The Rise of Reddit:
How Social Media Affects Belief Formation and Price Discovery

Danqi Hu, Charles M. Jones, Siyang Li, Valerie Zhang and Xiaoyan Zhang*

February 2025

ABSTRACT

Using submission level data from the social media platform Reddit, we rely on the theoretical framework of Pedersen (2022) to examine how social media affects belief formation, price discovery, and trading dynamics. Consistent with the predictions on network belief spillover, we find that the opinions of hardheaded investors (rational or fanatic) significantly predict future opinions of naïve investors, especially when these investors have larger influence. For return predictions, social media tones positively and significantly predict future returns, and more so when agents' influence is higher. Finally, for trading dynamics, higher agent tones in networks with higher agent influence increase retail flows and deter shorting flows. Short sellers' consideration of agent influence in deciding to ride or burst bubbles enhances their abilities to predict negative returns. These findings generally support Pedersen's predictions.

Keywords: social media, belief formation, return prediction, short selling, retail investors.
JEL codes: G11, G12, G14, G23.

* Danqi Hu (email: danqi.hu@gsm.pku.edu.cn) is at Guanghua School of Management, Peking University. Charles M. Jones (email: cj88@gsb.columbia.edu) is at Columbia Business School. Siyang Li (email: lisy.20@pbcsf.tsinghua.edu.cn) is at the PBC School of Finance at Tsinghua University. Valerie Zhang (email: valerie.zhang@haas.berkeley.edu) is at the Haas School of Business at the University of California, Berkeley, and Xiaoyan Zhang (Corresponding author, email: zhangxiaoyan@pbcsf.tsinghua.edu.cn) is at the PBC School of Finance at Tsinghua University. We would like to thank Lin Tan and Longjiang He for their excellent research assistance. The paper benefits from comments from Matthew Ringgenberg, Zhi Da, Marina Niessner, Byoung-Hyoun Hwang, seminar participants at Tsinghua University, Harvard Business School, Northwestern University, as well as conference participants at CFEA, WFA, EFA, CICF, and ABFER. Jones thanks the Norwegian Finance Initiative at Norges Bank for helpful funding. Zhang acknowledges financial support from China NSF Project 71790605.

The Rise of Reddit:
How Social Media Affects Belief Formation and Price Discovery

February 2025

ABSTRACT

Using submission level data from the social media platform Reddit, we rely on the theoretical framework of Pedersen (2022) to examine how social media affects belief formation, price discovery, and trading dynamics. Consistent with the predictions on network belief spillover, we find that the opinions of hardheaded investors (rational or fanatic) significantly predict future opinions of naïve investors, especially when these investors have larger influence. For return predictions, social media tones positively and significantly predict future returns, and more so when agents' influence is higher. Finally, for trading dynamics, higher agent tones in networks with higher agent influence increase retail flows and deter shorting flows. Short sellers' consideration of agent influence in deciding to ride or burst bubbles enhances their abilities to predict negative returns. These findings generally support Pedersen's predictions.

Keywords: social media, belief formation, return prediction, short selling, retail investors.
JEL codes: G11, G12, G14, G23.

1. Introduction

GameStop Corp. is an American video game retailer. Over a short period from January 4, 2021, to January 29, 2021, its closing share price rockets from $17.25 to $325.00, an increase of almost 18-fold. This enormous upswing in price forms a powerful short squeeze, directly leading to the failures of some short-selling institutions, such as Melvin Capital, and threatens the liquidity of many institutions who have leveraged short positions on GameStop. The extremely high share price of GameStop does not last long. Two weeks later, the share price plummets to $40.59. This enormous volatility of the GameStop share price attracts substantial attention, and most investors connect the dramatic ups and downs in share prices to retail investors gathering and investing together, and to a discussion hub, the "WallStreetBets" forum, on the social media platform Reddit, where retail investors share opinions on stocks. Many regulators and investors are wondering: can social media significantly affect how beliefs and prices are formed and how different types of investors behave?

A recent study by Pedersen (2022) provides a coherent and comprehensive theoretical framework for understanding social network dynamics. By separating investors into three categories, rational, fanatic and naïve, all of whom interact in a social network, Pedersen (2022) derives closed-form solutions for the dynamics of beliefs and prices and proposes three main testable predictions to answer the above question. First, Pedersen (2022) shows that there are belief spillovers from social network interactions, and naïve investors' beliefs can be predicted by both rational and fanatic views, even more so if the rational and fanatic investors have higher influence in the network. Second, the views held by different investors can directly predict future share price movements, especially when these investors are more influential. In some cases, social media views can drive prices away from the rational price, leading to potential price bubbles. Finally, investors' demand for company shares changes as they observe social network views and anticipate price movements away from fundamentals. They optimize their trading behaviors by balancing between riding the bubble and bursting the bubble. In particular, rational investors might choose to ride or burst the bubble depending on the costs and benefits of doing so. When they ride the bubble, it may result in prices further deviating from fundamentals for an extended period of time. The length and magnitude of the price deviation from fundamental values thus depend on the mixture of different types of investors and the relative importance of their influences in the market.

Relying on the predictions from Pedersen's theoretical framework, we begin our empirical investigation by collecting data from the social media platform Reddit from January 2020 through December 2023. Following Pedersen's definitions, we first separate all investors into hardheaded (investors who do not change their opinions) and naïve investors (investors who have fluid opinions). We further separate hardheaded investors into rational (who pay attention to firm fundamental values) and fanatic investors (who do not pay attention to fundamentals). With our empirical categorization following Pedersen's assumptions, about 3% of Reddit users are identified as rational, 5% are fanatics, and 92% are naïve.

To account for the complex dynamics among agents' beliefs, returns, and trading from different market participants, we choose panel vector auto-regression (PVAR) as our main empirical method, which is designed to capture dynamics and interactions among different variables. In addition, PVAR allows us to infer which variable may be important to the future outcomes of another variable through Granger-causal relations, and to quantify the responses of the variables to innovations in other variables using impulse response functions.

We first examine how social networks affect belief formation and price movements. Using textual analysis, we measure each investor's opinion by her tone. The results show that rational and fanatic views significantly and positively predict future naïve agent views, which supports Pedersen's prediction on network belief spillover. One key variable for the social network structure is an agent's influence, which captures how much attention each investor attracts from others in the network. In the case of Reddit, we measure different agents' influence by the sum of the number of direct commenters these agents receive. We further document that the predictive power of rational and fanatic agents' views on future naïve investors' views is stronger in the network where these agents have higher influence. For price movements, we directly use investors' tones to predict future returns. As predicted by Pedersen, higher fanatic and naive tones are associated with higher future returns. More interestingly, the predictive power of rational, fanatic, and naïve tones on future returns is stronger when these agents have higher influence.

Next, we investigate how other important market participants, such as retail investors and short-sellers, trade in the presence of social media activities. Retail investors are generally viewed as less sophisticated than institutional investors. They tend to follow social media trends and have played a unique role in the GameStop episode. Our empirical results show that Reddit tones significantly predict future retail trading, in the sense that higher naïve tones predict higher future retail flows, and the predictive power is even stronger for networks with higher influence. In contrast, short sellers are generally believed to be informed and rational, and our results show that the predictive relation between Reddit tones for future shorting flows is complicated and intriguing. When agents' influence is higher, shorting flows significantly decrease in response to positive fanatic and naïve tones. By contrast, when agents' influence is lower, short sellers' trading is not related to Reddit tones. This result indicates that short sellers, to some extent, are deterred by

Reddit tones when Reddit agents are more influential, but when influence is low, short sellers might not feel threatened by agents' positive views.

We further look into short sellers' negative predictive power for future returns for these cases. Consistent with the prior literature, shorting flows significantly and negatively predict future returns, even with the inclusion of social media variables. Surprisingly, the negative predictive power of shorting flows is stronger in the subsample of stocks with higher agent influence. That is to say, when agents are more influential on Reddit, short sellers become even more informative and predict even lower future returns. Combined with earlier results showing that short sellers may generally shy away from short selling when social media tones are higher (riding the bubble) in the high influence network, when they do choose to short (bursting the bubble), their shorting flows are more informative about future negative returns.

Our study naturally connects to three strands of literature: social media, retail investors and short sellers. Earlier studies on social media, such as Tumarkin and Whitelaw (2001), Antweiler and Frank (2004), Das and Chen (2007), Chen et al. (2014), and Bartov et al. (2018), provide suggestive evidence that users' social media activities, sentiment, and dispersion of sentiment are correlated with stock returns, trading volume, and volatility. More recent studies delve deeper into investors' belief formation in the social network. For instance, Cookson and Niessner (2020) find that investor disagreement attributed to information differences is more important for trading than disagreement stemming from different investment approaches; and Cookson, Engelberg, and Niessner (2023) find evidence of investors' selective exposure to confirmatory information and echo chambers in the network. For retail investors, many earlier papers focus on whether retail investors are pure noise traders or informed traders. For instance, Barber and Odean (2008), Barber, Odean, and Zhu (2009), find that retail investors are generally uninformed, while Kaniel et al.

(2008), Kelley and Tetlock (2013), Fong, Gallagher, and Lee (2014) and Boehmer, Jones, Zhang and Zhang (2021) show that retail investors' trading can predict future stock returns. For the most recent period around the COVID-19 pandemic, Ozik, Aadka and Shen (2021) show large increases in retail trading, and research interest shifts to the new generation of retail investors on Robinhood including Welch (2022), Eaton, Green, Roseman and Wu (2022), and Barber, Huang, Odean and Schwarz (2022). There is also a vast literature on short sellers. Theoretical work by Diamond and Verrecchia (1987) argues that short sellers are more informed than average traders. Empirically, previous work such as Desai, Ramesh, Thiagarajan, and Balachandran (2002), Asquith, Pathak, and Ritter (2005), Boehmer, Jones, and Zhang (2008) and Boehmer, Huszár, and Jordan (2010) show that high trading activity by short sellers predicts low future stock returns. Engelberg, Reed, and Ringgenberg (2012) report that the information advantage of short sellers arises partly from their superior public information-processing skills. For the case of GameStop, Allen, Nowak, Pirovano, and Tengulov (2022) provide evidence that the January 2021 episode is a short squeeze.

Compared to the large volume of previous literature on social media, retail investors, and short sellers, our paper makes three distinct contributions to the field. First, we innovatively classify different types of agents in the social network by empirically adapting Pedersen's definitions to our data. This provides an important foundation for future empirical analyses focused on different types of agents. Second, following the theoretical guidance of Pedersen (2022), we thoroughly examine the *joint dynamics* of different types of agents' beliefs, price formation, and trading behaviors of retail investors and short sellers in the social network. This provides a comprehensive perspective for better understanding the role of social media in the financial market. Third, we provide many novel and unique empirical findings, such as how influences interact with social media tones in predicting belief, price and trading formations, and how short-sellers respond to social media, given contingencies related to social media tone and influence. In addition, we leverage both time periods before and after the GME episode, showing that the interaction between social media tones and shorting activities changes from the pre-GME period to the post-GME period. This suggests that short sellers become more cautious about going against social media tone after the GME episode highlights the costs and risks of doing so. Our unique insights and timely answers to various important questions regarding the interactions of multiple participating parties are significantly helpful for all market participants.

2. Pedersen's Model and Empirical Hypotheses Development

Pedersen (2022) is one of the first theoretical studies to provide a comprehensive framework for understanding social networks and their implications for asset prices. Here we introduce the assumptions and propositions in Pedersen's model and develop our empirical hypotheses accordingly. We refer readers to the original paper for more details on derivations.

2.1 Assumptions and Model Setup

Pedersen (2022) makes two assumptions about assets and signals in the economy. First, there is one asset with a supply of shares *s*. The asset's fundamental value is $v + u(t)$, where $u(t)$ is a publicly observed random walk that has an innovation of constant variance σ_u^2, and v is an unobserved random variable that investors try to learn about. Second, the economy has N investors who communicate with each other. At time 0, each investor i is endowed with a signal about the value *v*, i.e. $x_i(0) = v_i$. All signals collectively reflect the true value of *v*, $v = \sum_{i=1}^{N} k_i v_i$, where individual investors' weights, k_i, sum up to one, or $\sum_{i=1}^{N} k_i = 1$. The objective of the model is to form the dynamics of belief formation, price discovery, and trading behaviors before the value is revealed.

Pedersen assumes an exogenous social network with different types of agents interacting with each other. There are three types of investors: rational, fanatic and naïve. Rational investors learn from everybody in the first round, gaining information on v and maintaining their opinions in later rounds. Fanatic investors learn only from themselves, and do not change their opinions. Both rational and fanatic investors do not change their views after the first round, so these two types of agents are labeled as "hardheaded". Naïve investors constantly learn from investors they follow, and update their views accordingly. At each time t, everyone states their current views, collected in the $N \times 1$ vector $x(t) = (x_1(t), \dots, x_N(t))$.

