

**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*

April 2024

ABSTRACT

Using submission level data from 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 opinions of hardheaded investors (fanatic or rational) significantly predict future opinions of naïve investors, especially when these investors have larger influences. For return predictions, social media tones positively and significantly predict future returns, and more so when agents' influences are 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@pbcfs.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@pbcfs.tsinghua.edu.cn) is at the PBC School of Finance at Tsinghua University. We would like to thank Lin Tan for her excellent research assistance. The paper benefits from comments from Matthew Ringgenberg, Zhi Da, Marina Niessner, seminar participants at Tsinghua University, Harvard Business School, Northwestern University, as well as conference participants at CFEA, WFA, EFA, and CICF. 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**

April 2024

ABSTRACT

Using submission level data from 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 opinions of hardheaded investors (fanatic or rational) significantly predict future opinions of naïve investors, especially when these investors have larger influences. For return predictions, social media tones positively and significantly predict future returns, and more so when agents' influences are 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 leads 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. Interestingly, as we write the first version of this article on March 1, 2021, the share price quietly moves back to \$101.74. This enormous volatility of the GameStop share price attracts substantial attention, and most of the 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, at a 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, fanatic, rational and naïve, all of whom interact in a social network, Pedersen (2022) derives closed-form solutions for dynamics of beliefs and prices and proposes three main testable predictions in answering the above question. First, Pedersen (2022) shows that there are belief spillovers from social network interactions, with echo-chamber effects, and naïve investors’ beliefs are affected by both fanatic and rational views, more so if the fanatic and rational have higher influences in the network. Second, the social

network views by different investors can directly affect 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, and lead 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 price deviating from fundamental value 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, as well as from standard capital market data sources, over the sample period from January 2020 through February 2021. Following Pedersen's theoretical framework, we first separate all investors into hardheaded (investors who don't change their opinions) and naïve investors (investors who have fluid opinions). We further separate hardheaded investors into rational investors (who pay attention to firm fundamental values) and fanatic investors (who don't pay attention to fundamentals). With our empirical categorization following Pedersen's assumptions, about 8% of Reddit users are identified as fanatics, 4% are rational, and 88% 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 by Granger-causal relations and to quantify the responses of the variables to innovations in other variables by impulse response functions.

We first examine how social networks affect belief formation. Using textual analysis, we measure each investor's opinion by her tone. Estimation from PVAR provides strong evidence that the views from fanatic and rational agents significantly and positively predict next-day 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 the influences of different types of investors using the sum of the number of direct commenters for each agent in the agent type. We further document that the impacts of fanatic and rational agents' views on next-day naïve investors' views are stronger in the network where these agents have higher influences.

Next, we study how social network dynamics affect price movements. Specifically, we use investors' tones to predict next-day returns, where tones represent investors' beliefs about the stocks they are discussing. We provide direct and strong evidence that Reddit tones can significantly predict future stock price movements. To be more specific, higher agent tones are associated with higher future returns. More interestingly, the impacts of fanatic, rational, naïve tones on next-day returns are higher when agent influence in the network is higher. These findings suggest that social network interactions have significant influences on price formation.

Last, we investigate how other important market participants, such as retail investors and short-sellers, trade in the existence of social media activities. Retail investors trading is influenced by Reddit tones, in the sense that higher social media tones predict higher next-day retail flows. Specifically, fanatic, rational, and naïve tones Granger cause next-day retail flows with positive coefficients. Moreover, the impact of agent tone on retail flows is larger in networks with higher influences. Within the framework of Pedersen (2022), the fact that a shock in positive social media tones lead to more bullish retail trading suggests retail investors buy and potentially profit from the short-term uptrend in price driven by social media activities.

For short sellers, our results show that Reddit tones significantly impact shorting flows, and whether short sellers short against bullish social media tones (i.e., burst or ride the bubble) depends on agent influence in the social network. In particular, in networks with lower agent influence, short sellers may choose to short more against agents' positive views (i.e. burst the bubble). By contrast, when agents' influences are higher, the impulse response function shows that shorting flows significantly decrease in response to a positive shock of agent tone. It indicates that short sellers, to some extent, are deterred by Reddit tones when Reddit agents are more influential, which suggests that short sellers worry about the risk of a short squeeze. In terms of short sellers' negative predictive power for future returns, we find that, 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 the earlier results that short sellers may generally shy away from short selling when social media tones are higher (ride the bubble) in the high influence network, when they do choose to short (burst 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. Most existing research on social media provides suggestive evidence that users' social media activities, sentiment, and dispersion of sentiment are correlated with stock returns, trading volume, and volatility.¹ In the event of Gamestop and Reddit, a few contemporaneous papers examine how social media sentiment directly affects GameStop's spike in price in January and find that investor attention and sentiment significantly predict stock returns.²

In terms of retail investors, before 2010, many papers, such as Barber and Odean (2008), Barber, Odean, and Zhu (2009), find that retail investors are generally uninformed. However, evidence after 2010, including 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. During the Covid-19 pandemic in 2020, 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).

¹ Earlier papers, such as Tumarkin and Whitelaw (2001), Antweiler and Frank (2004), Das and Chen (2007), Chen et al. (2014), and Bartov et al. (2018), establish links between social media sentiment and stock returns, volatility and earnings news.

² Contemporaneous Reddit papers include Betzer and Harries (2021), Diangson and Jung (2021), Long, Lucey, and Yarovaya (2021), Bradley, Hanousek, James and Xiao (2023), Lyocsa, Baumohl and Vyroost (2022), Vasileiou, Bartzou, and Tzanakis (2022), Strych and Reschke (2022).

There is also a vast literature on short sellers. Theoretical work by Diamond and Verrecchia (1987, the DV model hereafter) argues that the high costs of short selling and the resulting absence of liquidity-motivated short selling makes short sellers more informed than average traders. Empirically, 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, and Fusari, Jarrow and Lamichhane (2022) propose that the January GME event represents a bubble.

Despite the large volume of previous literature on social media, retail investors, and short sellers, none of the existing literature examines the joint dynamics of different types of agents' beliefs, retail investors and short sellers' trading behaviors, and price formation in the social network, for all listed stocks in the U.S. market (rather than just for GME and a few others). Our study, by relying on concrete theoretical predictions, provides unique insights and timely answers to various questions on the interactions of multiple participating parties during the belief and price formation processes, and these answers can be helpful for all market participants.

2. Pedersen's Model and Empirical Hypotheses Development

Pedersen (2022) is one of the first theoretical research to provide a comprehensive framework to understand social networks and its implications for asset prices. Here we introduce the assumptions and propositions in Pedersen's model, and develop our empirical

hypotheses on belief formation, price discovery and trading behaviors in the social network accordingly. We refer readers to the original paper for more details in 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. They have information on v after the first round, and do not change their opinions in later rounds. Fanatic investors learn only from themselves, and do not change their opinions. 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 $x(t) = (x_1(t), \dots, x_N(t))$. Both fanatics and rational investors do not change their views after the first round, so these two types of agents are labeled as "hardheaded".

