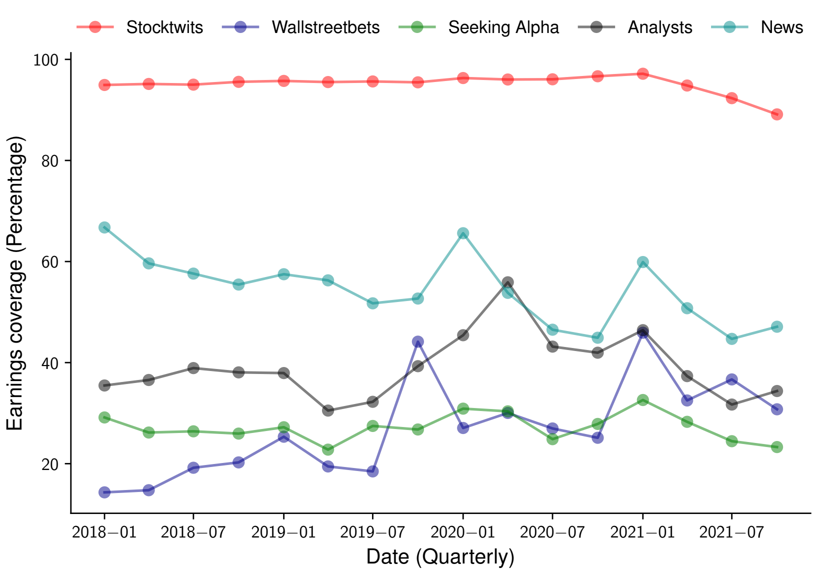
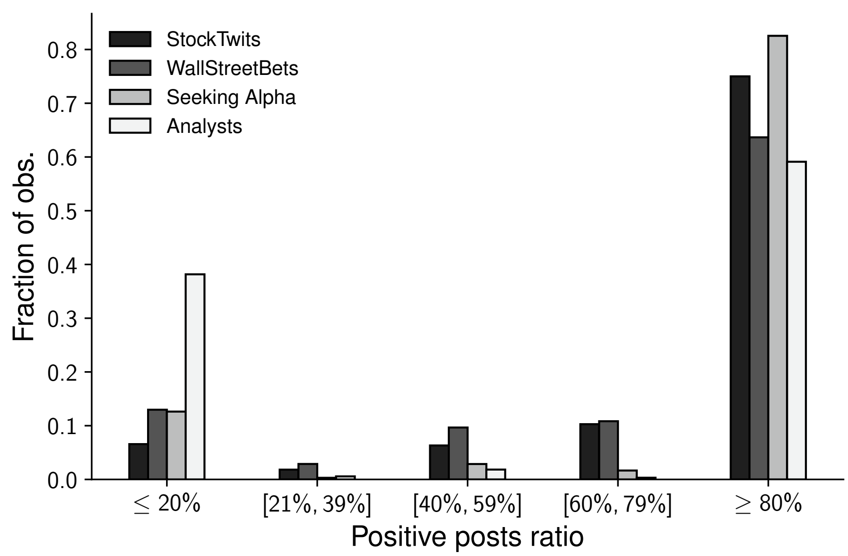


Figure 4. Positive Skewness in Sentiment

Panel A shows the fraction of stock-earnings announcements observation with positive sentiment posts ratio $\leq$ 20%, between 20 and 40%, 40 and 60%, 60 and 80%, and greater than 80%. Panel B shows the fraction of aggregated monthly posts with post ratio $\leq$ 20%, between 20 and 40%, 40 and 60%, 60 and 80%, and greater than 80%. 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.

Panel A. Daily frequency
(average of five days prior to earnings announcement)