Investors' belief update is modeled as: $x(t+1) = Ax(t)$ (a VAR setup), where the $N \times N$ weighting matrix A has each i-th row summing up to one, or $\sum_j A_{ij} = 1$. In other words, the social network is characterized by matrix A, which captures how agent i's view is influenced by other investors. Suppose we use subscript "h" to denote hardheaded agents (i.e., rational and fanatic agents), and subscript "n" to denote naïve agents. Then A_{nh} is the matrix that defines how naïve agents listen to hardheaded agents, and A_{nn} is the matrix that defines how naïve agents listen to each other.[1]

In this network, if one agent is influential, depending on her type, she can affect others through two channels: thought leadership and influencer value. Thought leadership measures how much one agent's view attracts other's attention. For instance, rational and fanatic agents can affect naïve agents through thought leadership. Influencer value measures how much attention naïve agents attract from other naïve agents. That is, naïve agents do not have thought leadership since

[1] To highlight the social network effect on belief formation, price discovery and trading behaviors, Pedersen (2022) abstracts away certain aspects of real-world data. For example, the model assumes that the realization of the value of the firm does not change, nor do investors receive new information. We relax many of these model assumptions based on our data observations to design empirical proxies that better align with Pedersen's theoretical predictions.

their views are affected by others, but the connectedness among naïve agents themselves, measured by influencer value, also affects the information dynamics in the network. Notice that thought leadership and influencer value both capture how influential each agent is in the social network.

2.2 Belief Formation Dynamics

In the model's equilibrium, every naïve agent's view is a convex combination of views of rational and fanatic agents (Proposition 1 of Pedersen, 2022):

$$x_n(t) \rightarrow (I - A_{nn})^{-1} A_{nh} x_h, \tag{1}$$

This proposition indicates that the long-run views of naïve investors reflect the views of rational and fanatic investors, weighted by their relative influences in the network. We develop two testable hypotheses based on Proposition 1:

H1: Naïve investors' views can be predicted by views from rational and fanatic investors.

H2: The more influence rational and fanatic agents have, the stronger the predictive power of their views on the views of naive agents.

2.3 Price Formation Dynamics

As agents trade following their beliefs after learning in the social network, equilibrium asset price for period *t* is determined as follows (Proposition 4 of Pedersen 2022):

$$p(t) = p_r(t) + p_n(t) = p_r(t) + function(a, b, \bar{x}_n(t), \bar{x}_f(t), x_r). \tag{2}$$

The equilibrium price $p(t)$ has two components, the rational price $p_r(t)$, formed in the special case where all wealth is in the hands of rational investors, and the network price, $p_n(t)$, which is a function of the relative wealth of naïve investors (parameter a), the relative wealth of fanatic investors (parameter b), the average view among naïve investors $\bar{x}_n(t)$, the average view among fanatic investors $\bar{x}_f(t)$ and the rational view x_r. In Proposition 5, Pedersen defines the long-term equilibrium price as a function of agent views, which is positively affected by agent influence. In other words, the contribution of agent j to long term price is higher if she has higher influence. Following propositions 4 and 5, we develop two testable hypotheses:

H3: Agent views from the social media network predict future stock returns.

H4: Agents with higher influence have stronger predictive power on stock returns.

2.4 Trading Dynamics

In Propositions 7 and 8, Pedersen (2022) discusses trading behaviors of various investors. Investors' demand for shares shifts as they observe the expected network price, $p_n(t)$, change. For instance, if investors anticipate that social media views drive up the price, causing it to be much higher than the true value of the stock, they optimize their trading behaviors by weighing the benefits and costs of riding a potential positive bubble. If prices remain high for the holding horizon of the investors, and the benefits of riding the bubble outweigh the benefits of bursting the bubble, the investors might choose to ride the bubble, or at least not to burst the bubble, and vice versa.

Here we choose to examine the trading behaviors of two important groups of investors: retail investors and short sellers. Retail investors are generally viewed as less sophisticated, who might follow the social media tone; while short sellers are generally believed to be informed and rational, who may ride or burst the bubbles depending on the expected short-term and long-term price dynamics discussed above. With the increasing importance of social media and its significant influences on investors' views and prices, we form the following testable hypotheses following Propositions 7 and 8:

H5: Social media views predict future retail flows, and views from more influential agents have stronger predictive power on retail flows.

H6: Whether short sellers ride or burst social-media-induced bubbles depends on the costs and benefits of doing so.

3. Data and Empirical Method

3.1 Reddit data

Reddit is a social media platform with 100,000+ communities, or "subreddits", each of which focuses on a different topic.[2] In this article, we focus on one forum on Reddit, "WallStreetBets", on which participants discuss the trading of stocks and equity options. As of February 2025, the subreddit has a total of 18 million subscribers, making it one of the largest social media forums for financial news and trading strategies.

Previous literature examines several major social media platforms, including X (formerly Twitter), StockTwits, Seeking Alpha, and Reddit's WallStreetBets.[3] We choose to focus on Reddit's "WallStreetBets" because it has the following two distinctive features. First, unlike X, StockTwits, or Seeking Alpha, users on Reddit can view all posts on the front page without needing to subscribe to or follow individual accounts. This unrestricted access enhances the spread of posts, making Reddit an ideal platform for studying the dissemination of investor opinions. Second, comments on X or StockTwits can be accessed independently of the original post, whereas comments on Reddit can only be viewed through the original post. This thread-based structure enables investors to engage in focused discussions, further allowing them to influence one another.

[2] These communities, which we refer to as "forums", attract more than 52 million daily active users, and more than 50 billion monthly views. These numbers clearly show that Reddit receives substantial attention and is an influential social media platform.

[3] For example, Bianchi et al. (2023) and Bianchi et al. (2024) use Twitter; Cookson and Niessner (2020) and Cookson et al. (2023) use stocktwits; Chen et al. (2014) and Farrell et al. (2022) use Seeking Alpha; and Bradley et al. (2024) use Reddit. Other social media platforms, such as Yahoo! Finance and myForexBook, have also been studied in previous literature, including Antweiler and Frank (2004), Das and Chen (2007), and Heimer (2016).

That is, these features may make inexperienced new investors more susceptible to being influenced by opinions shared on the forum.[4]

We collect all submissions and comments on this subreddit between January 1, 2020, and December 31, 2023. Each submission, created by Reddit users, has its own web page with comments displayed beneath it. We assign each submission or comment to specific companies using tickers and collect unique Reddit IDs to identify authorship. We exclude bot accounts and moderators from our sample since their posting behaviors differ from those of typical users. Altogether, we have 10,274,035 agent*stock*day observations from Reddit data.

Adapting Pedersen's definitions to our empirical data, we identify hardheaded agents on Reddit as users who express more opinions (so they can be heard) and whose tones remain stable over a period of time. Specifically, we define that an agent k is a hardheaded agent for firm i in the week t if the following conditions are met: 1) agent k posts more submissions and comments about firm i than 95% of all other agents over the week; 2) agent k's posts have stable tones, with either non-positive or non-negative tones for at least 75% of his posts during the week. Once agent k is identified as hardheaded for firm i, they remain hardheaded for firm i for as long as their tone remains stable, as specified in condition (2). [5]

Rational investors base their views on fundamental, value-relevant information, while fanatic investors do not. We distinguish between them based on whether they use value-relevant information in their posts, using a manually created dictionary of such words.[6] For instance,

[4] Contemporaneous Reddit papers include Betzer and Harries (2021), Diangson and Jung (2021), Long, Lucey, and Yarovaya (2021), Bradley et al. (2024), Lyocsa, Baumohl and Vyrost (2022), Vasileiou, Bartzou, and Tzanakis (2022), Strych and Reschke (2022).

[5] We consider alternative ways of identifying different types of agents by varying the parameters in both conditions 1 and 2, and the results, reported in Section 5.2, stay qualitatively similar.

[6] We first create a dictionary of the most frequently used words in Reddit submissions and comments, then manually inspect and select those related to firm financial and accounting information. In Appendix A Panel B, we present the list of value-relevant words. We also include fundamental words from Bradley et al. (2024) to form a more comprehensive dictionary of value-relevant words.

"YOLO" (You Only Live Once) is often used on Reddit to describe bold, risky investments. Pedersen (2022) also discusses how social media investors use YOLO to signal stubborn and extreme views. Thus, we define rational agents as hardheaded agents who use at least one value-relevant word and do not mention "YOLO" in their posts for the week. The other hardheaded agents are classified as fanatics. Reddit users who are not hardheaded are classified as naïve investors.

We create indicator variables for rational, fanatic, and naïve agents, which equal 1 if the Reddit user belongs to the respective agent category, and 0 otherwise. Table 1 Panel A shows that 2.57% of agents are rational, 6.80% are fanatics, and 90.63% are naïve investors, indicating that most Reddit users on r/wallstreetbets behave like naïve investors. To validate our measure, we look at specific individuals who are known to be fanatics. Pedersen (2022) uses Keith Gill as an example of a fanatic for his unwavering belief in GameStop. We confirm that Keith Gill is classified as a fanatic for GameStop for 94.38% of the time during our sample period, supporting our measurement approach.

We measure each agent's beliefs using the tone of their submissions and comments, identified using the modified Loughran and McDonald (LM) dictionary. Since r/wallstreetbets has its own lingo (e.g., emojis, slang, memes), we modify the LM dictionary to better capture this language. Our modified LM dictionary includes a custom Reddit dictionary of the 1,000 most frequently used words and 100 most frequently used emojis on r/wallstreetbets, each manually categorized as positive, neutral, or negative based on context.[7] For firm *i* on day *d*, for each submission/comment *m*, we first compute,

$$SubmissionTone_{imd} = \frac{\#\,of\ positive\ words/emojis_{imd} - \#\,of\ negative\ words/emojis_{imd}}{total\ \#\,of\ words/emojis_{imd}}. \tag{3}$$

[7] We provide a list of these jargons and emojis and their assigned sentiment in Appendix A.

The submission tone ranges between -1 and 1, and higher tone indicates more positive views. For firm-level tone across agent types, we average across all submissions/comments (M_{ikd}) of an agent k for all agents (K_{ild}) within each agent type l for the stock i on day d:

$$AgentTone_{ild} = \frac{1}{\sum_{k=1}^{K_{ild}} M_{ikd}} \sum_{k=1}^{K_{ild}} (\sum_{m=1}^{M_{ikd}} SubmissionTone_{imd}). \tag{3'}$$

Since not every firm has social media posts every day, we average agent tones across all days of a week t and compute agent-type*stock*week level variables.[8] [9] Following this method, we compute $FanaticTone_{it}$, $RationalTone_{it}$, and $NaiveTone_{it}$ for stock i on week t.

The importance of an agent in the social network is measured by her influence, defined by how much attention the agent attracts from other users.[10] We define social networks for every firm-week pair in our sample. The social network for firm i on week t includes all Reddit users discussing firm i during week t. User k is connected to user j if k captures j's attention. Since we do not have data on who views k's submissions or comments, we use the number of users who comment on k's posts as a proxy for how many users pay attention to k, which is also the influence measure of agent k. A higher value of the influence variable means the agent has more direct commenters and thus attracts more people's attention.[11]

To compute agent-type*stock*week level influence measures, we sum across the number of commenters of all agents (K_{ilt}) belonging to an agent group l for every stock i on week t as $NCommenter_{ilt} = \sum_{k=1}^{K_{ilt}} NCommenter_{ikt}$. We normalize this variable to compute our influence

[8] As Cookson et al. (2024) note, social media is an abundant but noisy source of data on retail investors. Averaging the tone on a weekly basis helps filter out this noise, allowing for more precise estimates of social media tone.
[9] We also report results using influence-weighted tone, where we weight an individual's agent tone by influence, in Table 6 Panel B. The main results remain robust.
[10] As mentioned earlier, Pedersen's model mathematically defines two types of influence, thought leadership and influencer value. For parsimony in empirical estimation, we do not separately estimate thought leadership and influencer value.
[11] In Section 5.4, we also use an agent's PageRank score in the network as a proxy for influence and report results in Table 6 Panel C. Our main results remain robust under this alternative definition.

measure in two steps. First, we compute the natural logarithm of one plus the raw measure, to address the skewness in its distribution. Second, we transform the logged number to a domain of [0,1] for ease of interpretation. That is, for an agent group *l* for every stock *i* on week *t,* influence is calculated as,

$$Influence_{ilt} = \frac{log(1+NCommenter_{ilt}) - \min_{i}\{log(1+NCommenter_{ilt})\}}{\max_{i}\{log(1+NCommenter_{ilt})\} - \min_{i}\{log(1+NCommenter_{ilt})\}}. \tag{4}$$

A higher value of this measure indicates that this agent type for a firm receives more attention from all agent types, thereby exerting higher influence in the social network. For cross-sectional comparisons, we split our sample firms each week into high-influence and low-influence networks. Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. [12]

Table 1 Panel B presents summary statistics on social media activity. The average tones for rational, fanatic, and naïve agents are 0.0016, 0.0025, and 0.0052, respectively, indicating that naïve agents have the most positive tones, while rational investors have the least. The average influence values for rational, fanatic, and naïve agents are 0.0125, 0.0147, and 0.0202, respectively, with naïve agents having the highest influence, possibly due to their larger numbers. Tone measures have low correlations, around 0.27, suggesting that different agent groups have different views. Influence variables have high correlations, mostly above 0.70, indicating that influence structures are stable across stocks and over time.