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

another word, 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.³

In this network, if one agent is influential, he 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 don't have thought leadership since 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 fanatics and rational agents (Proposition 1 of Pedersen 2022):

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

³ To highlight the social network effect on the belief formation, price discovery and trading behaviors, Pedersen (2022) abstracts away certain aspects of the 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.

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 fanatic and rational investors.

H2: The more influences fanatic and rational agents have, the greater impacts they have 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):

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

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 another word, the contribution of agent j on long term price is higher if she has higher influence. Following propositions 4 and 5, we develop two testable hypotheses:

H3: Agent views from social media network predict next day stock returns.

H4: Agents with higher influences have larger impacts 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 change before the fundamental value of the stock is revealed. For instance, if investors anticipate that social media views drive up the price, causing it to deviate from the true value of the stock, they optimize their trading behaviors by weighing the benefits and costs of riding a potential positive bubble. If the 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 than institutional investors. They tend to follow social media trends and have played a unique role in the Gamestop episode. In contrast, 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 their significant influences on investors' views and prices, we form the following testable hypotheses following Propositions 7 and 8:

H5: Social media views predict next day retail flows, and views from more influential agents have larger impacts on retail flows.

H6: Social media views predict next day shorting flows, and 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. 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. In this article, we focus on one forum in Reddit, “WallStreetBets”, on which participants discuss the trading of stocks and equity options. As of March 2024, the subreddit has a total of 15 million subscribers, making it one of the largest social media forums for financial news and trading strategies. Other than its popularity, we choose to study Reddit data in this article for two additional reasons. First, the GameStop phenomenon of January 2021 was instigated by discussions on r/wallstreetbets. Second, Reddit users see everyone else’s posts on the front page of the subreddit without having to subscribe or follow other user accounts. This feature differs from other social media platforms such as X (formerly Twitter), making Reddit an ideal platform for investors to directly listen to and respond to other’s beliefs.

We collect all submissions and their comments on this subreddit between January 1, 2020, and February 15, 2021. Submissions are posts initiated by Reddit users, which are usually presented in chronological order on a forum’s front page. Each submission also has its individual web page. Comments made in response to the content of the submission are positioned below each submission. Following previous papers in the literature such as Cookson and Niessner (2020), we assign messages that are posted in the interval of day $t-1$ after 4 p.m. EST to day t before 4 p.m. EST to trading day t (because trading stops at 4

p.m. EST). For each submission or comment, we attribute it to specific companies by tickers, which is normally contained in the title of the submission or the text of the comment. We also collect unique Reddit IDs to identify the authorship of every submission or comment. Clearly, the granularity of the Reddit data allows us to classify agents into different types of investors and study their interactions. Altogether, we have 5,315,487 agent*stock*day observations from Reddit data.

To have a tractable social media information structure, Pedersen (2022) separates all investors into two broad types: hardheaded investors whose opinions do not change, and naïve investors who learn via social networks. 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. Given the fast-paced nature of social media, we use 5 days as the period in which we examine agents’ characteristics. Specifically, we define that an agent k is a hardheaded agent for firm i on day 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 past 5-day window; 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 past 5 days. We consider alternative ways of identifying different types of agents by varying the parameters in both condition 1 and 2, and the results, reported in section 5.1, stay qualitatively similar.

Pedersen further classifies hardheaded investors into rational investors and fanatics. Rational investors gather fundamental-related and value-relevant information in the first stage and don’t change their views for the following periods because they know the truth about firm value. Fanatic agents are also stubborn about their own personal view, but their

views are generally formed without considering any value-relevant information. To define “value-relevant” information, we form a dictionary of words that appear most frequently in Reddit submissions/comments, and hand select value-relevant ones, which are related to firm financial and accounting information.⁴ We define rational agents as hardheaded agents who write at least one post that includes a value-relevant word during the 5-day window. Hardheaded agents whose posts contain no value-relevant words are identified as fanatics. Having identified both types of hardheaded agents, Reddit users who are not hardheaded are classified as naïve investors in our sample.

To examine the reasonableness of our empirical identification design, we create indicator variables for fanatic, rational, and naïve agents, which equals 1 if the Reddit user belongs to each respective agent category, and 0 otherwise. In Table 1 Panel A, we present summary statistics on the agent category measures, by pooling observations across stocks and days. The means of the fanatic and rational indicator variables are 0.079 and 0.044, indicating about 7.9% of agents are fanatics, and 4.4% are rational. The naïve indicator variable has a mean of 0.877 and a standard deviation of 0.161, indicating that 87.7% of Reddit users are naïve investors. In our opinion, the ratio of fanatic, rational, and naïve investors are reasonable as we expect the majority of Reddit users on r/wallstreetbets to behave like naïve investors.

We measure each agent’s beliefs/views using the tones of the submissions and comments, in terms of whether they are positive or negative. To identify the tone, we rely on the word count method as in traditional textual analysis, using the Loughran and McDonald (LM) dictionary. Users of r/wallstreetbets also have their own lingo (e.g.,

⁴ In Appendix A Panel B, we present a list of value-relevant words.

emojis, slang, jokes, and special meaning words), therefore we modify the LM dictionary to better capture the language used in this specific forum. Our modified LM dictionary combines a custom Reddit dictionary of 1,000 most used words and 3 most used emojis on r/wallstreetbets, where each word or emoji is manually assigned positive, neutral, or negative tone based on its context.⁵ For the tone of each submission/comment, we count the number of positive and negative words/emojis in the submission/comment, compute the difference between the two, and divide it by the sum of total number of words and total number of emojis of this submission/comment. That is, for firm i on day t , for each submission/comment m , we first compute,

$$SubmissionTone_{imt} = \frac{\# \text{ of positive words/emojis}_{imt} - \# \text{ of negative words/emojis}_{imt}}{\text{total \# of words/emojis}_{imt}}. \quad (3)$$

The submission tone ranges between -1 and 1, and higher tone indicates more positive views. For agent-level tone on a particular stock across submissions, we take an average across all submissions/comments (M_{ikt}) of an agent k for the stock i on day t . Finally, for firm-level tone across agent types, we average across all agents (K_{ilt}) within each agent type l for the stock i on day t , and compute agent-type*stock*day level variables:⁶

$$AgentTone_{ilt} = \frac{1}{K_{ilt}} \sum_{k=1}^{K_{ilt}} \left(\frac{1}{M_{ikt}} \sum_{m=1}^{M_{ikt}} SubmissionTone_{imt} \right). \quad (3')$$

Following this method, we compute $RationalTone_{it}$, $FanaticTone_{it}$, and $NaiveTone_{it}$.