Panel B. Monthly frequency

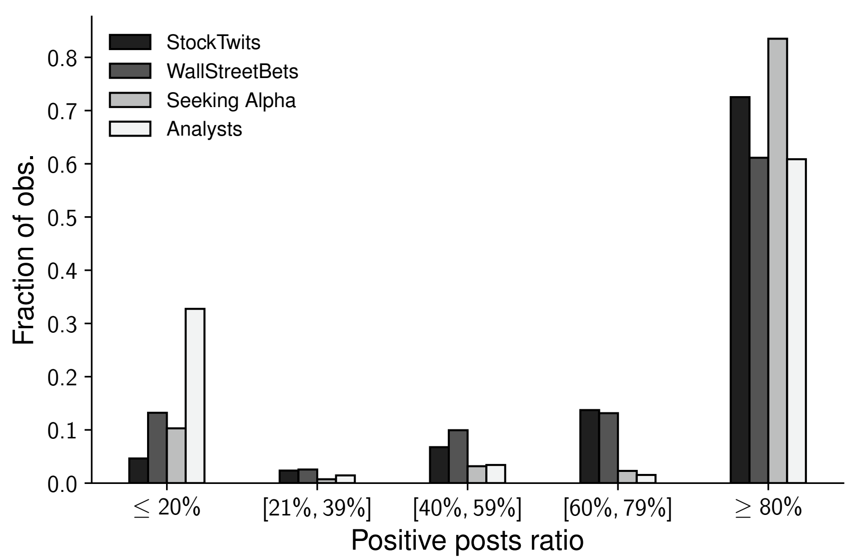
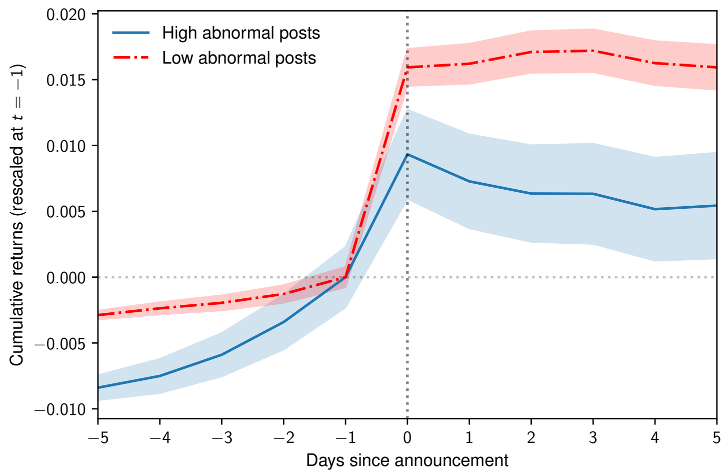


Figure 5. Cumulative Returns Around Earnings Announcement

This figure shows the abnormal cumulative returns five days before to 5 days after earnings announcements for stocks with high and low social media abnormal information production for positive and negative earnings surprises in Panels A and B, respectively. High (low) abnormal posts are defined as stocks with abnormal number of posts (less or equal to) zero for either StockTwits, WallStreetBets, or Seeking Alpha. We rescaled the plot such that the 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

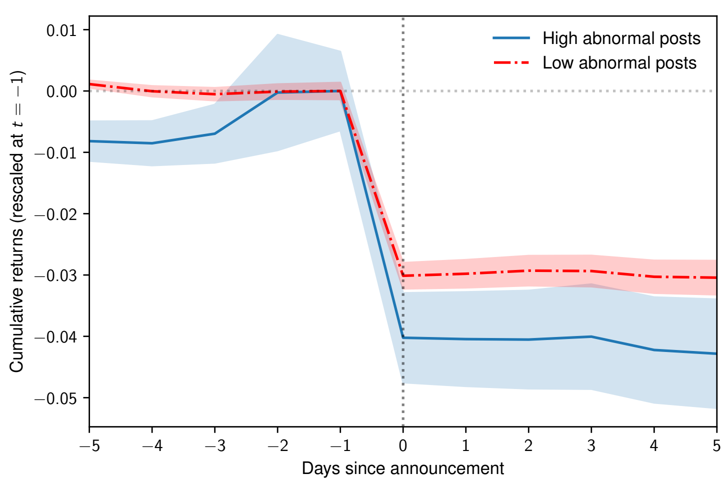
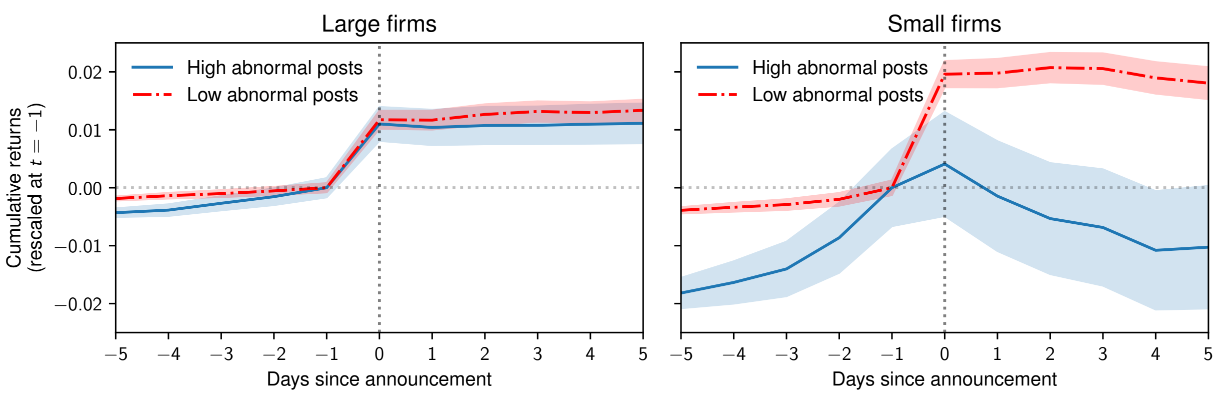