To provide a heuristic understanding of the social media measures, we plot their time series in Figure 1. Panels A and B show the proportion of the three user types over time, with a slight upward trend in fanatic agents from 2020 to 2021. Panels C and D provide the average tones and

[12] We also use the 90th percentile of influence for each agent type (rational influence, fanatic influence, and naïve influence) separately to divide our sample, and the results are similar. Our main results are also robust when we use the 95th percentile or 85th percentile.

the number of commenters (our influence measure before normalization) for each user type. All user types show a dramatic increase in tone and the number of commenters in January 2021, especially among naïve agents. These patterns indicate a significant rise in social media activity and bullish sentiment in January 2021, coinciding with the enormous upswing in GameStop's price.

3.2 Data for returns, retail flows, and shorting flows

Stock data are obtained from CRSP. We retain only common stocks (CRSP share code equal to 10 or 11) listed on NYSE, NYSE MKT, or NASDAQ. We cross-match the CRSP data with Reddit data using ticker symbols. To ensure sufficient Reddit discussion, we limit the sample to firms with at least 10 submissions or comments during the sample period. We identify retail investors using sub-penny price improvements in FINRA trade data following Barber et al. (2023), and calculate the daily retail trading for stock *i* on day *d* as $\frac{TotalRetailBuyVolume_{id} - TotalRetailSellVolume_{id}}{TotalRetailBuyVolume_{id} + TotalRetailSellVolume_{id}}$. For short sellers, we define their activity, following Boehmer et al. (2022), using the days-to-cover ratio (*DTCR*) as the total number of shares on loan from IHS Markit scaled by the daily trading volume, $\frac{QuantityOnLoan_{id}}{TradingVolume_{id}}$. We scale the variable by dividing it by 10,000. We calculate the weekly average of daily returns, daily retail trading, and daily shorting flows. In total, our sample has 472,765 stock*week observations.

3.3 Empirical method

In Pedersen's model, agents' views both predict and are predicted by other agents' views; agents' beliefs are predicted by stock prices and at the same time, have predictive power on prices; and investors' trading behaviors are also interdependent with agents' beliefs and stock returns. To estimate the dynamic interactions among agent tones, stock prices, and investors' trading behaviors, we follow Hendershott et al. (2015) and Hollifield et al. (2017) and adopt the panel vector autoregressions (PVAR) approach.[13] Compared to alternative empirical methods, such as Fama-MacBeth regression and panel regression, PVAR, with its matrix setup, is more consistent with the setup in Pedersen's theory and better at capturing intricate interactions and dynamics for multiple moving components within a system simultaneously.

To be specific, we specify the PVAR system as follows,

$$y_{i,t} = \sum_{l=1}^{L} A_l y_{i,t-l} + \gamma_t + \alpha_i + \varepsilon_{i,t}. \quad (5)$$

Here $y_{i,t}$ is the vector of dependent variables. Given Pedersen's framework, we set $y_{i,t} = \left(RationalTone_{i,t}, FanaticTone_{i,t}, NaiveTone_{i,t}, Return_{i,t}, RetailFlow_{i,t}, ShortFlow_{i,t}\right)'$. Matrix A_l is a coefficient matrix for lag l, and $l = 1, \ldots, L$, is lag length. Since the dependent variables may be related to certain firm characteristics and time periods, we include γ_t as time fixed effects, and α_i as firm fixed effects. We estimate the coefficients in (5) using the generalized method of moments (GMM), and the standard errors are double clustered by firm and date.[14]

The PVAR estimation provides three rich sets of statistics to facilitate the interpretation of the results. The first set includes the estimates of the A_l matrix, which shed light on the predictive patterns among various variables and the parameters' statistical significance. The second set contains Granger causality tests, which examine whether the past values of one variable provide statistically significant information about the future values of another variable. For example, let $y(m)$ be the *m*-th element of vector *y*, and let $A_l(m,n)$ be the element in the *m*-th row and *n*-th column of matrix A_l, then variable $y(n)$ is considered to Granger cause variable $y(m)$, if the elements $A_1(m,n)$, $A_2(m,n), \ldots, A_L(m,n)$ are jointly significant according to a Wald test. The third set is the impulse response functions (IRF) associated with PVAR, which describe how a

[13] Lee et al. (2018) and Stefano, Corradin and Maddaloni (2020) also adopt PVAR to capture financial variable dynamics.
[14] More details about the estimation is provided in Appendix B.

dependent variable responds to a one-time shock from one independent variable. The IRFs illustrate the economic magnitude of the relation between variables. Specifically, let $y_{i,t-1}(n)$ be the *n*-th element of vector $y_{i,t-1}$ (the independent variable), the impulse response of variable $y_{t+k}(m)$ (the dependent variable) measures how it responds to a one standard deviation change in $y_{i,t-1}(n)$, during week $t+k$. We choose $k = 1, \ldots, 10$ to capture the dynamics over 10 weeks. We generate the 5% confidence bounds for the IRFs using Monte-Carlo simulations with 1000 draws.

Previous literature points out two potential concerns regarding the PVAR approach. The first is that the estimation results of PVAR may be sensitive to the choice of lag length. To cope with this concern, we first compute the optimal length of lag using the Bayesian information criterion (BIC), and find the optimal lag length is 1. The second concern with PVAR is that it only allows linear relations in the system and cannot capture nonlinear relation among variables. To overcome this restriction in our case, we conduct subsample analyses by splitting the sample into different groups to allow flexible nonlinear relations among variables for different subgroups. Overall, the above two concerns regarding the PVAR do not significantly affect our results.

4. Empirical Results

4.1 Social Network Dynamics and Belief Formation

In this section, we examine how different agents' opinions propagate through the social network and how these agents form their beliefs, as specified in H1 and H2 in Section 2. If H1 is true that naïve investors' views are predicted by rational and fanatic investors' views, we expect the coefficients linking current naïve views and past rational and fanatic views to be statistically significant.

Table 2 Panel A presents the first 3 rows of the matrix A_1 in Pedersen (2022), which capture the dynamics of views among three groups of investors for the whole sample. All nine coefficients are positive and significant, suggesting that past tones, no matter whether they are from rational, fanatic or naïve agents, are all positively associated with future tones. We conduct the Granger causality tests and report the p-values at the bottom of the table. The p-values for all past tones Granger causing current tones are 0.0%, which indicates significant Granger-causal relations. To examine H1, we direct our discussion to column III, which describes how naïve tone in week t is related to rational and fanatic tone in week t-1. The coefficients on RationalTone (t-1) and FanaticTone (t-1) are 0.1387 (t-stat=21.55) and 0.0840 (t-stat=20.14), implying that both rational and fanatic tones are important for forecasting naïve tone, which supports H1.

To heuristically understand how different agents' tones within a dynamic system interact with each other, we plot the IRFs in Figure 2. The IRFs show the next 10-week reaction of each response variable corresponding to a one standard deviation shock to each impulse variable. Panels A and B present the response of shocks to rational tone and fanatic tone on naïve tone for the next 10 weeks. For a one standard deviation shock to rational tone or fanatic tone, the naïve tone increases by 0.1% for the first week. Given that the mean of the naïve tone is 0.005 from Table 1, this effect is economically meaningful. The response of naïve tone to rational and fanatic tone shocks gradually dies out after about 4 weeks.

Hypothesis H2 predicts that the greater the influence of a group of agents, the more likely they are to predict the tones of the other agents. To test this hypothesis, we estimate specification (5) for two subsamples, the high influence group and the low influence group, separately, where the separation of the two groups is based on the 90th percentile of overall influence (total number of commenters) across all firms for each week, as described in Section 3.1. If H2 is true, then past rational and fanatic views should have stronger predictive power on future naïve views in the high influence group than in the low influence group.

We report the estimation results for the high influence group in Panel B of Table 2. Compared to the coefficients in Panel A, the coefficients in Panel B are all positive and significant, with slightly higher magnitudes, suggesting that the interaction among tones are more pronounced in the high influence sample. For instance, in column III, the coefficients on RationalTone (t-1) and FanaticTone (t-1) in Panel B are 0.1556 (t-stat=16.48) and 0.1191 (t-stat=16.04) for the high influence subsample, which are higher than the coefficients in Panel A.

The estimation results for the low influence group are reported in Panel C of Table 2. Interestingly, 6 of 9 coefficients are positive and significant in Panel C, and the magnitudes are generally smaller than those in Panels A and B, implying that the interaction among tones is weaker for the low influence sample. For instance, again for column III, the coefficients on RationalTone (t-1) and FanaticTone (t-1) are 0.0854 (t-stat=8.04) and 0.0266 (t-stat=4.44), substantially smaller than those in Panel B. Granger causality tests at the bottom provide consistent implications.

The drastic differences between high and low influence samples are vividly plotted in Figure 2 Panels C-D and Panels E-F. For the high influence subsample in Panel C, the reaction of naïve tone in response to a one standard deviation shock to rational tone is a 0.40% increase for the following week, and it remains significant for the next 8 weeks. However, in the low influence subsample in Panel E, the response of naïve tone to the same rational tone shock on day 1 is merely a 0.03% increase, and it quickly drops to zero after 2 weeks. Similar patterns can be observed for fanatic tone in Panels D and F. Clearly, the magnitude of the response is much larger in the high influence subsample than in the low influence subsample. Overall, these patterns suggest that rational tone and fanatic tone have stronger predictive power on naïve tone when rational and fanatic agents have higher influence, which is consistent with H2.

To summarize, we document empirical support for H1 and H2, regarding belief formation in a social network. Views from rational and fanatic agents are strong predictors of future naïve agent views, and the predictive power of these agents' tones on future naïve tone is stronger when they have high influence.

4.2 Social Media Views and Price Discovery

In this section, we test H3 and H4 and examine whether and how the beliefs of different agents, along with their influences, predict next-week stock returns in the dynamic system. If H3 is true, social media views predict future returns, and the coefficients linking current returns to past agent views are expected to be significant.

Table 3 reports the PVAR coefficients connecting past agent views and current returns. For the whole sample results in column I, the coefficients on RationalTone (t-1), FanaticTone (t-1) and NaiveTone (t-1) are 0.0022 (t-stat=0.58), 0.0075 (t-stat=1.89) and 0.0068 (t-stat=2.57), respectively, suggesting that past fanatic tone and naïve tone contain predictive information about next-week returns. Also, the Granger Causality tests at the bottom of the table show that fanatic tone and naïve tone Granger cause future returns with a p-value of 5.8% and 1.0%, respectively. That is, fanatic and naïve tones, but not rational tone, predict future returns with statistical significance, which supports H3. In Panels A to C of Figure 3, we present the corresponding IRFs, showing the response of shocks to rational, fanatic, and naïve tones on returns for the next 10 weeks. For economic magnitude, a one standard deviation shock to fanatic tone or naïve tone is associated with a small 1 bps increase in the daily average return for the following week.

Hypothesis H4 predicts that the interaction among agents' tones and returns is stronger in networks with higher influence. We report the parameter estimates for the high influence sample and low influence sample in columns II and III of Table 3. For the high influence subsample in column II, the coefficients on RationalTone (t-1), FanaticTone (t-1), and NaiveTone (t-1) are 0.0131 (t-stat=2.40), 0.0222 (t-stat=3.52), and 0.0685 (t-stat=8.36), all much larger and more significant than those in Panel A. For Granger causal relations, rational tone, fanatic tone, and naïve tone all positively Granger cause next-week returns in the high influence subsample with p-values of less than 5%. Interestingly, for the low influence subsample in column III, the coefficients on RationalTone (t-1) and FanaticTone (t-1) are insignificant, while the coefficient for NaiveTone (t-1) is negative and significant, which is in sharp contrast with the positive and significant coefficients in column II. These results show that agents' tones positively predict future stock returns only when the agents exert high influence in the network, suggesting that influence is a key determinant of the relation between agent tones and stock returns. The negative prediction of naïve tone in the low influence subsample may be explained by the fact that when naïve tone lacks influence, it may lead to a temporary increase in price that is followed by a subsequent reversal.

Panels D-F and G-I of Figure 3 report the responses of returns to shocks in different tone variables for the high and low influence groups, respectively. For the high influence subsample in Panels D-F, a one standard deviation shock to rational, fanatic, naïve tone is associated with a higher daily average return of 5, 8, 20 bps for the following week, respectively, and the effects remain significant for the next 6 weeks. The magnitude of the response is much larger for the shocks from naïve tone, highlighting the importance of naïve tone in predicting future returns for the high influence sample, possibly because agents on social media are predominantly naïve agents (as in Table 1). In contrast, for the low influence group in Panels G-I, the shock to rational tone or fanatic tone is associated with only minor effects, which last for just 1 week.

To summarize, we find supportive evidence for hypotheses H3 and H4. Agent tones significantly predict future price movements. More importantly, agents' tones in a high-influence network are more predictive of future returns than in a low-influence network.

4.3 Social Network and Trading Dynamics of Retail investors

H5 states that social media views predict retail trading, and views from more influential agents have stronger predictive power on retail flows. If retail investors follow social media tones to trade, the coefficients linking current retail flows to past agent views are expected to be positive and significant.