The importance of an agent in the social network is measured by her influence. As mentioned earlier, Pederson's model mathematically defines two types of influences, thought leadership and influencer value. For the parsimony of empirical estimation, we don't separately estimate thought leadership and influencer value. Instead, we rely on the

⁵ We provide a list of these jargons and emojis and their sentiment values in Appendix A.

⁶ We also report results using influence-weighted tone, where we weight individual's agent tone by influence, in Table 6 Panel A. Main results remain robust.

model’s intuition, and define an influence variable for each agent, which measures how much attention the agent attracts from other users, or how connected this agent is to other agents. We start with defining social networks for every firm-day pair in our sample. The social network for firm i on day t consists of all Reddit users who talk about firm i on day t . User k is connected to user j if k captures j ’s attention. Since we do not have data on the viewership of k ’s submissions or comments, we use the number of commenters on k ’s posts as a close 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 she attracts more people’s attention.⁷

To compute agent-type*stock*day 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 day t as $NCommenter_{ilt} = \sum_{k=1}^{K_{ilt}} NCommenter_{ikt}$. We normalize this variable to compute our influence measure in two steps. In the first step, we compute the natural logarithm of one plus the raw measure, to address the skewness in its distribution. In the second step, 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 day t , influence is calculated as,

$$Influence_{ilt} = \frac{\log(1+NCommenter_{ilt}) - \min_t \{\log(1+NCommenter_{ilt})\}}{\max_t \{\log(1+NCommenter_{ilt})\} - \min_t \{\log(1+NCommenter_{ilt})\}} \quad (4)$$

A higher value of the measure indicates this agent-type for a firm receives more attention from all agent-types and consequently exerts higher influence in the social network. For cross-sectional comparisons, each day we split our sample firms into firms with high-influence networks, and firms with low-influence networks. We identify firms with high-

⁷ In section 5.2, we also use agent’s PageRank score in the network as a proxy for influence and report results in Table 6 Panel B. Our main results remain robust under this alternative definition.

influence networks for each day if the firm has overall influences (total number of commenters) to be above the 90th percentile of the cross-section of all firms, and other firms are classified as firms with low-influence networks.

We present summary statistics on social media activity measures in Table 1 Panel B. The tones of fanatics, rational, and naïve agents have means of 0.005, 0.003 and 0.012, respectively. That is, naïve agents' tones are on average more positive than other agents, while the rational investors have the least positive tones. In terms of standard deviations, naïve agents have more diverse views (a standard deviation of 0.049), while rational agents have the least dispersion (a standard deviation of 0.023). For the influences of different agent types in the network, the means of fanatic, rational and naïve agents are 0.018, 0.016 and 0.034, respectively. That is, the group of naïve agents has the highest value for the influence variable, possibly because there are more naïve agents in the network than other types of agents. Therefore, it is reasonable that naïve agents in aggregate exert more influences in the network. For correlations, the tone measures have relatively low correlations, ranging between 0.06 and 0.12, suggesting different groups of agents differ in their views. The correlations among different influence variables are all above 0.50, suggesting that their influences share similarities across stocks and over time.

To help provide a heuristic understanding of the social media measures, we plot the time series of these measures for January and February of 2021 for GameStop in Figure 1. Panel A presents the proportion of three types of users over time. Among all Reddit users who discuss GameStop, around 1% are fanatic investors, 3% are rational, and 96% are naïve investors, which clearly shows that most of the participants for GameStop discussion are naïve investors. We provide a close-up figure in Panel B to examine the time series

variation in the proportion of fanatic and rational agents. The number of fanatic agents rises sharply from January 24th to January 27th, during which GameStop's stock price increases from \$76.79 to \$347.51. In contrast, the proportion of rational agents falls during the same period. We also present the proportion of naïve agents in Panel C. Interestingly, the proportion of naïve agents exhibits an upward trend during this period, which suggests that GameStop's wild swings in price attract the attention of naïve investors more. We present the tones, and the number of commenters in the next two panels. For average tones in Panel D, the views of rational and naïve investors are relatively stable, while fanatic agents' tones are more volatile and extreme. We plot the number of commenters in Panel E to illustrate the time-series change of each agent type's influence. Between January 24 and January 31, all agent types exhibit similar heightened levels of number of commenters, indicating they receive more attention and thus have higher influences during this period. Overall, the empirical dynamic patterns of proportion and social media activity of each agent type in the case of GME well correspond to Pedersen's model, further supporting the assumption of different agent types in the Pedersen's model.

3.2 Data for returns, retail flows, and shorting flows

Stock data are obtained from CRSP. We retain only common stocks (those with a CRSP share code equal to 10 or 11) and exclude securities such as warrants, preferred shares, American Depositary Receipts, closed-end funds, and REITs. We require a minimum share price of \$1 for a stock to be included. We then cross-match the CRSP data with Reddit data using ticker symbols. To ease the concern that some firms do not have any Reddit activity during the sample period because they are not particularly favored by investors on r/wallstreetbets, we further restrict the sample to firms that have at least one

submission or comment during our sample period. In total, the merged sample has 308,044 stock*day observations.

For each stock i on each day t , we first compute stock returns as follows,

$$Return_{it} = \frac{BidAskAverage_{it} - BidAskAverage_{i,t-1}}{BidAskAverage_{i,t-1}}. \quad (5)$$

Here we choose daily bid-ask average prices, $BidAskAverage_{it}$, for return calculation, to minimize potential biases introduced by bid-ask bounces, as advocated by Blume and Stambaugh (1983).

To compute related retail investor measures, we identify retail investors using sub-penny price improvements in FINRA trade data following BJZZ (2021). That is, for trades with execution prices with a sub-penny portion between \$0.0001 and \$0.0040, we identify them as retail sells, and for trades with execution prices with a price ending between \$0.0061 and \$0.0099, we identify them as retail buys.⁸ We compute the scaled marketable retail flows variable as

$$RetailFlow_{it} = \frac{TotalMarketableRetailBuyVolume_{it} - TotalMarketableRetailSellVolume_{it}}{TotalMarketableRetailBuyVolume_{it} + TotalMarketableRetailSellVolume_{it}}. \quad (6)$$

For short sellers, we define their activity, following Boehmer, Jones and Zhang (2008), using the daily proportional shorting flows for stock i on day t as

$$ShortFlow_{it} = \frac{DailyShortVolume_{it}}{TotalTradingVolume_{it}}. \quad (7)$$

Due to data availability, we obtain daily shorting data from CBOE, the third largest exchange group in the U.S., which represents 20% of on-exchange shorting activity on average. Here the numerator is the total shares sold short in CBOE's short-sale transaction files for stock i on day t , and the denominator is that stock-day's CBOE trading volume.

⁸ In Section 5.3, we use a modified algorithm of Barber et al. (2023) to identify retail flows, and find the results are similar.

Table 1 Panel C provides summary statistics for these other variables. The average daily return is 0.004 and the standard deviation is 0.076. The average retail order imbalance is -0.020, and the standard deviation is 0.282. The average daily shorting flow is 0.522; the standard deviation is 0.161. All these numbers are consistent with previous literature.