Figure 6. Cumulative Returns Around Earnings Announcement for Large and Small Firms

This figure shows the abnormal cumulative returns five days before to 5 days after earnings announcements for stocks with high and low social media information production for positive and negative earnings surprises and for large and small firms. Panels A and B show the cumulative returns for positive and negative surprises, 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 as stocks with abnormal number of posts (less or equal to) zero for either StockTwits, WallStreetBets, or Seeking Alpha. We rescaled the plot such that the 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

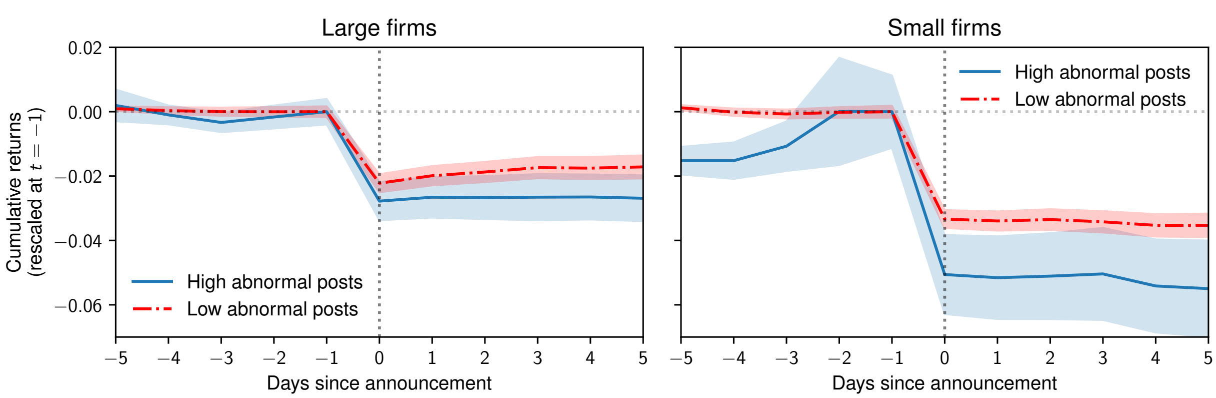
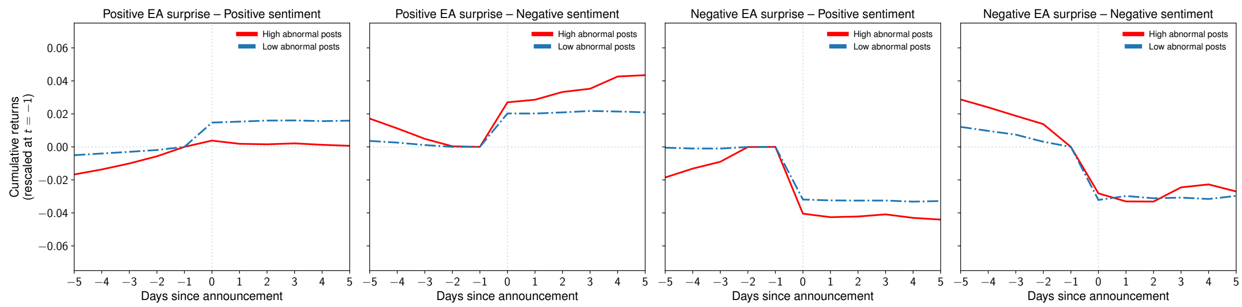