We report the coefficients linking past social media tones and future retail trading in Table 4 Panel A. From the whole sample estimates in column I, the coefficients on rational and fanatic tones are insignificant, while the coefficient on NaiveTone (t-1) is 0.0656 and is significant with a t-stat of 3.55, suggesting that retail investors are more bullish when naïve tone is higher. Naïve tone also Granger causes retail flows with p-value of less than 1%, supporting the first part of H5. The IRFs in Figure 4 Panels A-C show that a one standard deviation shock to naïve tone is associated with a 0.1% increase in retail flows for the next week, with the effect gradually dying out after 2 weeks.

For the second part of H5, if influence is important in affecting the predictive power of agent tones on retail order flows, we expect a stronger predictive relation between past agent views and future retail flows in the high influence group. Column II in Table 4 shows that the coefficient on NaiveTone (t-1) is 0.1055 (t-stat=4.86) for the high influence group, which is much higher than the 0.0497 (t-stat=2.21) for the low influence sample in column III, which supports H5. Granger causality tests provide similar inferences.

Figure 4 Panels D-F and G-I show the IRFs for the high and low influence subsamples, respectively, to further compare the economic magnitudes. In the high influence subsample, a one standard deviation shock to naïve tone is associated with a 0.4% increase in retail flows for the next week, with the effect lasting for more than 2 weeks. In the low influence subsample, the same shock results in only a 0.1% increase for retail flows, lasting just 1 week. These results suggest that naïve tone has stronger predictive power for future retail flows in high influence networks. When naïve agents exert influence, their positive views lead to more bullish retail trading. Overall, this supports our hypothesis that agent tones have stronger predictive power for retail flows in more influential networks.

Given prior findings that marketable retail flows significantly and positively predict future stock returns, how does the predictive power of retail flows change with social media activity? To answer this question, we examine how past retail flows predict future returns. Table 4 Panel B shows that the coefficient on RetailFlow (t-1) is 0.0012 (t-stat=4.21), and the Granger Causality test shows that retail flows Granger cause future returns with a p-value of less than 1%. For economic magnitude, Figure 4 Panel J shows that a one standard deviation shock to retail flows is associated with a 2 bps increase in the daily average return for next week, lasting for 2 weeks. These results show that retail flows positively predict future returns, even after we include social media variables in the dynamic system, suggesting that retail flows contain more relevant information for future price movements than what is contained in social media tones.

More interestingly, how does the return predictability of retail flows change for firms with different level of influence in the network. Columns II-III of Table 4 Panel B show that the coefficient on RetailFlow (t-1) is 0.0088 (t-stat=2.37) for the high influence subsample, which is substantially higher than the coefficient of 0.0009 (t-stat=3.72) for the low influence subsample.

Both Granger causality tests are significant at 5%. Figure 4 Panels K-L show the IRFs for the high and low influence subsample to further compare the economic magnitude. For the high influence subsample in Panel K, a one standard deviation shock to retail flows is associated with a 10 bps increase in the daily average return for week 1, with the effect fading after 4 weeks. Meanwhile, for the low influence subsample in Panel L, the same shock is associated with a 1 bps increase for the next week, which becomes insignificant by week 2. Overall, these results indicate that retail flows have stronger predictive power for future returns for firms in high influence networks.

4.4 Social Network and Trading Dynamics of Short Sellers

Short-sellers are generally considered to be more sophisticated than retail investors. H6 states that the relation between short-selling and social media might be more complicated, which might depend on the costs and benefits of short-selling. Meanwhile, these costs and benefits are likely correlated with agent influence. Therefore, if short sellers understand the social network's effect on prices and trade accordingly, the relation between agent views and future shorting flows may differ across networks with high and low influence.

In Table 5 Panel A, we present the coefficients of past tone variables predicting future shorting flows. The whole sample result in column I shows that the coefficients on agent tones are all negative and statistically significant, implying that social media tones do not predict future shorting flows. The Granger causality p-values at the bottom and the corresponding IRFs (Figure 5 Panels A-C) provide similar inferences.

Regarding whether the dynamics between agents' tones and shorting flows vary with agent influence, we report the results in column II and III of Table 5 Panel A. For high influence subsample in column II, the coefficients on FanaticTone (t-1) and NaïveTone (t-1) are -0.0004 (t-stat=-2.17) and -0.0009 (t-stat=-2.18), suggesting that short sellers significantly reduce their shorting activities in response to positive fanatic or naïve tones. The corresponding p-values for the Granger causality tests are both lower than 5%. In contrast, for the low influence subsample in column III, none of the coefficients is significant.

From Figure 5 Panel D-F, the response of a one standard deviation shock to fanatic or naive tone is a decrease of 0.03% or 0.04% in shorting flows for the next week in the high influence sample, with the effect staying significant over the next 4 weeks. In contrast, Figure 5 Panel G-I shows that for the low influence subsample, the responses of a one standard deviation shock in agent tones on shorting flows are all insignificant.

Comparing the results for the high and low influence subsamples, we find that the relation between short sellers' trading behavior and agent tones indeed varies with agent influence. In the high influence subsample, agents' views are negatively related to shorting flows, consistent with short sellers shying away from shorting (i.e., riding the bubble) when agents' influence is higher. However, in the low influence subsample, the negative relation is not significant, suggesting that short sellers may not feel threatened by positive agent tones when their influence is low. Therefore, short sellers may not decrease shorting as they believe they could profit from bursting the short-term bubbles. Considering the earlier results in section 4.2, which show that agent tones positively predict future returns only in the high influence subsample, it is reasonable to infer that short sellers choose to ride the bubble, rather than burst it, when agent influence is high and when higher social media tones are significantly connected to higher future returns.

That is, higher Reddit tones are associated with lower future shorting flows for the high influence subsample, which supports H6, in the sense that short-sellers trade by weighting costs and benefits of short-selling and considering the connections between social media tones and future returns.

Previous research finds that shorting flows negatively predict future stock returns (Boehmer, Jones, and Zhang, 2008). Given the current results showing that higher Reddit tones deter shorting flows in the network with high influence, how does the predictive power of shorting flows change with social media activity? We first present the coefficient in the column I of Table 5 Panel B, when we use short flows to predict future returns for the whole sample. The coefficient on ShortFlow (t-1) is -0.0011 (t-stat=-6.22), and shorting flows Granger causes future returns with a p-value of less than 1%. For economic magnitude, Figure 5 Panel J shows that a one standard deviation shock to shorting flows is associated with a 1 bps decrease in the daily average returns for the next week. These results show that shorting flows negatively predict future returns when including social media variables.

We are more interested in understanding how the return predictability of shorting flows changes with agent influence in the network. For the high influence subsample in column II of Table 5 Panel B, the coefficient on ShortFlow (t-1) is -1.7969 (t-stat=-3.53), which is much larger effect than -0.0010 (t-stat=-9.28) for the low influence subsample in column III. Shorting flows Granger causes returns with negative relations in both the high and low influence subsample.

To better compare the economic magnitudes, the corresponding IRFs are provided in Figure 5 Panel K-L. For the high influence subsample, a one standard deviation shock of shorting flows is associated with a 40 bps decrease in the daily average return for the following week. While for the low influence subsample, the same shock is associated with only a 1 bps decrease in returns. The results show that shorting flows are connected to a much larger decrease in future returns in the more influential network. This is quite intriguing, because earlier results suggest that short sellers on average shy away from shorting when they worry that more positive social media tones lead to higher returns and short squeeze risk, when agent influence is high. This particular pattern between shorting flows and future returns suggests that short sellers probably carefully consider the costs and benefits of shorting under heightened social media activity, and they only short if they are convinced that the benefits outweigh the costs, and hence stronger predictive power of shorts for future returns.

5. Further Discussion and Robustness Checks

5.1 Dynamics in Different Time Periods

In this subsection, we provide further discussion by considering how social media dynamics change in different time periods. Since the GME event attracts substantial investor attention and makes investors aware of social media's impact on prices, the dynamics might be different before, during and after the GME episode. Therefore, we divide the whole sample into three time periods: pre-GME (2020/1-2020/12), during-GME (2021/1-2021/2), and post-GME (2021/3-2023/12). After the GME episode, all investors become aware of the social network effect on prices, and they might tend to follow it rather than go against it. Therefore, we expect that the positive predictive power of agent tones for future returns to be stronger. Similarly, the relation between social media tones and shorting flows might be more negative during the post-GME period, after short sellers witness or experience the short squeeze and feel threatened by social media sentiments.

The estimation results for different time periods are reported in Table 6 Panel A.[15] For return prediction columns I-III, the coefficient on NaïveTone(t-1) is -0.0138 (t-stat=-2.18) for the pre-GME period, -0.0325 (t-stat=-1.34) during-GME period and 0.0130 (t-stat=4.53) for the post-GME period. That is, the naïve tone negatively predicts future returns in the pre-GME period, suggesting that the market perceives naïve agents as noise traders whose tones convey incorrect

[15] Here we choose to present return and short flow predictions. The complete results for subperiod analysis are presented in Appendix C Panel A-Panel F.

information about future returns. Interestingly, the coefficient switches sign in the post-GME period, suggesting that the market begins to view naïve tones as positive signals for future returns. This shift likely reflects the market's recognition of the impact that social media has on prices during the GME episode. Coefficients on rational tones and fanatic tones are mostly insignificant.

For shorting flows in columns IV-VI, the coefficients on agent tones are all positive and significant in the pre-GME and during-GME period, indicating that short sellers may view naïve investors as noise traders whose actions could lead to temporary positive bubbles and thus increase shorting as they believe they could profit from bursting the short-term bubbles. Surprisingly, all the coefficients become negative and two out of three are significant in the post-GME period, possibly because short sellers take social media tones seriously and become more cautious about going against them after the GME short squeeze.

5.2 Alternative Measures for Agents and Tone

In this subsection, we present robustness tests using different agent classifications and measures for tone in Table 6 Panel B. We start by considering three alternative cases to identify rational and fanatic agents and construct agent tones. For the first case, we identify hardheaded agents as those who post more than 99% of all other agents, instead of 95% as in the main results. We present the estimation results in columns I-II. For the second case, we require that hardheaded agents' posts have the same sign in tones (either positive or negative) for 100% of their posts, instead of only 75% of their posts during the week (columns III-IV). For the third case, we compute an influence-weighted tone to highlight the importance of agent influence in social networks, rather than computing average tones with equal weights across all individuals (columns V-VI). Furthermore, we control for common tones from other social media including X (formerly Twitter), StockTwits, and Seeking Alpha using data from Cookson et al. (2024) (columns VII-VIII). In almost all cases, rational tone and fanatic tone positively predict future naïve tone. Moreover, fanatic tone and naïve tone positively predict future returns. To summarize, the results remain largely robust to changing our definitions of hardheaded agents, or using a tone measure that is weighted by agent influence, or controlling for common tones from other social media platforms.

5.3 Use Traffic to Proxy for Social Media Activity

We consider an alternative measure for social media activity. Following Da et al. (2011), we compute a raw measure of general attention from agents, using the number of submissions and comments posted by each agent type. To reduce skewness in the raw data, we take the natural logarithm of one plus the number of submissions and comments posted by each agent, and denote it as *traffic*:

$$Traffic_{ilt} = log(1 + \sum_{k=1}^{K_{ilt}} \#Post_{ikt}), \tag{6}$$

where $\#Post_{ikt}$ is the number of submissions and comments posted by agent k for stock i in the week t, and K_{ilt} is the number of agents belonging to an agent group l for every stock i in the week t. The measure of traffic naturally reflects attention from submitters, where higher traffic means greater attention from that agent type.[16]

We re-estimate specification (5) by replacing agent tones with agent traffic. The estimation results are reported in Table 6 Panel C (columns I – II). We find that fanatic traffic positively predicts future naïve traffic. The coefficient of naïve traffic on future returns is positive and significant, indicating that higher naive traffic predicts higher future returns.

5.4 Use PageRank to Proxy for Influence

[16] We also compute other measures for social media activities, such as the concentration of the network and the dispersion among different types of investors. Since these measures are not directly linked to Pedersen's theory, we do not include them in the main text. These results are available on request.

Next, we consider Google PageRank as an alternative measure of influence. Google PageRank measures how connected a node is in the network (Page et al. 1999). In the context of Reddit influencers, the more central a user is, the more direct or indirect commenters she has, and thus the higher the PageRank value she has. We sum across the PageRank measures of all agents for every stock *i* in the week *t*, and compute stock*week level variables.

Results using PageRank are reported in Panel C of Table 6 (Columns III – VI). In the high influence subsample, we still find that agent tones positively predict returns. In the low influence subsample, we find that agent tones predict returns with a negative relation. Our main inference remains the same using this alternative measure of influence. That is, positive Reddit agent tones significantly drive up future returns only in networks with higher influences.

6. Conclusion

The volatile price movement of GameStop in January 2021, potentially driven by social media activity and retail trading, generates substantial interest in the capital market in understanding how social media affects information formation, price discovery and trading dynamics. Relying on the theoretical framework of Pedersen (2022), we systematically examine the social network structure and its influences on prices and trading, by collecting data directly from the social media platform Reddit.