3.3 Empirical method

We adopt the panel vector autoregressions (PVAR) approach to conduct our empirical analysis. Following Holtz-Eakin, Newey, and Rosen (1988), we specify the benchmark model as follows:

$$y_{i,t} = \sum_{l=1}^L A_l y_{i,t-l} + DayFE_t + FirmFE_i + \varepsilon_{i,t}, \quad (8)$$

where $y_{i,t}$ is the vector of dependent variables. In Pedersen's model, agents' views both influence and are influenced by other agents' views; agents' beliefs are affected by stock prices and at the same time, have an impact on prices; and investors' trading behaviors are also interdependent with agents' beliefs and stock returns. One advantage of PVAR is that it allows us to capture the dynamics of belief formation, price discovery and trading behaviors in the social network by including agents' beliefs, stock returns, and investor order flows simultaneously in the system. Therefore, we define $y_{i,t} = (FanaticTone_{i,t}, RationalTone_{i,t}, NaiveTone_{i,t}, Return_{i,t}, RetailFlow_{i,t}, ShortFlow_{i,t})'$.⁹ Matrix A_l is a coefficient matrix for lag l , and $l = 1, \dots, L$, is lag length. Given the panel data structure, we include both day fixed effects and firm fixed effects. We estimate the coefficients in (8) by the generalized method of moments (GMM) (Holtz-Eakin, Newey, and Rosen, 1988). The standard errors are double clustered by firm and date.

⁹ 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.

Another advantage of the PVAR approach is that it provides many test statistics for readers to understand the economic intuitions. To be specific, we examine three sets of test statistics. The first set includes the estimates of the A_l matrix, which shed light on the predictive patterns among various variables and the parameter's 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 $A_1(m, n), A_2(m, n), \dots, A_L(m, n)$ are jointly significant according to the Wald test. Intuitively, Granger causality tests help to infer which variable is important to the future outcomes of another variable. The third set is the impulse response functions (IRF) associated with PVAR, which describe how a dependent variable responds to a one-time shock from one independent variable. 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 one standard deviation changes in $y_{i,t-1}(n)$, on day $t+k$. We choose $k = 1, \dots, 10$ to capture the dynamics over 10 days. Examining the impulse response functions helps us to understand how variables within a dynamic system interact with each other.

Previous literature provides 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 3.¹⁰ The second concern for PVAR is that it only allows linear relation in the system and cannot capture the non-linear 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 relation 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 in 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 can be predicted by fanatic and rational investors' views, we expect the coefficients linking current naïve views and past fanatic and rational views to be statistically significant.

Given that the optimal lag length for our PVAR estimation is 3, and given that we have 6 variables in our depending variable vector, we have $6*6*3=108$ parameters in total. For the current section which focuses on belief formation, we only present in Table 2 the first 3 columns, and first 3 rows of the $A_1, A_2,$ and $A_3,$ which define the dynamics of views among different groups of investors.

In Panel A of Table 2, we focus on column III, which describes how naïve tone on day t is related to fanatic and rational tone from day $t-1, t-2$ and $t-3$. All six coefficients of fanatic tone and rational tone for the three lags are positive and significant, implying that

¹⁰ Although PVAR with one lag (PVAR(1)) is easier to read and interpret than PVAR with three lags (PVAR(3)), PVAR(3) is significantly better than PVAR(1), using LR test, with a p-value of less than 0.005. Therefore, we use PVAR(3) all through the paper. We also present results using PVAR(1) in Appendix B, and results are similar.

naïve tone can be predicted by fanatic tone and rational tone. For instance, the coefficients on FanaticTone (t-1) and RationalTone (t-1) are 0.0241 (t-stat=5.43) and 0.0394 (t-stat=6.63), respectively. We notice that the coefficients for the first lag have the largest magnitude and are most significant, suggesting that fanatic and rational views from previous day have the largest impact on the naïve views. We conduct the Granger causality tests and report the p-values at the bottom of the table. In column III, the p-values of past fanatic tone and rational tone Granger causing current naïve tone are both 0.0%, which indicates significant Granger-causal relations. That is, both fanatic and rational tones are important for forecasting naïve tone. Therefore, H1 is supported by the data.

To heuristically understand how different agents' tones within a dynamic system interact with each other, we plot the IRF in Figure 2. The IRFs show the next 10-day reaction of each response variable corresponding to one standard deviation shock of each impulse variable. The 5% confidence bounds (dashed lines) are generated using Monte-Carlo simulations with 1000 draws. Panel A and Panel B present how shocks to fanatic tone and rational tone affect naïve tone for the next 10 days. Both panels clearly show that a positive shock of fanatic tone or rational tone is followed by a significant positive response of naïve tone. For a one standard deviation shock of fanatic tone, the positive response of naïve tone is the highest for day 1, with an increase in the naïve tone of 0.1%. Given that the mean of the naïve tone is 0.012 from Table 1, this effect is economically meaningful. Based on the confidence interval, the response is significantly different from zero. The impact slowly dies out over the next 5 days, and becomes insignificant on day 6. For a one standard deviation shock of rational tone, the response is an increase in naïve

tone of 0.1% for day 1. The response of naïve tone with respect to rational tone shock stays high for the next 2 days and dies out after about 7 days.

Hypothesis H2 focuses on the influence of different agent groups and predicts that the higher the influences of a group of agents, the more they can impact the tones of the other agents. To test this hypothesis, we estimate the specification (8) for two subsamples, high influence group and low influence group, separately, where the separation of the two groups is described in Section 3.1.¹¹ If H2 is supported, then past fanatic and rational views should have a bigger impact on future naïve views in the high influence group than in the low influence group.

We report the estimation results in Panel B and C of Table 2. When we focus on the naïve tone in column III, all six coefficients of fanatic tone and rational tone for the three lags are positive and significant for the high influence subsample in Panel B, while four coefficients are significant for the low influence subsample in Panel C. For instance, the coefficients on FanaticTone (t-1) and FanaticTone (t-2) in Panel B are 0.0491 (t-stat=8.39) and 0.0301 (t-stat=6.01), while in Panel C the coefficients on FanaticTone (t-1) and FanaticTone (t-2) are 0.0176 (t-stat=2.60) and 0.0057 (t-stat=1.06). Similar patterns can be observed for lagged rational tones. From the Granger Causality tests at the bottom, fanatic tone and rational tone Granger cause the future naïve tone, with comparable p-values in both subsamples.