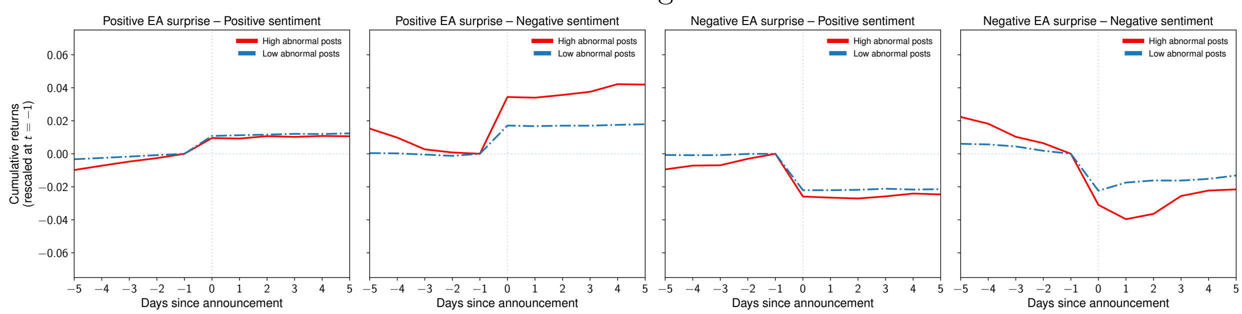
Figure 7. Cumulative Returns Around Earnings Announcement for High and Low Sentiment Stocks

This figure shows the abnormal cumulative returns five days before to 5 days after earnings announcements for stocks with high positive sentiment ratio (≥ 0.80) and low positive sentiment ratio (< 0.80) and for positive and negative earnings surprises for the full sample, large, and small firms in Panels A to C, respectively. We restrict our sample 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. We rescaled the plot such that the 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

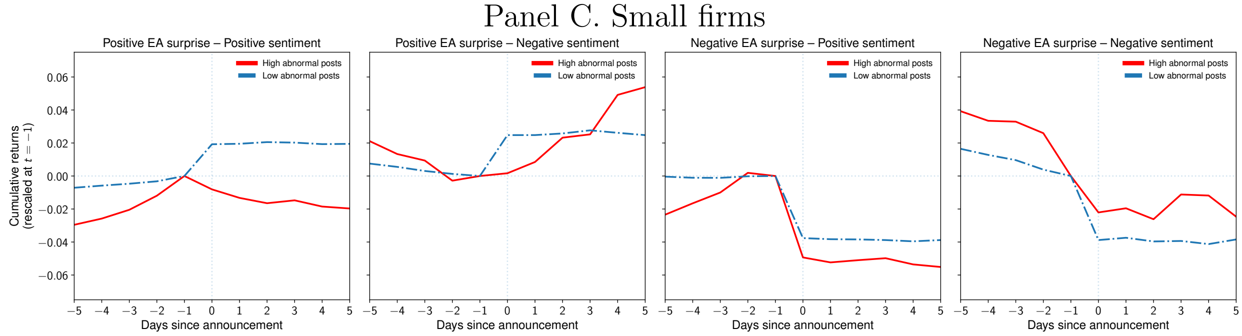


Figure 8. Positive Skewness and Abnormal Information Production before Earnings Announcements

The figure shows the cumulative distribution of the positive sentiment ratio for StockTwits with high and low abnormal post five days before earnings announcements. The positive 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.