Our results generally support the theoretical predictions. First, for belief formation, we find that opinions of rational and fanatic agents positively and significantly predict future opinions of naïve investors, especially in a network with higher agent influences. Second, for return predictions, more positive tones from social media significantly predict higher future returns, and more so when agents' influences are higher, demonstrating the importance of social media in the capital market. Finally, for trading dynamics, higher naïve tones generally increase retail flows.

More interestingly, whether short sellers short more or less against bullish social media tones (i.e., burst or ride the bubble) depends on agent influence in the social network. When agents are highly influential, short sellers shy away from more bullish tones, as these tones lead to more positive returns in the future and thus higher short squeeze risks; while networks with low influence agents, short sellers seem to not feel threatened by more bullish social media tones. More interestingly, short sellers' negative return predictive power is stronger in the social network with high influence, supporting Pedersen's prediction that sophisticated investors may ride or burst the bubble, depending on the balance between costs and benefits.

REFERENCES

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

Anderson, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61(1): 43-76.

Antweiler, W., and M. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59: 1259-1294.

Asquith, P., P. Pathak, and J. R. Ritter. 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78: 243-276.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention induced trading and returns: Evidence from Robinhood users. *Journal of Finance* 77(6): 3141-3190.

Barber, B. M., S. Lin, and T. Odean. 2023. Resolving a paradox: Retail trades positively predict returns but are not profitable. *Working Paper.*

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

Barber, B. M., X. Huang, P. Jorion, T. Odean, and C. Schwarz. 2024. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ. *Journal of Finance* 79: 2403-2427.

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

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

Betzer, A. and J. P. Harries. 2021. If he's still in, I'm still in! How Reddit posts affect Gamestop retail trading. *Working paper.*

Bianchi, F., R. Gómez-Cram, T. Kind, and H. Kung. 2023. Threats to central bank independence: High-frequency identification with Twitter. *Journal of Monetary Economics* 135: 37–54.

Bianchi, F., R. Gómez-Cram, and H. Kung. 2024. Using social media to identify the effects of congressional viewpoints on asset prices. *Review of Financial Studies* 37(7): 2244–2272.

Blume, M. E., and R. F. Stambaugh. 1983. Biases in computed returns: An application to the size effect. *Journal of Financial Economics* 12(3): 387-404.

Boehmer, E., Z. R. Huszar, and B. D. Jordan. 2010. The good news in short interest. *Journal of Financial Economics* 96(1): 80-97.

Boehmer, E., Z. R. Huszar, Y. Wang, X. Zhang, and X. Zhang. 2022. Can Shorts Predict Returns? A Global Perspective. *Review of Financial Studies* 35(5): 2428–2463.

Boehmer, E., C. M. Jones, and X. Zhang. 2008. Which shorts are informed? *Journal of Finance* 63: 491-527.

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

Box, T., R. Davis, R. Evans, and A. Lynch. 2021. Intraday arbitrage between ETFs and their underlying portfolios. *Journal of Financial Economics* 141: 1078-1095.

Bradley, D., J. Hanousek, R. Jame, and Z. Xiao. 2024. Place Your Bets? The Value of Investment Research on Reddit's Wallstreetbets. *Review of Financial Studies* 37(5): 1409–1459.

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

Cookson, J. A., J. E. Engelberg, and W. Mullins. 2023. Echo Chambers. *Review of Financial Studies* 36(2): 450–500.

Cookson, J. A., R. Lu, W. Mullins, and M. Niessner. 2024. The social signal. *Journal of Financial Economics* 158: 304-405.

Cookson, J. A., W. Mullins, and M. Niessner. 2024. Social Media and Finance. *Oxford Research Encyclopedia of Economics and Finance* (forthcoming).

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

Da, Z., J. Engelberg, and P. Gal. 2011. In search of attention. *Journal of Finance* 66(5): 1461-1499.

Das, S., and M. Chen. 2007. Yahoo! For Amazon: sentiment extraction from small talk on the web. *Management Science* 53: 1375-1388.

Desai, H., K. Ramesh, S. R. Thiagarajan, B. V. Balachandra. 2002. An investigation of the information role of short interest in the Nasdaq market. *Journal of Finance* 57(5): 2263-2287.

Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18: 277-311.

Diangson, B. and N. Jung. 2021. Bet if on Reddit: The effects of Reddit chatter on highly shorted stocks. *Working paper.*

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

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

Farrell, M., T.C. Green, R. Jame, S. Markov. 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics* 145 (2): 616–641.

Fong. K. Y. L., D. R. Gallagher, and A. D. Lee. 2014. Individual investors and broker types. *Journal of Financial and Quantitative Analysis* 49(2): 431-451.

Heimer, R.Z., 2016. Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies* 29(11): 3177-3209.

Hendershott, T., D. Livdan, and N. Schürhoff. 2015. Are institutions informed about news? *Journal of Financial Economics* 117: 249-287.

Hilscher, J., Pollet, J.M. and Wilson, M., 2015. Are credit default swaps a sideshow? Evidence that information flows from equity to CDS markets. J*ournal of Financial and Quantitative Analysis* 50(3): pp.543-567.

Hollifield, B., Neklyudov, A. and Spatt, C., 2017. Bid-ask spreads, trading networks, and the pricing of securitizations. *The Review of Financial Studies* 30(9): 3048-3085.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. Estimating Vector Autoregressions with Panel Data. *Econometrica* 56(6): 1371–1395.

Kaniel, R., G. Saar, and S. Titman. 2008. Individual investor trading and stock returns. *Journal of Finance* 63(1): 273-310.

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

Lee, J., Naranjo, A. and Velioglu, G., 2018. When do CDS spreads lead? Rating events, private entities, and firm-specific information flows. *Journal of Financial Economics* 130(3): 556-578.

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

Loughran, T. and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance 66*:135-165.

Lyosca, S., E. Baumohl, and T. Vyrost. 2021. YOLO trading: Riding with the herd during the GameStop episode. *Working paper.*

Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49(6): 1417-1426.

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

Page, L., S. Brin, R. M., and T. Winograd. The PageRank citation ranking: Brining order to the web. 1999. Technical Report, Stanford University, Stanford.

Pedersen, L. H. 2022. Game On: Social Networks and Markets. *Journal of Financial Economics* 146: 1097-1119.

Slezak, S.L. 1994. A theory of the dynamics of security returns around market closures. *Journal of Finance* 49:1163-1211.

Strych, J., and F. Reschke. 2022. Emojis and stock returns. *Working Paper.*

Tumarkin, R., and R. Whitelaw. 2001. News or noise? Internet message board activity and stock prices. *Financial Analysts Journal* 57: 41–51.

Vasileiou, E., E. Bartzou, and P. Tzanakis. 2022. Explaining GameStop short squeeze using intraday data and google searches. *Journal of Prediction Markets* 16(3): 67-79.

Welch, I. 2022. The wisdom of the Robinhood crowd. *Journal of Finance* 77: 1489-1527.

Table 1. Summary Statistics

This table presents summary statistics of main variables used in this study. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Detailed definitions of each variable are discussed in Section 3. Panel A presents the proportion of three agent groups from the Reddit sample with 468,274 firm-week observations. Rational agents are hard-headed agents with stable view that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. In this panel, we define indicator variables for the three types of agents, and present summary statistics for these indicator values. Panel B presents summary statistics for social media activities of each type of agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Influence measures each agent group's influence on investors, computed as the sum of the number of commentors of each agent group. To address the skewness in this distribution, we transform this measure to a domain of [0,1] for ease of interpretation based on equation (4). Panel C presents the summary statistics of the other main dependent variables. The variable *Return* is the daily return calculated for each trading day. The variable *RetailFlow* is the daily retail order imbalance measured in number of traded shares. The variable *ShortFlow* is the days-to-cover ratio, and is computed as the total number of shares on loan scaled by the daily trading volume.

Panel A. Proportion of three agent groups

	mean	std	median
Rational agent	0.0257	0.0152	0.0254
Fanatic agent	0.0680	0.0334	0.0625
Naïve agent	0.9063	0.0317	0.9091

Panel B. Social media activity measures

			correlation					
	mean	std	Rational Tone	Fanatic Tone	Naïve Tone	Rational Influence	Fanatic Influence	Naïve Influence
RationalTone	0.0016	0.0103	1					
FanaticTone	0.0025	0.0146	0.27	1				
NaïveTone	0.0052	0.0192	0.27	0.26	1			
RationalInfluence	0.0125	0.0707	0.51	0.32	0.28	1		
FanaticInfluence	0.0147	0.0726	0.38	0.41	0.30	0.72	1	

NaïveInfluence	0.0202	0.0866	0.46	0.38	0.37	0.85	0.85	1

Panel C. Other measures

	mean	std	P50
Return	0.0005	0.0280	0.0000
RetailFlow	-0.0282	0.1774	-0.0192
ShortFlow	0.0014	0.0831	0.0002

Table 2. Dynamics of the Social Network

This table presents results on the dynamics of social networks. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

	Panel A. Whole sample			Panel B. High influence subsample			Panel C. Low influence subsample		
	I	II	III	I	II	III	I	II	III
	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)
RationalTone(t-1)	0.2214***	0.1952***	0.1387***	0.2664***	0.2262***	0.1556***	0.0558***	0.0640***	0.0854***
	[21.30]	[22.25]	[21.55]	[21.85]	[22.12]	[16.48]	[8.04]	[6.05]	[8.04]
FanaticTone(t-1)	0.0864***	0.1519***	0.0840***	0.1268***	0.1969***	0.1191***	-0.0011	0.0613***	0.0266***
	[19.04]	[25.59]	[20.14]	[21.74]	[24.24]	[16.04]	[-0.49]	[7.46]	[4.44]
NaïveTone(t-1)	0.0386***	0.0488***	0.0945***	0.1632***	0.1840***	0.3063***	-0.0244***	-0.0175***	0.0120***
	[14.91]	[13.23]	[20.25]	[22.47]	[20.04]	[25.39]	[-15.33]	[-7.25]	[2.72]
Number of observations	472,765	472,765	472,765	48,310	48,310	48,310	424,455	424,455	424,455
p-value of Granger causality test	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)	Rational Tone(t)	Fanatic Tone(t)	Naïve Tone(t)
RationalTone(t-1)		0.0%	0.0%		0.0%	0.0%		0.0%	0.0%
FanaticTone(t-1)	0.0%		0.0%	0.0%		0.0%	62.3%		0.0%
NaïveTone(t-1)	0.0%	0.0%		0.0%	0.0%		0.0%	0.0%	

Table 3. Predicting Returns Using Social Media Views

This table presents results on predicting returns. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

	I. Whole sample	II. High influence subsample	III. Low influence subsample
	Return(t)	Return(t)	Return(t)
RationalTone(t-1)	0.0022	0.0131**	-0.0040
	[0.58]	[2.40]	[-0.71]
FanaticTone(t-1)	0.0075*	0.0222***	-0.0058
	[1.89]	[3.52]	[-1.62]
NaïveTone(t-1)	0.0068**	0.0685***	-0.0087***
	[2.57]	[8.36]	[-2.86]
Number of observations	472,765	48,310	424,455
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past RationalTone	56.0%	1.6%	47.9%
Past FanaticTone	5.8%	0.0%	10.5%
Past NaïveTone	1.0%	0.0%	0.4%

Table 4. Social Media Activity Associated with Retail Flows

This table presents results on retail flows. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

Panel A. How social media views relate to future retail flows

	I. Whole sample	II. High influence subsample	III. Low influence subsample
	RetailFlow(t)	RetailFlow(t)	RetailFlow(t)
RationalTone(t-1)	0.0122	0.0194	-0.0061
	[0.60]	[0.82]	[-0.20]
FanaticTone(t-1)	-0.0080	-0.0243	0.0083
	[-0.55]	[-1.49]	[0.35]
NaïveTone(t-1)	0.0656***	0.1055***	0.0497**
	[3.55]	[4.86]	[2.21]
Number of observations	472,765	48,310	424,455
p-value of Granger causality test	RetailFlow(t)	RetailFlow(t)	RetailFlow(t)
Past RationalTone	54.7%	41.5%	84.2%
Past FanaticTone	58.0%	13.6%	72.3%
Past NaïveTone	0.0%	0.0%	2.7%

Panel B. Retail flows' predictive power for returns with different agent influence

	I. Whole sample	II. High influence subsample	III. Low influence subsample
	Return(t)	Return(t)	Return(t)
RetailFlow(t-1)	0.0012***	0.0088**	0.0009***
	[4.21]	[2.37]	[3.72]
Number of observations	472,765	48,310	424,455
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past RetailFlow	0.0%	1.8%	0.0%

Table 5. Social Media Activity Associated with Shorting Flows

This table presents results on shorting flows. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