Figure 2 Panel C-D and Panel E-F depict the corresponding IRFs for high and low influence groups, respectively. In the high influence subsample in Panel C, the reaction of

¹¹ We also use the 90th percentile of influence for each agent type (fanatic influence, rational influence, and naïve influence) to divide our sample, and the results are similar. Our main results are also robust when we use the 95th percentile or 85th percentile.

naïve tone in response to a one standard deviation shock of fanatic tone is a 0.4% increase on day 1, and it remains high for the next 2 days. In fact, the impact of shock of fanatic tone on naïve tone is significant for the next 10 days for the high influence sample. However, in the low influence subsample in Panel E, the response of naïve tone to the same fanatic tone shock on day 1 is a 0.05% increase, and it quickly drops to zero after 3 days. Similar patterns can be observed for rational tone in Panel D and F. Clearly, the magnitude of response is much larger in the high influence subsample than in the low influence subsample. Overall, these patterns suggest that fanatic tone and rational tone have stronger impacts on naïve tone when fanatic and rational agents have higher influences, which is consistent with H2.

To summarize, we document empirical support for H1 and H2, regarding belief formation in a social network. Views from fanatic and rational agents are strong predictors of next-day naïve agent views. The impacts of these agents' tones on next-day naïve tone are stronger if they have high influences.

4.2 Social Media Views and Price Discovery

In this section, we test our H3 and H4 and examine whether and how beliefs of different agents, along with their influences, predict next-day stock returns in the dynamic system. We first focus on H3, which establishes that the equilibrium price in a market with naïve, fanatic, and rational agents is a combination of the rational price and a social network price component. If social media views can predict future returns, as in H3, the coefficients linking current returns and past agent views are expected to be significant.

For the current section which focuses on price discovery, we present in Table 3 the part of PVAR(3) that defines relation between past agent views and current returns. For

the whole sample results in column I, the coefficient on FanaticTone (t-1) for predicting Return (t) is 0.0089 (t-stat=1.99), suggesting that past fanatic tone contains predictive information about next-day returns. However, the predictive power diminishes quickly, and coefficients on FanaticTone (t-2) and FanaticTone (t-3) are neither positive nor significant. Also, the Granger Causality test at the bottom of the table shows that the three variables don't jointly predict Return(t). For rational tone variables, the coefficients on RationalTone (t-1), RationalTone (t-2) and RationalTone (t-3) are -0.0025 (t-stat=-0.38), 0.0119 (t-stat=2.25) and -0.0044 (t-stat=-0.89), which suggests that only RationalTone (t-2) has significant predictive power for Return (t). The Granger Causality test also shows that the three variables don't jointly predict Return(t). The pattern changes a bit for naïve tone variables. The coefficients on NaiveTone (t-1), NaiveTone (t-2) and NaiveTone (t-3) are 0.0060 (t-stat=1.43), 0.0059 (t-stat=2.40) and 0.0042 (t-stat=1.77). The Granger Causality test shows that naïve tone Granger causes future returns with a p-value of 1.3%. Overall, we observe that different lags of fanatic, rational and naïve tone predicts future returns.

In Figure 3, we present the corresponding IRFs. Panel A, B, and C present how shocks to fanatic, rational, and naïve tones affect returns for the next 10 days. A one standard deviation shock of fanatic tone is associated with a higher daily return of 3.1 bps increase for day 1. The impact quickly dies out and is insignificant on day 2. For a one standard deviation shock of rational tone or naïve tone, the impact on returns is insignificant for day 1, and becomes significant for day 2, with an increase of 2.8 bps or 2.9 bps, and then dies out. Overall, H3 is supported by our empirical results, in the sense

that different lags of fanatic tone, rational tone, and naïve tone, are significant predictors of future returns.

Hypothesis H4 tests whether the dynamic relation among agents' tones and returns is stronger in networks with higher influences. As before, if influence is important in affecting the predictive power of agent tones on future returns, we expect past agent views to impact future returns more in the high influence subsample.

We report the estimation results of whether agent influence changes tones' predictability on returns in column II and III of Table 3. In the high influence subsample in column II, all coefficients of agent tones are positive and significant. For Granger causal relations, fanatic tone, rational tone, and naïve tone all positively Granger cause next-day returns in the high influence subsample with p-values of 0.0%. The pattern differs significantly in the low influence subsample, however. In the low influence subsample in column III, most of the coefficients on the tone variables are insignificant. It is interesting to find that the coefficient on RationalTone (t-1) and NaiveTone (t-1) are both negative and significant, suggesting that they predict a negative relation in next-day return. Consistently, the Granger causality tests on past native and rational tones are significant.

The IRFs are reported in Figure 3. Panel D-F and Panel G-I display returns' responses to shocks to different tone variables for high and low influence groups respectively. For the high influence subsample in Panel D-F, a one standard deviation shock of fanatic/rational/naïve tone is associated with a higher return of 33/24/92 bps increase for day 1. The magnitude of response is much larger for a shock of naïve tone, indicating the importance of naïve tone in affecting future returns. The impacts of agent tones on returns remain significant over 10 days. In contrast, for the low influence group

in Panel G-I, the same shock leads to decreases in returns, and magnitudes of the changes are smaller. Moreover, the impact only lasts for less than 2 days for this low influence subsample.

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 next-day returns than in a low-influence network.

4.3 Social Network and Trading Dynamics of Retail investors

In this subsection, we investigate the impact of network belief dynamics on retail investors' trading. Our H5 states that social media views predict retail trading, and views from more influential agents have larger impacts on retail flows. If retail investors follow the social media tones to trade, the coefficients linking current retail flows and past agent views are expected to be positive and significant.

To examine the trading dynamics of retail investors, we present the coefficients on how past tone variables predict future retail trading in Table 4 Panel A. We report the estimation results for the whole sample in column I. The coefficients on agent tones are all positive and eight of nine coefficients are significant, suggesting that retail investors are more bullish when agent tones on Reddit are higher. The Granger causality tests for retail flows show that fanatic tone, rational tone, and naïve tone all Granger cause retail flows with p-values less than 1%, which supports the first part of H5. The corresponding IRFs are shown in Figure 4 Panel A-C. When retail order flows are the response variable, the impact of a one standard deviation shock of fanatic tone is initially small and insignificant on day 1, but becomes larger and more significant on day 2, and then gradually dies out

after 5 days. The positive response of retail flows to a one standard deviation shock of rational/naïve tone is 0.1%/0.2% for day 1, and then dies out.

The second part of H5 is about the dynamics among agents' tones and retail order flows in networks with higher or lower influences. If influence is important in affecting the predictive power of agent tones on retail order flows, we expect the predictive relation between past agent views for future retail order flows to be stronger in the high influence group. We report the estimation results in column II and III of Table 4. In the high influence subsample in column II, the coefficients on agent tones are all positive and significant. In contrast, for the low influence sample in column III, the coefficients are mostly insignificant. For Granger causal relations, agents' tones Granger cause retail flows with positive relations in the high influence subsample with p-values less than 1%, while such relations disappear in the low influence subsample. Our results suggest that agent tones positively predict next-day retail flows only if the agents can exert high influence in the network. Overall, the results support our hypothesis that agent tones have larger impacts on retail flows in the more influential network.