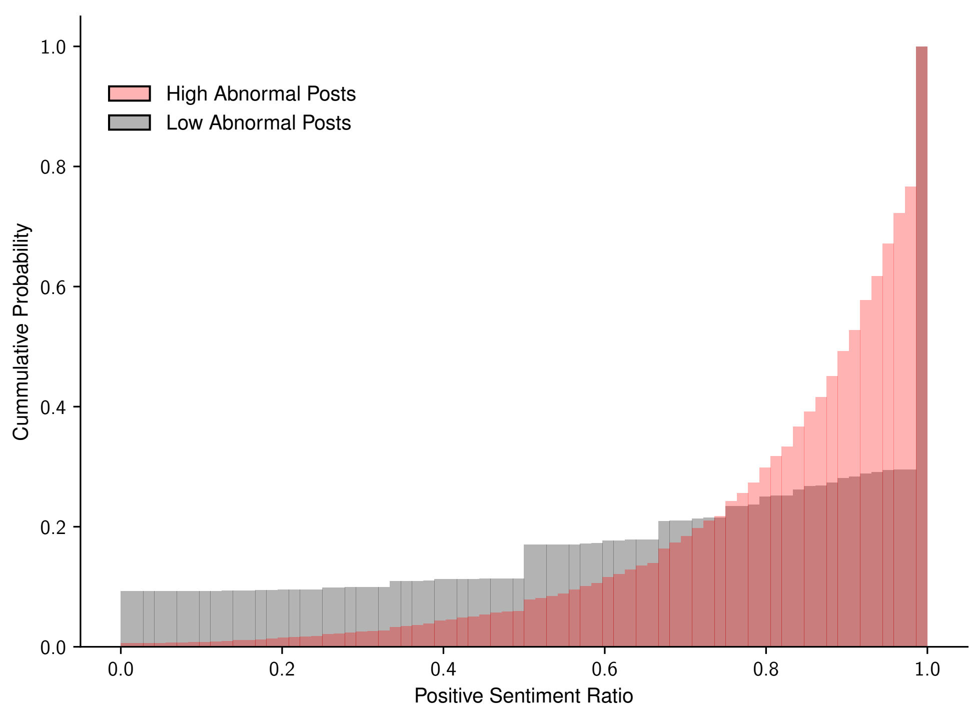
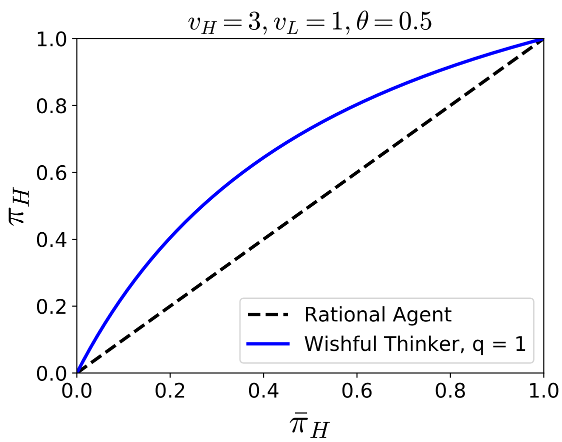
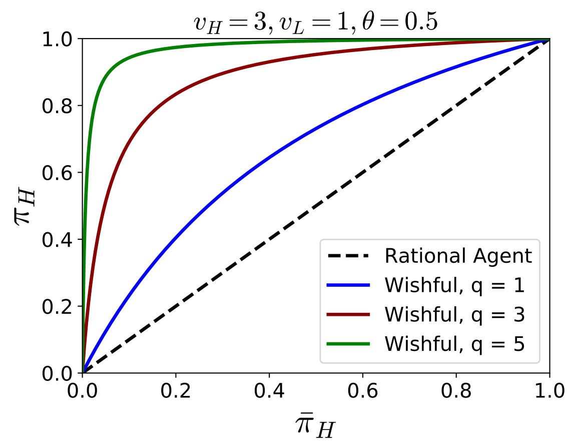


Figure 9. Plots of π_H for Wishful Thinking vs Rational Investors. Dashed black line represents a rational agent. Solid lines represent wishful thinking investors for different quantities q.



(a)



(b)

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 and 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
Drivers to Social Media Information Production Before Earnings Announcements

This table reports the coefficients of the following regression

$$\begin{aligned}Abn\ post_{[-5,-1],i,t} = &\beta_1 Abn\ ret_{[-30,-6]i,t} + \beta_2 |Abn\ ret|_{[-66,-6]} + \beta_3 Surprise_{[q-1],i,t} + \beta_4 Ln\ MCAP_{i,t} + \\ &\beta_5 N\ analysts_{i,t} + \beta_6 News\ count_{[-30,-6]i,t} + \beta_7 Ln\ vol_{[-30,-6],i,t} + \beta_8 Retail\ ln\ vol_{[-30,-6],i,t} + \\ &\beta_9 Abn\ post^{ST}_{[-30,-6],i,t} + \beta_{10} Abn\ post^{WSB}_{[-30,-6],i,t} + \beta_{11} Abn\ post^{SA}_{[-30,-6],i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}\end{aligned}$$

where *Abn post* corresponds to the average daily abnormal number of posts defined in Equation (1) five to one day before earnings announcement for stock-earnings announcement i on social network platform j =StockTwits, WallStreetBets, or Seeking Alpha. *Abn ret* and $|Abn\ ret|$ correspond to the abnormal return over 30 days until 5 days and its absolute value before earnings announcement, respectively. *Surprise* is the earnings announcement surprise of the previous quarter. $LnMCAP$ is the natural logarithm of market capitalization. *Analysts* corresponds to the total number of analysts issuing a forecast for earnings announcement i. *News count* is the total number of news articles from RavenPack from 30 to 5 days before the earnings announcement. $LnVol$ and *Retail* $LnVol$ is the average natural logarithm of the total daily volume and retail volume from 30 to 5 days before the announcement. $Abn\ post^j_{[-30,-5],i}$ is the average daily abnormal posts from 30 to 5 days before announcements for social platform j. α_i and α_t correspond to the stock and year-quarter fixed effects, respectively. All independent variables except the log-transformed variables are standardized. 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.