Panel A. How social media views relate to future shorting flows

	I. Whole sample	II. High influence subsample	III. Low influence subsample
	ShortFlow(t)	ShortFlow(t)	ShortFlow(t)
RationalTone(t-1)	-0.0003	-0.0001	-0.0006
	[-0.64]	[-0.62]	[-0.28]
FanaticTone(t-1)	-0.0004	-0.0004**	-0.0003
	[-1.11]	[-2.17]	[-0.45]
NaïveTone(t-1)	-0.0006	-0.0009**	-0.0004
	[-0.38]	[-2.18]	[-0.26]
Number of observations	472,765	48,310	424,455
p-value of Granger causality test	ShortFlow(t)	ShortFlow(t)	ShortFlow(t)
Past RationalTone	52.1%	53.7%	78.2%
Past FanaticTone	26.5%	3.0%	65.2%
Past NaïveTone	70.1%	2.9%	79.8%

Panel B. Shorting flows' predictive power for returns with different agent influence

	I. Whole sample	II. High influence subsample	III. Low influence subsample
	Return(t)	Return(t)	Return(t)
ShortFlow(t-1)	-0.0011***	-1.7969***	-0.0010***
	[-6.22]	[-3.53]	[-9.28]
Number of observations	472,765	48,310	424,455
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past ShortFlow	0.0%	0.0%	0.0%

Table 6. Further Discussion and Robustness Checks

This table presents results on robustness check and further discussion. Our sample period is Jan 2, 2020 to Dec 31, 2023, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Panel A reports results in different time periods: pre-GME (2020/01-2020/12), during-GME (2021/01-2021/02), and post-GME (2021/03-2023/12). Column I-III report results on predicting returns. Column IV-VI report results on shorting flows. Panel B reports the estimation results using alternative agent classifications and alternative measures for tones. In column I-II, we identify hardheaded agents as those who post more than 99% of all other agents, instead of 95% as in the main results; in column III-IV, we require that hardheaded agent's posts have the same sign in tone (either positive or negative) for 100% of their posts, instead of only 75% of their posts during the 5-day window; in column V-VI, we compute the influence-weighted tone to highlight the importance of agent influence in social networks, rather than defining tones of an agent-type as the average tone across all individuals in that type; in column VII-VIII, we control for other social media tone using the PC1 of sentiment on Twitter, StockTwits, and Seeking Alpha, which is measured by Cookson et al. (2024). Panel C reports the estimation results using alternative proxies for social media activity and influence. Column I-II report the estimation results using traffic as an alternative measure for social media activity. Traffic measures investors' attention towards the firm, computed as the natural logarithm of one plus the number of posts and comments discussing the firm. Column III-VI report the estimation results using PageRank as an alternative influence measure. To reduce the fat tail and make it easy to interpret, we take logarithm, rank the variables each day, and match them to the [0,1] interval. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

Panel A. Different time periods

	Return(t)			ShortFlow(t)		
	I	II	III	IV	V	VI
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.0023	0.0261	-0.0018	0.0018***	0.0006***	-0.0010
	[0.30]	[1.40]	[-0.43]	[8.81]	[3.35]	[-1.31]
FanaticTone(t-1)	0.0035	0.0339	0.0052*	0.0024***	0.0010***	-0.0014***
	[0.58]	[0.83]	[1.78]	[10.31]	[5.65]	[-3.57]
NaïveTone(t-1)	-0.0138**	-0.0325	0.0130***	0.0067***	0.0037***	-0.0032**
	[-2.18]	[-1.34]	[4.53]	[12.55]	[10.15]	[-2.11]
Number of observations	105,764	21,066	345,935	105,764	21,066	345,935
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	76.3%	16.3%	28.7%	0.0%	0.1%	28.4%
Past FanaticTone	56.4%	40.8%	9.1%	0.0%	0.0%	0.0%
Past NaïveTone	2.9%	17.9%	0.0%	0.0%	0.0%	3.6%

Panel B. Alternative measures for agents and tone

	I	II	III	IV	V	VI	VII	VIII
	Use P99 of number of posts as threshold of hardheaded		Require stable tones for past 5 days for hardheaded		Influence-weighted tone		Control for other social media tone	
	NaïveTone(t)	Return(t)	NaïveTone(t)	Return(t)	NaïveTone(t)	Return(t)	NaïveTone(t)	Return(t)
RationalTone(t-1)	0.1522***	0.0035	0.1402***	-0.0020	0.1311***	0.0021	0.1450***	0.0054
	[17.22]	[0.79]	[16.64]	[-0.47]	[19.88]	[0.57]	[17.13]	[0.97]
FanaticTone(t-1)	0.1058***	0.0108	0.1042***	0.0094*	0.0765***	0.0077*	0.1103***	0.0152**
	[18.52]	[1.19]	[20.06]	[1.79]	[18.14]	[1.85]	[18.09]	[2.06]
NaïveTone(t-1)	0.1587***	0.0082***	0.1506***	0.0094***	0.0899***	0.0072***	0.1559***	0.0222***
	[27.64]	[2.91]	[28.50]	[2.79]	[19.94]	[2.76]	[17.38]	[3.82]
Number of observations	472,765	472,765	472,765	472,765	472,765	472,765	81,156	81,156

Panel C. Alternative proxies for social media activity and influence

	I	II		III	IV	V	VI
	Traffic as a proxy for social media activity			Pagerank as a proxy for influence, high influence		Pagerank as a proxy for influence, low influence	
	NaïveTraffic(t)	Return(t)		NaïveTone(t)	Return(t)	NaïveTone(t)	Return(t)
Rational Traffic(t-1)	-0.0086	-0.0172	Rational Tone(t-1)	0.1713***	0.0155**	0.0577***	-0.0012
	[-0.79]	[-0.58]		[15.13]	[2.49]	[6.15]	[-0.22]
Fanatic Traffic(t-1)	0.2168***	-0.0166	Fanatic Tone(t-1)	0.1233***	0.0272***	0.0204***	-0.0083**
	[17.16]	[-0.53]		[14.40]	[4.09]	[3.85]	[-2.37]
Naïve Traffic(t-1)	0.6492***	0.0526**	Naïve Tone(t-1)	0.3834***	0.0869***	-0.0028	-0.0089***
	[51.13]	[2.18]		[23.75]	[8.14]	[-0.80]	[-2.99]
Number of observations	472,765	472,765		41,067	41,067	431,698	431,698

Figure 1. Distribution of Reddit Activities of Agent Group

These graphs present the distribution of Reddit activities of three agent groups for whole sample from Jan 2, 2020 to Dec 31, 2023. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents.

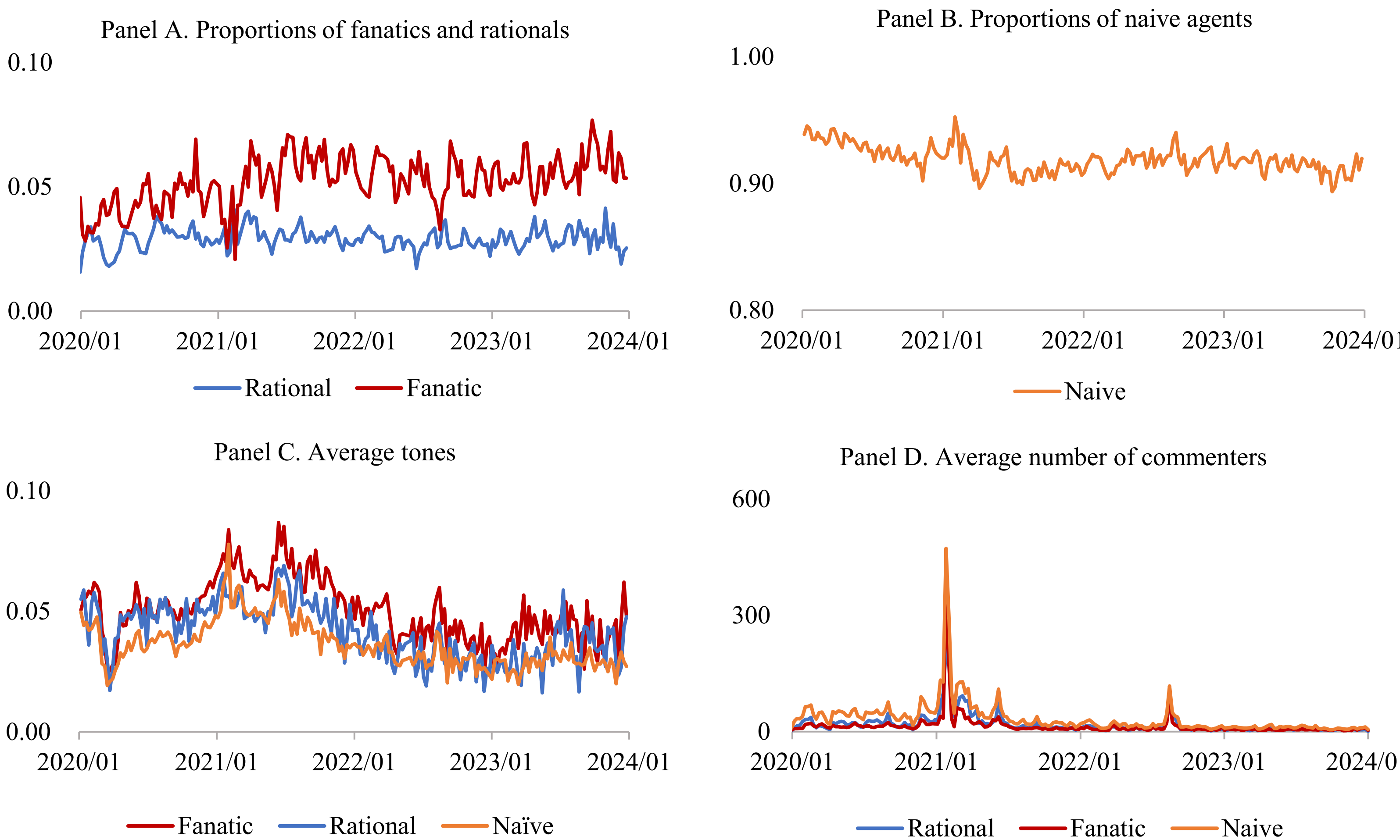


Figure 2. Impulse Responses for Agents' Tones

The figure reports the impulse response functions (IRF) corresponding to the PVAR estimation in Table 2. Impulse responses correspond to a one standard deviation shock. Error bands at 5% level for the impulse responses (dashed lines) are generated using Monte-Carlo simulations with 1000 draws.

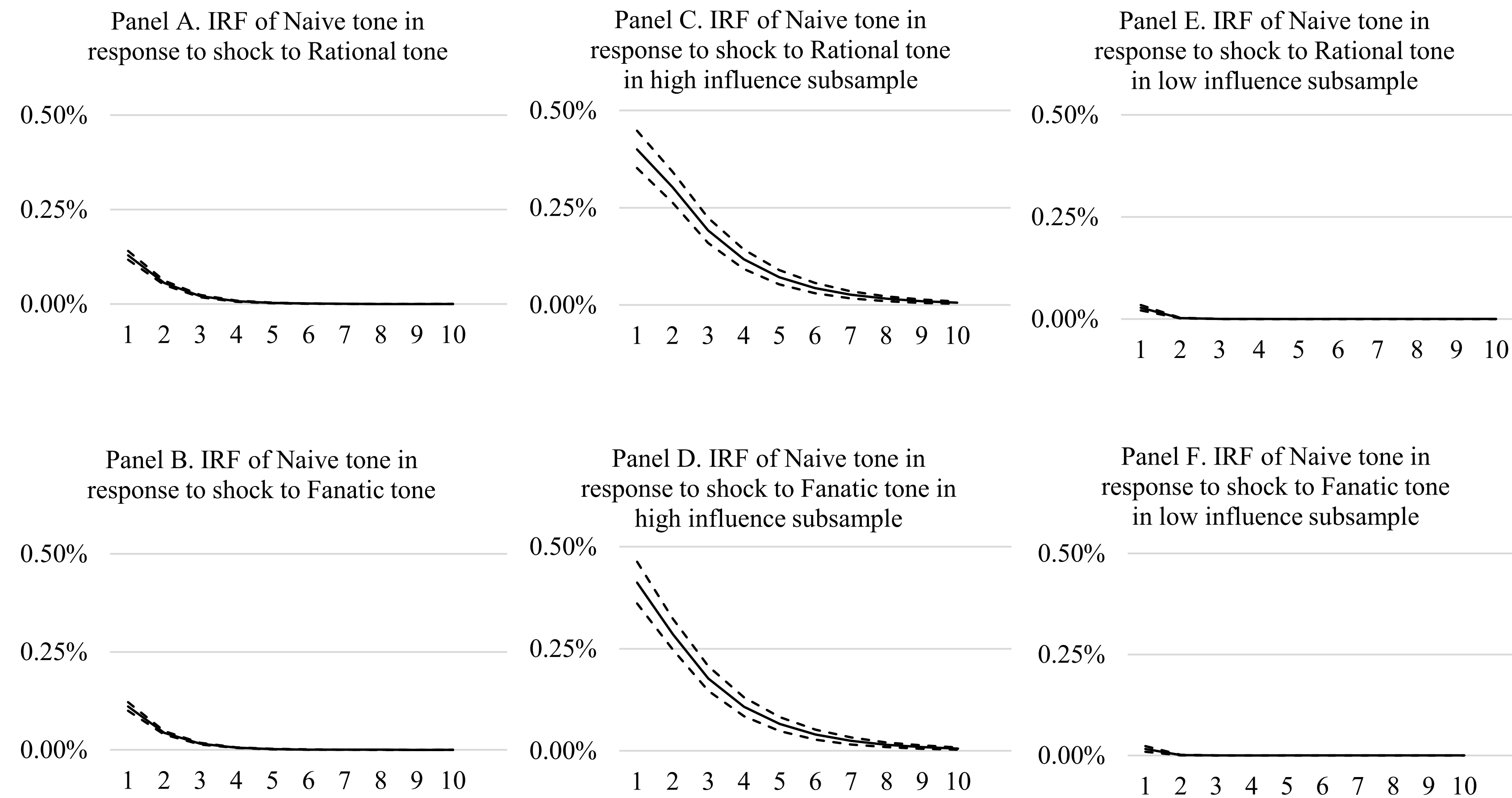


Figure 3. Impulse Responses for Agents' Tones and Returns

The figure reports the impulse response functions (IRF) corresponding to the PVAR estimation in Table 3. Impulse responses correspond to a one standard deviation shock. Error bands at 5% level for the impulse responses (dashed lines) are generated using Monte-Carlo simulations with 1000 draws.