The IRFs are shown in Figure 4 Panel D-F and Panel G-I for high and low influence subsamples respectively. In the high influence subsample, the positive response of retail flows to a one standard deviation shock of fanatic/rational/naïve tone is 0.2%/0.4%/0.7% for day 1, goes up for the next 2 days, and then slowly declines. Nonetheless, the positive impact lasts for over 10 days. However, for the low influence sample, the same shock has no significant impact on retail flows. The above results suggest that user tones positively predict next-day retail flows, especially in networks with high influences. That is, when

reddit users come together to exert influence in the social network, their positive views lead to more bullish retail trading.

Given prior findings that marketable retail flows significantly and positively predict future stock returns, how does retail flows' predictive power change with social media activity? To answer the question, we test how past retail flows predict future returns. We report the estimation results for the whole sample in column I of Table 4 Panel B. The coefficients on RetailFlow (t-1), RetailFlow (t-2) and RetailFlow (t-3) are 0.0023 (t-stat=3.98), 0.0006 (t-stat=0.88) and 0.0017 (t-stat=2.37), which are all positive and two of three are significant. The Granger Causality test shows that retail flows Granger cause future returns with a p-value of 0.0%. Figure 4 Panel J shows the corresponding IRF. A one standard deviation shock to retail flows is associated with a higher daily return of 6.5 bps increase for day 1. The impact quickly dies out and is insignificant on day 2. Overall, the results show that retail flows positively predict future returns, even after we include social media variables in the dynamic system.

More interesting, how does retail flows' return predictability change for firms with different agent influences in the network. We report the estimation results for high and low influence subsamples in column II-III of Table 4 Panel B. For the high influence subsample in column II, the coefficients on retail flows are all positive and significant. In contrast, for the low influence subsample in column III, only the coefficient on RetailFlow (t-1) is significant. Both Granger causality tests are significant at 5%.

The corresponding IRFs are shown in Figure 4 Panel K-L. For the high influence subsample, a one standard deviation shock of retail flows is associated with a higher daily return of 47 bps increase for day 1. The impact dies out after 4 days. While for the low

influence subsample, the same shock is associated with a higher daily return of 5 bps for day 1. The impact becomes insignificant on day 2. Overall, the results clearly show that retail flows have stronger predictive power for future returns for firms in social networks with higher influences.¹²

4.4 Social Network and Trading Dynamics of Short Sellers

Turning now to whether short sellers understand the social network effects on prices and how they respond to changes in network belief dynamics. Our H6 states that social media views predict shorting flows, and views from more influential agents have larger impacts on shorting flows. If short sellers understand social network's effect on prices and trade accordingly, either to ride or to burst the bubble, the coefficients linking current shorting flows and past agent views are expected to be significant.

In Table 5 Panel A, we present the coefficients of past tone variables predicting future shorting flows, which reflects the dynamics between agent views and shorting flows. We report the estimation results for the whole sample in column I. The coefficient on NaïveTone (t-1) is positive and significant, suggesting that short sellers increase their shorting flows when naïve view is more positive. This result suggests that short sellers may view naïve investors as noise traders whose actions could lead to temporary positive bubbles; therefore, short sellers increase shorting as they believe they could profit from bursting the short-term bubbles. The Granger causality tests for shorting flows show that naïve tone positively Granger causes shorting flows with p-value of 3.8%, further supporting the first part of H6. The corresponding IRFs are shown in Figure 5 Panel A-C.

¹² We also show the cumulative impulse response functions corresponding to Figure 4 in Appendix C. The results are similar, but the magnitudes are larger.

The response of shorting flows to a one standard deviation shock of naïve tone is the highest on day 1, with an increase of 0.1%. The impact lasts for 5 days, and then gradually declines.

The second part of H6 is regarding whether the dynamics between agents' tones and shorting flows vary with agent influence. We report these results in column II and column III of Table 5 Panel A. For high influence subsample in column II, the coefficients on agent tones are all negative. However, the Granger tests don't show statistical significance. For low influence subsample in column III, most of the coefficients are positive and two of them are significant. Specifically, the coefficient on FanaticTone (t-1) is 0.0205 (t-stat=2.42), and the coefficient on NaïveTone (t-1) is 0.0287 (t-stat=3.48). The Granger causality tests further show that naïve tone positively Granger causes shorting flows with p-value of 0.2%. The results show that the previous positive relation between naïve agent tones and shorting flows is mainly driven by the low influence subsample, suggesting that short sellers may choose to trade against agents' positive views (i.e. burst the bubble) only when agents' influences are lower. Considering the earlier results in section 4.2 that when agent influence is lower, naïve agents' views predict negative next-day returns, short sellers' action of bursting the bubble in this situation is thus consistent with their goal of profiting from the price decline. When agents' influences are higher however, agents' views are negatively related to shorting flows, consistent with short sellers shying away from shorting (i.e., ride the bubble) when agents' influences are higher, although the coefficients are not statistically significant.

The negative relation between shorting flows and agents' views in the high influence subsample becomes more significant in IRFs in Figure 5 Panel D-F (high influence subsample) In the high influence subsample, the impact of a one standard

deviation shock to fanatic/rational/naive tone is initially small and insignificant on day 1 and 2, but becomes significantly negative on day 3. The magnitude of the decrease in returns on day 3 for the shock to fanatic/rational/naive tone is 0.2%/0.2%/0.3%, respectively. The negative reaction stays significant from day 3 to day 10, indicating that the agent tones have a relatively long-lasting negative effect on shorting flows in the high influence subsample. These results further support that short sellers shy away from shorting when reddit tones are more positive and when agent influence in the network is higher.

We present the results on IRFs of the low influence subsample in Panel G-I. The impact of a one standard deviation shock to naïve tone on shorting flows is positive on day 1, with an increase of 0.1%, and then gradually dies out, suggesting that agent tones not only do not deter shorting flows, but they may even have a small effect of increasing shorting flows in the low influence subsample. That is, in a network with lower influence, short sellers, instead of feeling threatened by agents' bullish sentiment, increase shorting flows when naïve view is positive. To summarize, we find supporting evidence for our hypothesis H6 that Reddit tones significantly impact shorting flows and short sellers ride or burst bubbles depending on the costs and benefits of doing so, and one type of such costs/benefits relates to agent influence.

Previous research finds that shorting flows negatively predict future stock returns (Boehmer, Jones, and Zhang, 2008). Given the current results in Figure 5 which tentatively show that higher Reddit tones increase shorting flows in network with low influence and deter shorting flows in network with high influence, how does shorting flows' predictive power change with social media activity? To answer the question, we present in Table 5 Panel B the coefficients on past shorting flows predicting future returns. We report the

estimation results for the whole sample in column I. The coefficient on ShortFlow (t-1) is -0.0043 (t-stat=-2.74), which is negative and significant. The Granger causality test shows that shorting flows Granger cause future returns with a p-value of 1.0%. Figure 5 Panel J shows the corresponding IRF. A one standard deviation shock of shorting flows is associated with a lower daily return of 6.9 bps decrease for day 1. The impact quickly dies out and becomes insignificant on day 2. The results for the whole sample suggest that shorting flows still negatively predict future returns when including social media variables in the dynamic system.