	Dependent variable:					
	Abn postST		Abn post^{WSB}		Abn post^{SA}	
	(1)	(2)	(3)	(4)	(5)	(6)
Abn ret$_{[-30,-6]}$	0.210***	0.155**	0.020	0.021	-0.049*	-0.040*
	(0.079)	(0.061)	(0.033)	(0.023)	(0.027)	(0.025)
$\lvert$Abn ret$\rvert_{[-30,-6]}$	0.342***	0.286***	0.056	0.049	0.098**	0.090**
	(0.095)	(0.109)	(0.046)	(0.036)	(0.039)	(0.039)
Surprise$_{[q-1]}$	0.772	0.160	0.174	0.134	-0.029	-0.085
	(0.545)	(0.249)	(0.112)	(0.082)	(0.132)	(0.135)
Ln MCAP	0.024	0.030	-0.006	0.001	0.011	0.014
	(0.059)	(0.035)	(0.030)	(0.027)	(0.018)	(0.013)
N analysts	-0.004	-0.000	-0.001	-0.000	0.004	0.004
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
News count$_{[-30,-6]}$	0.337***	0.006	0.370***	0.036	0.156	-0.043
	(0.105)	(0.046)	(0.141)	(0.058)	(0.100)	(0.083)
Ln vol$_{[-30,-6]}$	0.197***	-0.008	0.060***	0.024***	0.033***	0.004
	(0.026)	(0.022)	(0.013)	(0.008)	(0.009)	(0.006)
Retail ln vol$_{[-30,-6]}$	0.013**	0.004	0.018***	0.011***	0.007	0.002
	(0.005)	(0.003)	(0.005)	(0.003)	(0.008)	(0.004)
Abn post$^{ST}_{[-30,-6]}$		0.743***		-0.010		0.049*
		(0.144)		(0.037)		(0.026)
AbnPost$^{WSB}_{[-30,-6]}$		-0.078		0.622***		0.029
		(0.074)		(0.047)		(0.034)
Abn post$^{SA}_{[-30,-6]}$		-0.036		0.008		0.264***
		(0.030)		(0.041)		(0.078)
N	36,268	36,268	36,268	36,268	36,268	36,268
R^2	0.0395	0.4611	0.0111	0.3268	0.0022	0.0582
Firm FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y

Table 3
Forecasting Earnings Surprise and Announcement Day Returns with Sentiment

This table reports the coefficients of the following regression:

$$
\begin{aligned}
Surprise_{i,t} =& \beta_1 Sent_{i,j,t} + \beta_2 AbnPost_{i,j,t} + \beta_3 Sent_{i,j,t} \times AbnPost_{i,j,t} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel A,} \\
BHAR[0,1]_{i,t} =& \beta_1 Sent_{i,j,t} + \beta_2 AbnPost_{i,j,t} + \beta_3 Sent_{i,j,t} \times AbnPost_{i,j,t} + BHAR[-5,-1]_{i,t} + \\
& \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel B.}
\end{aligned}
$$

$Surprise$ is the earnings surprise for stocks' earnings announcement i and $BHAR[0,1]$ corresponds to the buy-and-hold abnormal returns on the earnings announcement date and the following day. $Sent_{i,j,t}$ corresponds to the average sentiment of posts from 5 to 1 days before earnings announcement t. $AbnPost_{i,j,t}$ is the average abnormal measure from 5 to 1 days before earning announcement t of stock i, for j = {StockTwits, WallStreetBets, Seeking Alpha, and Analysts}. All independent variables except the log-transformed variables are standardized. 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.

Panel A. Forecasting earnings surprise

	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. Forecasting BHAR[0,1]

	StockTwits		WallStreetBets		Seeking Alpha		Analyst	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sent	-0.002	-0.004***	-0.000	0.000	0.003	0.005	-0.003	-0.001
	(0.001)	(0.001)	(0.003)	(0.003)	(0.007)	(0.021)	(0.004)	(0.005)
AbnPost		0.001		-0.002**		-0.002		0.000
		(0.002)		(0.001)		(0.005)		(0.001)
Sent× AbnPost		-0.006***		-0.001		-0.000		-0.001
		(0.002)		(0.002)		(0.005)		(0.001)
$BHAR[-5,-1]$	-0.036*	-0.030	-0.056*	-0.054*	0.085	0.084	-0.023	-0.023
	(0.020)	(0.020)	(0.031)	(0.031)	(0.140)	(0.140)	(0.052)	(0.053)
N	27,054	27,054	5,569	5,569	1,988	1,988	3,533	3,533
R^2	0.0009	0.0020	0.0022	0.0031	0.0058	0.0068	0.0008	0.0009