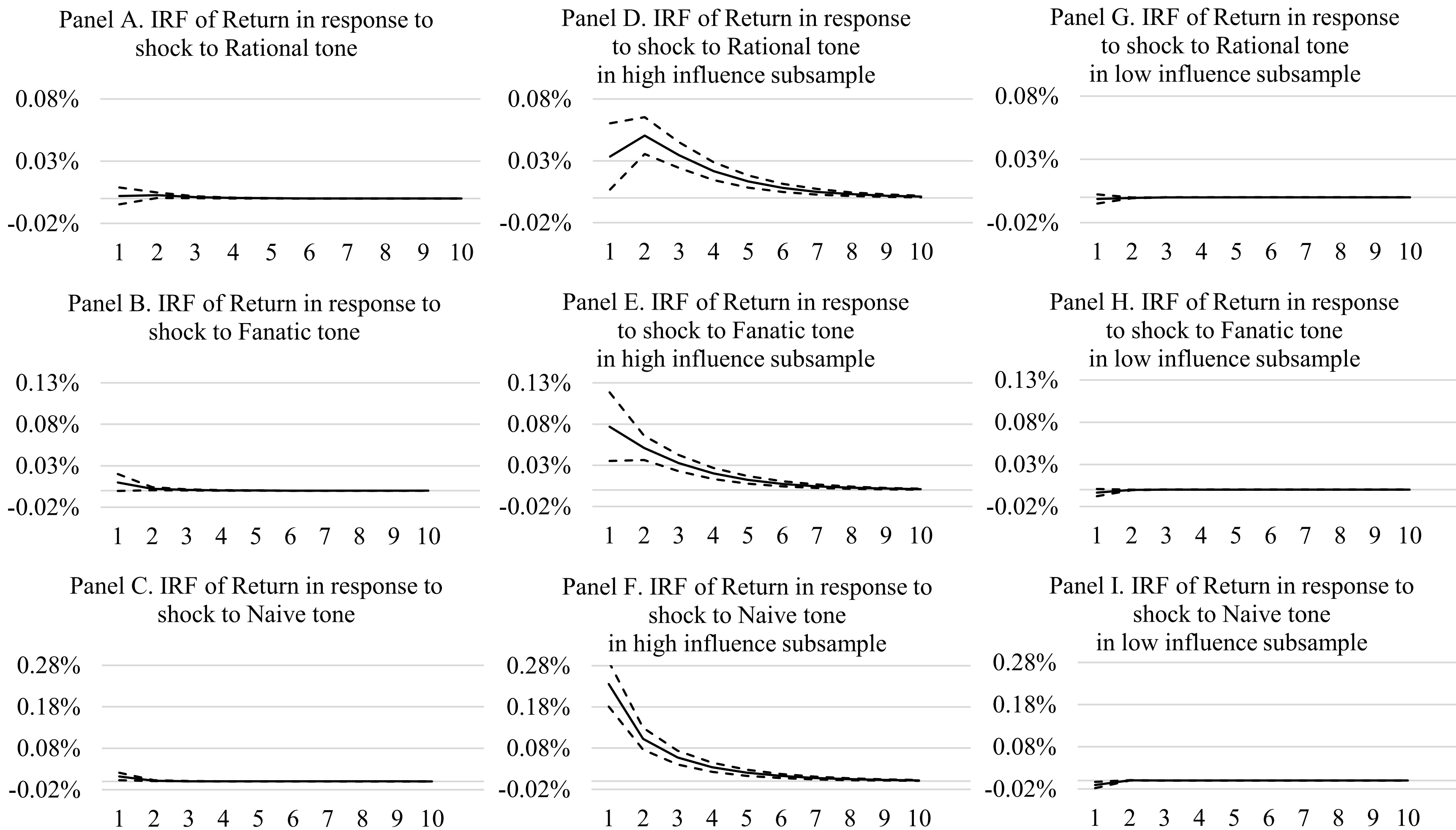


Figure 4. Impulse Responses for Agents' Tones and Retail Flows

The figure reports the impulse response functions (IRF) corresponding to the PVAR estimation in Table 4. Impulse responses correspond to a one standard deviation shock. Error bands at 5% level for the impulse responses (dashed lines) are generated using Monte-Carlo simulations with 1000 draws.

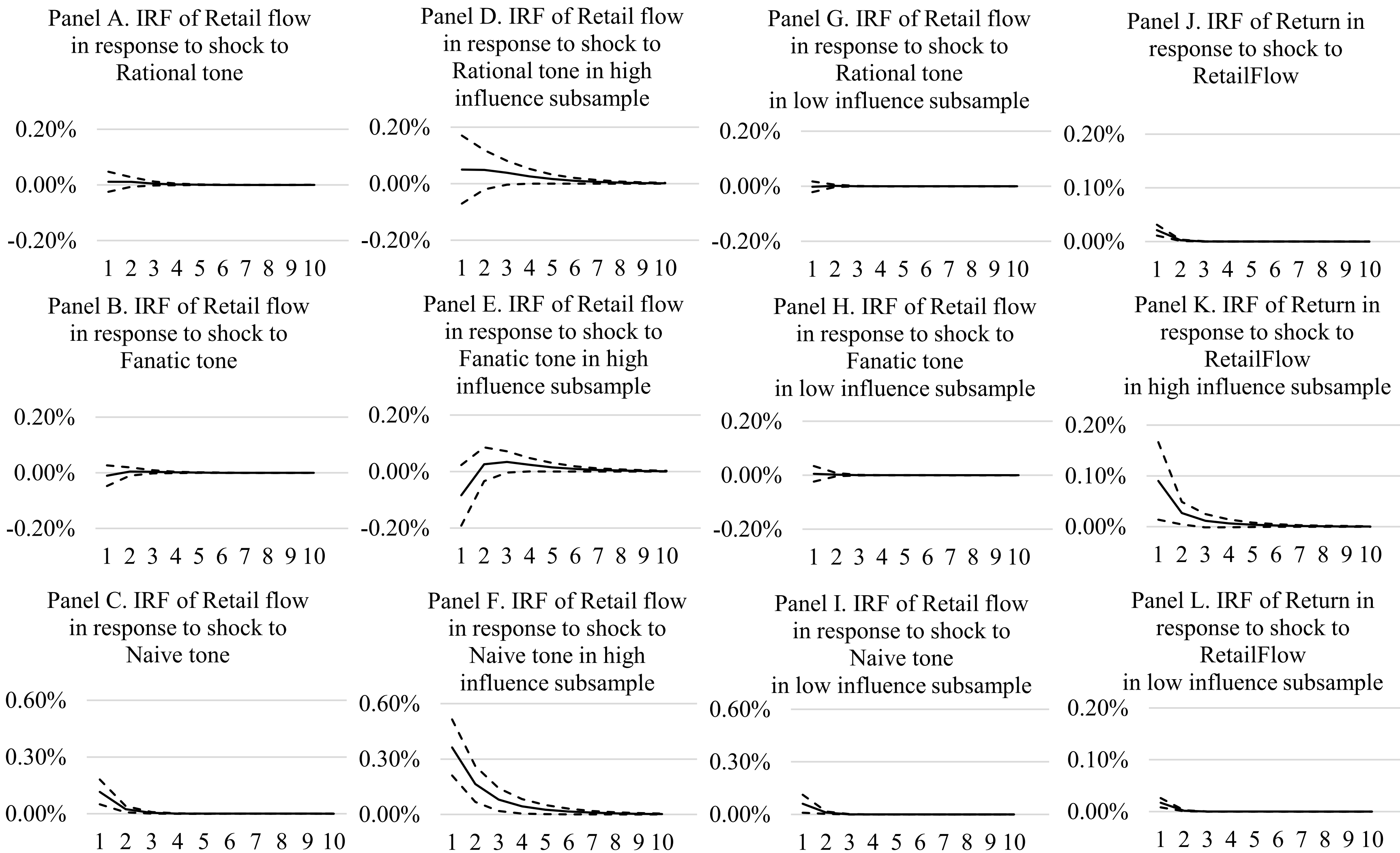
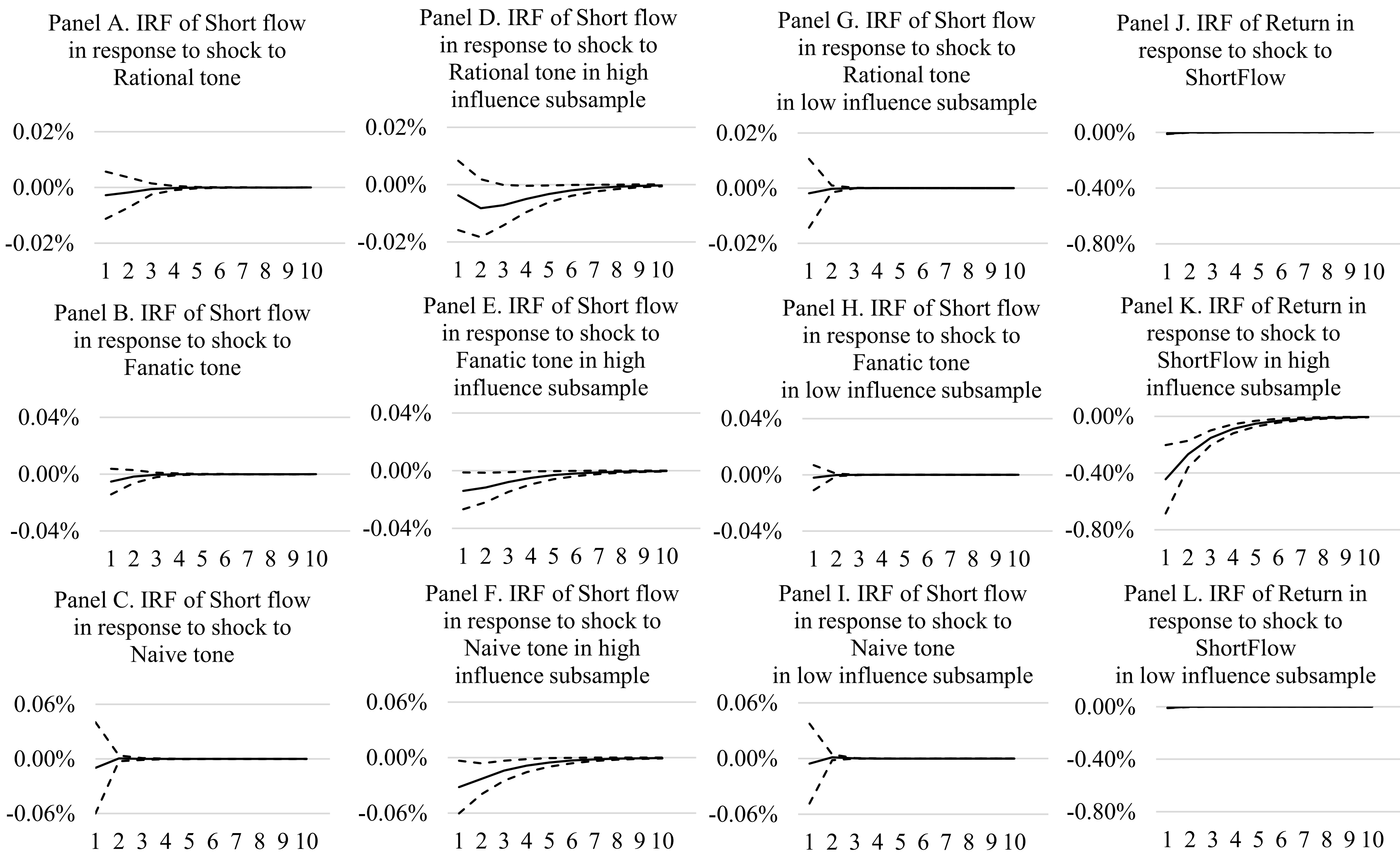


Figure 5. Impulse Responses for Agents' Tones and Shorting Flows
The figure reports the impulse response functions (IRF) corresponding to the PVAR estimation in Table 5. Impulse responses correspond to a one standard deviation shock. Error bands at 5% level for the impulse responses (dashed lines) are generated using Monte-Carlo simulations with 1000 draws.