We are more interested in understanding how the return predictability of shorting flows changes with agent influence in the network. We report the estimation results for high and low influence subsample in column II-III of Table 5 Panel B. For the high influence subsample in column II, the coefficient on ShortFlow (t-1) is negative and significant. The predictive power dies away quickly, since the coefficient on ShortFlow (t-2) is insignificant, and the coefficient on ShortFlow (t-3) even becomes positive. For Granger causal relations, shorting flows Granger cause returns with negative relations in the high influence subsample with p-value of 0.6%. In contrast, for the low influence subsample in column III, the coefficient on ShortFlow (t-1) is negative but insignificant, and the coefficients on ShortFlow (t-2) and ShortFlow (t-3) are both positive. The Granger causality tests also show no significant Granger causal relations of shorting flows and returns in the low influence subsample. It indicates that the negative predictive power of short sellers is stronger when agent influence is higher.

The corresponding IRFs are shown in Figure 5 Panel K-L. For the high influence subsample, a one standard deviation shock of shorting flows is associated with a lower

daily return of 59 bps decrease for day 1. While for the low influence subsample, the same shock is associated with insignificant response in returns. The results show that shorting flows have larger impact on future returns in the more influential network. This is quite intriguing because earlier results in Figure 5 suggest that short sellers on average may shy away from shorting when they worry that more positive social media tones could lead to short squeeze risk when agent influence is high. The results in this section further indicate that networks with high agent influence may not hurt but rather enhance short sellers' return predictability. Under heightened social media activity, shorts sellers carefully consider the costs and benefits of shorting and will short only if they are convinced that the benefits outweigh the costs.¹³

5. Robustness Checks and Further Discussion

In this section, we conduct several robustness checks and provide further discussion in addition to our main results.

5.1 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 A. We start by considering three alternative cases to identify fanatic and rational agents. For the first case, we identify hardheaded agents using the submission and comment activity of the previous 10 days, rather than the previous 5 days. We present the estimation results in column I-II. For the second case, we identify hardheaded agents as those who post more than 99% of all other agents, instead of 95% as in the main results (column III-IV). For the third case, we require that hardheaded agents' posts have the same sign in tones (either positive or negative) for 100% of their posts,

¹³ We also show the cumulative impulse response functions corresponding to Figure 5 in Appendix D. The results are similar, but the magnitudes are larger.

instead of only 75% of their posts during the 5-day window (column V-VI). For the fourth case, we compute influence-weighted tone to highlight the importance of agent influence in social networks, rather than use the tone as average across all individuals (column VII-VIII). In almost all cases, fanatic tone and rational tone positively predict future naïve tone. Moreover, fanatic tone positively predicts next-day return. 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.

5.2 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 the 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 *traffic*:

$$Traffic_{ilt} = \log(1 + \sum_{k=1}^{K_{ilt}} \#Post_{ikt}), \quad (9)$$

where $\#Post_{ikt}$ is the number of submissions and comments posted by agent k for stock i on day t , and K_{ilt} is the number of agents belonging to an agent group l for every stock i on day t . The measure of traffic naturally reflects attention from submitters, with higher traffic means higher attention from that agent type.¹⁴

We re-estimate specification (8) by replacing agent tones with agent traffic. The estimation results are reported in Table 6 Panel B. The coefficients of past agent traffic for

¹⁴ We also compute other measures for social media activities, such as concentration of network, dispersion among different type of investors. Since these measures are not directly linked to Pederson's theory, we don't include them in the main text. These results are available on request.

future returns are mostly insignificant, indicating that agents' traffic cannot predict next-day returns in the dynamic system.¹⁵

5.3 Use PageRank to Proxy for Influence

Next, we consider Google PageRank as an alternative measure of influence. The 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. To consider the potential different network effects of a small network from that of a big network, we use the network size (i.e., the number of agents in the network) weighted PageRank instead of the raw measure. We also sum across the PageRank measures of all agents for every stock i on day t , and compute stock*day level variables.

Results using PageRank are reported in Panel B of Table 6 (Columns III – VI). In the high influence subsample, we still find that agent tones positively predict return. In the low influence subsample, we find that agent tones predict return 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 next-day returns only in networks with higher influences.

5.4 Use BHOS Algorithm for Retail Order Flows

In the main results, we use BJZZ (2021) algorithm to identify retail investors and their order flows. A recent study by Barber et al. (2023) provides a modified algorithm to

¹⁵ As shown in Cookson et al. (2023), sentiment and attention contain different return-relevant information. They find that sentiment-induced retail trading imbalance predicts positive next-day returns while attention-induced retail trading imbalance predicts negative next-day returns. It is consistent with our results that agents' tones predict next day return while agents' traffic does not, as sentiment-induced retail trading may contain return-relevant information while attention-induced retail trading may be just noise trading.

identify retail trades (BHOS algorithm). The two algorithms have different methods of signing a trade. Specifically, BJZZ (2021) use the sub-penny digit to sign a trade as a buy or sell, while Barber et al. (2023) modify the algorithm by signing trades using the quoted spread midpoints. As a robustness check, we re-estimate the results for retail flows using the new algorithm and present the results in Panel B of Table 6 (Column VII). We find that agent tones still positively predict future retail flows. Overall, our main results remain robust to this alternative algorithm used to identify retail trades.

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 directly collecting data from social media platform Reddit. Our results generally support the theoretical predictions. First, for belief formation, we find opinions of fanatic and rational agents positively and significantly predict future opinions of naïve investors, especially in a network with higher agents' 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 capital market. Finally, for trading dynamics, higher 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. In networks

with low influence agents, short sellers seem to short more against more bullish social media tones, which are associated with a negative return on the next day, making shorts potentially profitable. By contrast, 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. In addition, we find that short sellers' negative return predictive power is stronger in the social network with high influence, suggesting that high agent influence does not necessarily hurt but rather enhances short sellers' return predictability. These patterns support Pedersen's prediction that rational investors may ride the bubble 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. 2023. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ. *Journal of Finance, forthcoming*.
- 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*.
- 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., 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. 2023. Place Your Bets? The Value of Investment Research on Reddit's Wallstreetbets. *Review of Financial Studies*.
- 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., R. Lu, W. Mullins, and M. Niessner. 2023. The social signal. *Working paper*.
- 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.
- 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.
- Fusari, N., R. Jarrow, S. Lamichhane. 2022. Testing for asset price bubbles using options data. *Working paper*.
- Hendershott, T., D. Livdan, and N. Schürhoff. 2015. Are institutions informed about news? *Journal of Financial Economics* 117: 249-287.
- 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.
- 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*.
- 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 Feb 15, 2021, 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 161,599 firm-day observations. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable view that are 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: $(\text{number of positive words and emojis} - \text{number of negative words and emojis}) / (\text{number of words} + \text{number of emojis})$. 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 daily CBOE short volume divided by total CBOE trading volume.