Table 4
Retail Trading and Social Media Information Production

This table reports coefficients of the following regression

$$\Delta \textit{Retail trading} = \beta_1 AbnPost_{i,t}^{ST} + \beta_2 AbnPost_{i,t}^{WSB} + \beta_3 AbnPost_{i,t}^{SA} + \beta_4 |Surprise|_{i,t} + \beta_5 |News\ sent|_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel A,}$$
$$\textit{Retail OI} = \beta_1 AbnPost_{i,t}^{ST} + \beta_2 AbnPost_{i,t}^{WSB} + \beta_3 AbnPost_{i,t}^{SA} + \beta_4 Surprise_{i,t} + \beta_5 News\ sent_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ in Panel B,}$$

where the dependent variables Δ*Retail trading* is the change in the average retail trading, retail volume, and retail option volume from $t = [-30, -6]$ to $t = [-5, -1]$ for stock i on time t in columns (1)-(4), columns (5)-(8), and columns (9)-(12), respectively. *Retail OI* is the average daily retail trading imbalance, retail volume imbalance, and option volume order imbalance for $t = [-5, -1]$ in columns (1)-(4), columns (5)-(8), and columns (9)-(12), respectively. $AbnPost_{i,t}^{ST}$, $AbnPost_{i,t}^{WSB}$, and $AbnPost_{i,t}^{SA}$ is the average of abnormal number of posts on StockTwits, WallStreetBets, and Seeking Alpha, respectively. *Surprise* (|*Surprise*|) is the earnings announcement (absolute) surprise. *News sent* (|*News sent*|) is the average (absolute average) news sentiment from RavenPack for $t = [-5, -1]$. All independent variables except the log-transformed variables are standardized. 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.

Panel A. Retail trading

	Dependent variable:											
	ΔRetail trading				ΔRetail volume				ΔRetail option volume			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AbnPost^{ST}	0.119***			0.122***	0.132***			0.133***	0.078***			0.069***
	(0.010)			(0.011)	(0.011)			(0.011)	(0.009)			(0.007)
AbnPost^{WSB}		0.030***		-0.008		0.036***		-0.006		0.040**		0.018
		(0.008)		(0.006)		(0.008)		(0.007)		(0.016)		(0.017)
AbnPost^{SA}			0.010***	0.003			0.011***	0.004*			0.013***	0.009***
			(0.002)	(0.002)			(0.002)	(0.002)			(0.003)	(0.003)
|Surprise|	-0.004	0.118	0.110	-0.009	0.002	0.136	0.127	-0.003	-0.233	-0.151	-0.162	-0.225
	(0.294)	(0.311)	(0.313)	(0.295)	(0.317)	(0.326)	(0.329)	(0.317)	(0.436)	(0.431)	(0.430)	(0.432)
$\|\text{News sent}\|_{[-5,-1]}$	0.272**	0.303**	0.290**	0.268**	0.267*	0.302*	0.286*	0.264*	0.217*	0.247**	0.229*	0.228**
	(0.130)	(0.141)	(0.143)	(0.131)	(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	40,254	40,254	40,254	40,254
R^2	0.0275	0.0027	0.0008	0.0277	0.0206	0.0022	0.0006	0.0206	0.0038	0.0013	0.0003	0.0041

Panel B. Retail order imbalance

	Dependent variable:											
	Retail trading OI				Retail volume OI				Retail option volume OI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AbnPost^{ST}	0.011***			0.011***	0.004***			0.004***	0.002			0.000
	(0.001)			(0.001)	(0.001)			(0.001)	(0.002)			(0.002)
AbnPost^{WSB}		0.003***		0.000		0.001*		0.000		0.003*		0.003*
		(0.001)		(0.001)		(0.001)		(0.001)		(0.002)		(0.002)
AbnPost^{SA}			0.001***	0.001*			-0.000	-0.000			0.002	0.002
			(0.000)	(0.000)			(0.001)	(0.001)			(0.001)	(0.001)
Surprise	-0.006	-0.008	-0.007	-0.006	0.110***	0.110***	0.110***	0.110***	0.042	0.041	0.042	0.042
	(0.046)	(0.046)	(0.046)	(0.046)	(0.040)	(0.040)	(0.040)	(0.040)	(0.059)	(0.059)	(0.059)	(0.059)
$\text{News sent}_{[-5,-1]}$	-0.003	-0.000	-0.002	-0.003	0.004	0.005	0.005	0.004	-0.024	-0.022	-0.023	-0.022
	(0.016)	(0.015)	(0.015)	(0.016)	(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	39,259	39,259	39,259	39,259	40,254	40,254	40,254	40,254
R^2	0.0025	0.0003	0.0001	0.0025	0.0004	0.0002	0.0002	0.0004	0.0000	0.0001	0.0001	0.0001