Appendix A Sentiment and Value Dictionaries

In this Appendix, we outline the methods we used to define our sentiment and value dictionary. Traditional text analysis often uses word counts, and here we apply the same method. Since users on r/wallstreetbets have their own lingo (e.g., emojis, slang, jokes, and special meaning words), traditional measures of sentiment which uses specialized financial dictionaries, such as the Loughran and McDonald dictionary (LM), are not well suited for calculating the tone of posts and comments on Reddit (Bradley et al., 2023). We create a modified LM dictionary to better capture Reddit sentiment. We first gather all the text from the titles of submissions and strip the text of punctuation and numbers. Next, we remove stop words, set all words to lower case letters, lemmatize and finally tokenize each word. We identify the 1,000 most important words using the tfidf algorithm and manually classify each word as a positive, neutral, or negative word. We took special care to examine every word in the context that it is used on Reddit by surveying randomly selected posts or comments which contain the word, before assigning sentiment. In Panel A, we list all positive or negative words that are not included in the traditional LM dictionary. Next, we combine our manually classified 1,000-word sentiment dictionary with other words in the LM dictionary and we use this modified LM dictionary to calculate sentiment. We use a similar approach to assess whether a specific word is value relevant. We manually tag every word from the list of 1,000 most frequently appearing words and determine whether they contain information about firm fundamentals. To help make this decision, we also read randomly selected posts or comments to better understand the context under which these words are used on the reddit forum. We present the list of value-relevant words in Panel B. We also identify the 100 most popular emojis and manually classify each emoji as a positive, neutral, or negative emoji. We include them in our sentiment dictionary as well, and list positive and negative emojis in Panel C.

Panel A. Additional positive and negative words in our Reddit dictionary

Positive words not recognized in LM							Negative words not recognized in LM				
appreci	climb	get	hold	love	rise	trust	baghold	death	hit	sad	wtf
ath	congrat	glad	hope	million	rocket	upvot	bear	delet	idiot	scare	
beat	correct	go	invest	moon	safe	well	bitch	die	issu	sell	
big	crush	gold	join	nice	smart		bleed	dont	kill	shit	
bought	decent	got	jump	power	solid		boomer	dumb	pain	shitti	
break	diamond	grow	launch	pump	super		broke	fake	piss	sold	
bull	energi	growth	leap	purchas	sure		bullshit	fall	put	sorri	
bullish	fine	hand	legit	recov	tendi		crash	fomo	restrict	stupid	
buy	free	high	like	rich	thank		cri	fucker	revers	suck	
call	fun	higher	long	right	true		dead	hate	rip	tank	

Panel B. Words in value dictionary

announc	ceo	debt	fundament	industri	manufactur	quarter	revenu
asset	cut	demand	grow	info	news	releas	sharehold
bankrupt	data	dividend	growth	ipo	pe	report	store
cap	dd	earn	guidanc	loan	product	research	suppli

Panel C. Most used emojis

Positive emojis						Negative emojis		
🚀	🌈	🍀	❤	👌	😉	😭	😳	💩
💎	🔥	💪	✊	😆	🤞	🤦	📉	🥺
🙌	📈	👍	🤝	💯	✅	😤	🖕	🙄
🌕	😎	👏	😍	😄	😜	💀	😩	😔

Appendix B. Technical Details about PVAR

There are three technical details for our PVAR estimation. First, we follow Hendershott et al. (2015) and remove firm fixed effect by applying the forward orthogonal deviations transformation. After transformation,

$$y_{i,t}^{*} = \left(\frac{\mathrm{T-t}}{\mathrm{T-t+1}}\right)^{\frac{1}{2}} \left(y_{i,t} - \frac{1}{T-t}\sum_{s=t+1}^{T} y_{i,s}\right), \tag{A1}$$

and equation (5) is reduced to

$$y_{i,t}^{*} = \sum_{l=1}^{L} A_l\, y_{i,t-l}^{*} + \varepsilon_{i,t}^{*}, \tag{A2}$$

where $\varepsilon_{i,t}^{*} = \left(\frac{\mathrm{T-t}}{\mathrm{T-t+1}}\right)^{\frac{1}{2}} \left(\varepsilon_{i,t} - \frac{1}{T-t}\sum_{s=t+1}^{T} \varepsilon_{i,s}\right)$ is the transformed error term. Second, since $y_{i,t-l}^{*}$ correlates with $\varepsilon_{i,t}^{*}$, we follow Holtz-Eakin, Newey, and Rosen (1988) and use the lagged untransformed variables $y_{i,t}$ as instruments to obtain unbiased estimates. Third, if an agent type's tone measure is missing, we replace it with zero and create a corresponding indicator variable, in order to minimize the impact of missing variables on the estimation results.

Appendix C. Social Media Dynamics Across Different Samples and Time Periods

In this Appendix, we present detailed results on social media dynamics across different time periods. We divide the whole sample into three time periods: pre-GME (2020/01-2020/12), during-GME (2021/01-2021/02), and post-GME (2021/03-2023/12). Our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each agent group measures their views, and is computed using the text in each submission/comment: (number of positive words and emojis - number of negative words and emojis)/(number of words + number of emojis). Firms are classified as high-influence if their overall influence (total number of commenters) is above the 90th percentile of all firms for that week; others are classified as low-influence. Parameters are estimated using PVAR with GMM estimation with lag length L=1. We subtract from each variable in the model its cross-sectional mean before estimation to remove common time fixed effects from all the variables. Following Hendershott et al. (2015), we apply the forward orthogonal deviations transformation to eliminate firm fixed effects. The standard errors are clustered on date and firm. T-statistics are reported in brackets. Levels of significance are denoted by * (10%), ** (5%), and *** (1%).

Panel A. Dynamics of rational tones in different time periods

	Rational Tone (t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.2369***	0.2471***	0.2000***	0.2991***	0.2177***	0.2470***	0.0193	0.1611***	0.0460***
	[19.57]	[8.32]	[14.17]	[17.87]	[5.85]	[14.47]	[1.44]	[3.94]	[6.15]
FanaticTone(t-1)	0.1131***	0.1281***	0.0704***	0.1798***	0.1877***	0.1051***	-0.0153***	0.0234*	-0.0037*
	[13.21]	[6.78]	[12.70]	[12.18]	[6.12]	[14.20]	[-2.94]	[1.77]	[-1.68]
NaïveTone(t-1)	0.0651***	0.0735***	0.0261***	0.3547***	0.4201***	0.1052***	-0.0585***	-0.0419***	-0.0167***
	[9.51]	[5.98]	[9.76]	[10.87]	[10.03]	[14.44]	[-13.20]	[-4.56]	[-10.57]
Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past FanaticTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	7.7%	50.7%
Past NaïveTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Panel B. Dynamics of fanatic tones in different time periods

	Fanatic Tone (t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.1990***	0.1964***	0.1848***	0.2452***	0.1680***	0.2195***	0.0507***	0.1288*	0.0494***
	[13.45]	[6.19]	[15.30]	[12.12]	[4.65]	[14.94]	[2.64]	[1.96]	[4.60]
FanaticTone(t-1)	0.1674***	0.1580***	0.1444***	0.2371***	0.2334***	0.1822***	0.0521***	0.0513***	0.0631***
	[15.93]	[8.78]	[18.16]	[13.90]	[7.65]	[16.93]	[3.84]	[2.62]	[5.78]
NaïveTone(t-1)	0.0724***	0.0704***	0.0380***	0.3332***	0.3472***	0.1427***	-0.0272***	-0.0207	-0.0178***
	[8.65]	[4.61]	[9.31]	[8.59]	[6.19]	[14.22]	[-4.11]	[-1.64]	[-6.95]
Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	5.0%	0.0%
Past NaïveTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.2%	0.0%

Panel C. Dynamics of naive tones in different time periods

	Naïve Tone(t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.1469***	0.1643***	0.1183***	0.2008***	0.1295***	0.1275***	0.0880***	0.1564***	0.0560***
	[14.51]	[7.14]	[13.97]	[10.36]	[4.29]	[10.57]	[4.24]	[2.96]	[4.57]
FanaticTone(t-1)	0.0914***	0.1272***	0.0714***	0.1638***	0.1684***	0.0987***	0.0341***	0.0745***	0.0115*
	[11.99]	[8.67]	[14.03]	[8.75]	[5.99]	[11.43]	[2.58]	[3.83]	[1.70]
NaïveTone(t-1)	0.1223***	0.1334***	0.0776***	0.5396***	0.5174***	0.2361***	0.0065	0.0484**	0.0041
	[13.99]	[6.06]	[14.53]	[11.30]	[9.24]	[19.30]	[0.70]	[1.99]	[0.79]

Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%
Past FanaticTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	0.0%	2.6%

Panel D. Predicting returns in different time periods using social media views

	Return(t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.0023	0.0261	-0.0018	0.0837***	0.0202	0.0003	-0.0197	-0.0071	-0.0029
	[0.30]	[1.40]	[-0.43]	[3.58]	[0.61]	[0.06]	[-1.53]	[-0.30]	[-0.47]
FanaticTone(t-1)	0.0035	0.0339	0.0052*	0.0901***	0.1324	0.0064	-0.0294***	-0.0353**	0.0038
	[0.58]	[0.83]	[1.78]	[4.06]	[1.61]	[1.51]	[-3.75]	[-2.23]	[0.92]
NaïveTone(t-1)	-0.0138**	-0.0325	0.0130***	0.3223***	0.2955*	0.0395***	-0.0464***	-0.0688***	0.0041
	[-2.18]	[-1.34]	[4.53]	[4.49]	[1.69]	[6.11]	[-6.97]	[-2.74]	[1.19]
Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	76.3%	16.3%	28.7%	0.0%	54.4%	51.8%	12.5%	76.3%	50.5%
Past FanaticTone	56.4%	40.8%	9.1%	0.0%	10.8%	10.9%	0.0%	2.6%	58.3%
Past NaïveTone	2.9%	17.9%	0.0%	0.0%	9.1%	0.0%	0.0%	0.6%	33.5%

Panel E. Social media activity associated with retail flows in different time periods

	Retail Flow(t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME

RationalTone(t-1)	0.0809***	-0.0881*	-0.0146	0.1090***	-0.0615	-0.0241	-0.0020	-0.0987	0.0027
	[2.97]	[-1.95]	[-0.55]	[3.15]	[-1.25]	[-0.74]	[-0.03]	[-1.18]	[0.07]
FanaticTone(t-1)	0.0241	-0.0131	-0.0254	0.0504*	0.0084	-0.0547**	-0.0257	0.0065	0.0086
	[1.09]	[-0.30]	[-1.39]	[1.75]	[0.16]	[-2.55]	[-0.62]	[0.11]	[0.29]
NaïveTone(t-1)	0.1155***	-0.0301	0.0555***	0.2796***	0.1204	0.0620***	0.0580	-0.0048	0.0508**
	[2.78]	[-0.27]	[2.78]	[3.42]	[0.79]	[2.63]	[1.21]	[-0.04]	[2.00]
Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	0.3%	5.1%	51.8%	0.2%	21.1%	41.6%	97.2%	23.7%	99.6%
Past FanaticTone	27.6%	76.1%	18.1%	8.0%	86.9%	0.8%	53.5%	91.3%	65.6%
Past NaïveTone	0.5%	79.0%	0.8%	0.1%	43.0%	0.8%	22.6%	96.8%	6.8%

Panel F. Social media activity associated with shorting flows in different time periods

	Short Flow(t)								
	Whole sample			High influence subsample			Low influence subsample		
	I	II	III	IV	V	VI	VII	VIII	IX
	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME	Pre-GME	During-GME	Post-GME
RationalTone(t-1)	0.0018***	0.0006***	-0.0010	0.0005***	0.0006**	-0.0010***	0.0018***	0.0012***	-0.0002
	[8.81]	[3.35]	[-1.31]	[6.46]	[2.14]	[-3.17]	[4.24]	[2.94]	[-0.06]
FanaticTone(t-1)	0.0024***	0.0010***	-0.0014***	0.0004***	0.0012***	-0.0011***	0.0035***	0.0010***	-0.0014
	[10.31]	[5.65]	[-3.57]	[4.14]	[3.74]	[-4.50]	[8.22]	[4.31]	[-1.39]
NaïveTone(t-1)	0.0067***	0.0037***	-0.0032**	0.0008***	0.0049***	-0.0019***	0.0071***	0.0038***	-0.0035*
	[12.55]	[10.15]	[-2.11]	[2.82]	[7.07]	[-4.59]	[11.97]	[13.43]	[-1.70]
Number of observations	105,764	21,066	345,935	10,659	2,077	35,569	95,105	18,989	310,366
p-value of Granger causality test	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)	Fanatic Tone(t)	Rational Tone(t)	Naïve Tone(t)
Past RationalTone	0.0%	0.1%	28.4%	0.0%	3.2%	0.3%	0.0%	0.3%	99.9%
Past FanaticTone	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	16.7%
Past NaïveTone	0.0%	0.0%	3.6%	0.5%	0.0%	0.0%	0.0%	0.0%	9.0%