Table 5
Pre-Earnings Announcement Cumulative Returns and Social Media Information Production

This table reports the coefficients of the following regression:

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

where $\mathbb{1}_{[ST]}$, $\mathbb{1}_{[WSB]}$, and $\mathbb{1}_{[SA]}$ are indicator variables equal to one if the social media abnormal number of posts at time t is positive, zero otherwise, for Stocktwits, WallstreetBets, and Seeking Alpha, respectively. $\mathbb{1}_{[Small]}$ 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 $\mathbb{1}_{Ana}$, $\mathbb{1}_{Ana} \times \mathbb{1}_{Small}$, $\mathbb{1}_{News}$, $\mathbb{1}_{News} \times \mathbb{1}_{Small}$, $Surprise$, $News\ sent_{[-5,-1]}$, and $Analyst\ sent_{[-5,-1]}$. Similar to how we define abnormal, posts on social media (see Equation 1), $\mathbb{1}_{Ana}$ ($\mathbb{1}_{News}$) is an indicator variable if the number of abnormal analyst recommendations (newswire article) is positive, zero otherwise. $Surprise$ is the earnings surprise of earnings announcement i. $News\ sent$ is the average news sentiment in RavenPack five to one day before the earnings announcement. α_i and α_t correspond to firm 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$_{[-5,-1]}$				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 6
Average Cumulative Returns Prior to Earnings Announcements Conditioned on Stocktwits' Information Production

This table reports the average buy-and-hold abnormal returns five to one day 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 the days prior to 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 7
Forecasting Cumulative Return after EA

This table reports the coefficients of the following regression:

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

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 social media abnormal number of posts at time t is positive, zero otherwise, for StockTwits, WallStreetBets, and Seeking Alpha, respectively. $\mathbb{1}_{[Small]}$ is an indicator variable equal to one if the stock-earnings announcement i belongs to the bottom two NYSE market capitalization quintiles. 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]×$\mathbb{1}_{AbnPost}$		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]×$\mathbb{1}_{[Small]}$			-0.041			-0.029			0.002
			(0.004)			(0.003)			(0.003)
BHAR[-5,-1] ×$\mathbb{1}_{AbnPost}$ × $\mathbb{1}_{[Small]}$			0.038			0.008			-0.112
		(0.031)	(0.032)		(0.020)	(0.031)		(0.027)	(0.058)
$\mathbb{1}_{AbnPost}$ × $\mathbb{1}_{[Small]}$			-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

Table 8
Forecasting Earnings Surprise with Induced and Residual Order Imbalance

This table reports the coefficients of the following regression:

$$Surprise_{i,t} = \beta_1 OIInduced_{i,t} + \beta_2 OIResidual_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where *Surprise* is the earnings surprise for stocks' earnings announcement i. *OIInduced* is the order imbalance induced from the information production and sentiment of Stocktwits. *OIResidual* is the order imbalance not related and independent from the information of Stocktwits. All independent variables except the log-transformed variables are standardized. 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, 2013, to December 31, 2021.

	Retail volume				Option volume			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Volume OI Total	0.001**							
	(0.001)							
Volume OI Induced		-0.037		-0.035				
		(0.025)		(0.025)				
Volume OI Residual			0.001**	0.001**				
			(0.001)	(0.001)				
Option OI Total					0.000			
					(0.000)			
Option OI Induced						0.004		0.004
						(0.018)		(0.018)
Option OI Residual							0.000	0.000
							(0.000)	(0.000)
N	91,580	91,580	91,580	91,580	91,580	91,580	91,580	91,580
R^2	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000

Appendix

Derivation of Results The wishful thinking investor will choose subjective beliefs π_H and π_L by maximizing expected utility in (6) 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

$$\pi_H = \bar{\pi}_H \exp\left(\theta q v_H - \theta\mu - 1\right) \qquad \text{and} \qquad \pi_L = \bar{\pi}_L \exp\left(\theta q v_L - \theta\mu - 1\right). \tag{8}$$

Plugging (8) 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 (8), we get (7).