Panel A. Proportion of three agent groups

	mean	std	median
Fanatic agent	0.079	0.146	0.000
Rational agent	0.044	0.107	0.000
Naïve agent	0.877	0.161	0.947

Panel B. Social media activity measures

	mean	std	correlation					
			Fanatic Tone	Rational Tone	Naïve Tone	Fanatic Influence	Rational Influence	Naïve Influence
FanaticTone	0.005	0.034	1					
RationalTone	0.003	0.023	0.06	1				
NaïveTone	0.012	0.049	0.12	0.12	1			
FanaticInfluence	0.018	0.082	0.27	0.17	0.13	1		
RationalInfluence	0.016	0.082	0.13	0.35	0.12	0.53	1	
NaïveInfluence	0.034	0.109	0.22	0.33	0.19	0.68	0.76	1

Panel C. Other measures

	mean	std	P50
Return	0.004	0.076	0.000
RetailFlow	-0.020	0.282	-0.013
ShortFlow	0.522	0.161	0.535

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 Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are 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). Influence measures each agent group's influence on investors, computed as the sum of the number of commentors of each agent group. Parameters are estimated using PVAR with GMM estimation with lag length $L=3$. 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
	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)
FanaticTone(t-1)	0.0692*** [10.79]	0.0113*** [4.40]	0.0241*** [5.43]	0.1202*** [9.90]	0.0337*** [6.46]	0.0491*** [8.39]	0.0398*** [5.74]	-0.0050* [-1.83]	0.0176*** [2.60]
RationalTone(t-1)	0.0079 [1.37]	0.1491*** [17.44]	0.0394*** [6.63]	0.0530*** [5.33]	0.2050*** [14.30]	0.0830*** [10.08]	-0.0194*** [-3.03]	0.0987*** [9.25]	0.0278*** [2.81]
NaïveTone(t-1)	0.0164*** [5.95]	0.0068*** [4.20]	0.0198*** [5.42]	0.1480*** [9.92]	0.1259*** [12.17]	0.1855*** [13.01]	0.0013 [0.46]	-0.0096*** [-5.73]	0.0028 [0.71]
FanaticTone(t-2)	0.0282*** [6.31]	0.0088*** [3.57]	0.0109*** [2.82]	0.0577*** [6.33]	0.0333*** [5.54]	0.0301*** [6.01]	0.0137*** [2.74]	-0.0054** [-2.34]	0.0057 [1.06]
RationalTone(t-2)	0.0175*** [2.74]	0.0593*** [10.61]	0.0314*** [5.76]	0.0546*** [4.59]	0.0969*** [9.97]	0.0485*** [6.85]	-0.0075 [-1.17]	0.0258*** [3.53]	0.0316*** [3.72]
NaïveTone(t-2)	0.0087*** [3.81]	0.0089*** [6.64]	0.0176*** [5.45]	0.0960*** [8.89]	0.0979*** [12.69]	0.0918*** [9.32]	0.0003 [0.14]	-0.0018 [-1.40]	0.0123*** [3.53]
FanaticTone(t-3)	0.0203*** [5.24]	0.0090*** [3.62]	0.0105*** [2.65]	0.0476*** [6.20]	0.0265*** [4.58]	0.0270*** [5.72]	0.0049 [1.14]	-0.0025 [-0.87]	0.0051 [0.87]
RationalTone(t-3)	0.0046 [0.80]	0.0362*** [7.39]	0.0341*** [5.88]	0.0219** [2.00]	0.0665*** [6.99]	0.0406*** [6.25]	-0.0053 [-0.92]	0.0114** [2.30]	0.0400*** [4.32]
NaïveTone(t-3)	0.0115*** [5.48]	0.0070*** [5.50]	0.0207*** [6.95]	0.0824*** [8.33]	0.0821*** [10.21]	0.0838*** [9.06]	0.0041** [1.97]	-0.0026** [-2.32]	0.0153*** [4.76]
Number of observations	245002	245002	245002	25554	25554	25554	219448	219448	219448
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%		3.6%	3.4%
Past RationalTone	0.9%		0.0%	0.0%		0.0%	1.1%		0.0%
Past NaïveTone	0.0%	0.0%		0.0%	0.0%		27.3%	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 Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are 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). Influence measures each agent group's influence on investors, computed as the sum of the number of commentors of each agent group. Parameters are estimated using PVAR with GMM estimation with lag length L=3. 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)
FanaticTone(t-1)	0.0089** [1.99]	0.0449*** [4.44]	-0.0071 [-1.58]
RationalTone(t-1)	-0.0025 [-0.38]	0.0426*** [2.87]	-0.0186*** [-2.89]
NaïveTone(t-1)	0.0060 [1.43]	0.1775*** [6.02]	-0.0121*** [-3.43]
FanaticTone(t-2)	0.0042 [0.65]	0.0295* [1.71]	-0.0048 [-1.30]
RationalTone(t-2)	0.0119** [2.25]	0.0379*** [3.54]	0.0034 [0.65]
NaïveTone(t-2)	0.0059** [2.40]	0.0606*** [4.23]	0.0036 [1.45]
FanaticTone(t-3)	-0.0005 [-0.12]	0.009 [1.06]	-0.0035 [-0.92]
RationalTone(t-3)	-0.0044 [-0.89]	0.0183** [2.06]	-0.0113* [-1.80]
NaïveTone(t-3)	0.0042* [1.77]	0.0695*** [5.87]	-0.0012 [-0.53]
Number of observations	245002	25554	219448
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past FanaticTone	22.4%	0.0%	19.7%
Past RationalTone	14.1%	0.0%	1.2%
Past NaïveTone	1.3%	0.0%	0.2%

Table 4. Social Media Activity Associated with Retail Flows

This table presents results on retail flows. Our sample period is Jan 2, 2020 to Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are 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). Influence measures each agent group's influence on investors, computed as the sum of the number of commentors of each agent group. Parameters are estimated using PVAR with GMM estimation with lag length L=3. 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)
FanaticTone(t-1)	0.0191* [1.67]	0.0276** [2.28]	0.0114 [0.70]
RationalTone(t-1)	0.0458*** [3.11]	0.0790*** [4.56]	0.0148 [0.64]
NaïveTone(t-1)	0.0328** [2.32]	0.1284*** [4.40]	0.0191 [1.27]
FanaticTone(t-2)	0.0238** [2.54]	0.0228* [1.87]	0.0232* [1.81]
RationalTone(t-2)	0.0318** [2.48]	0.0435*** [2.97]	0.0172 [0.86]
NaïveTone(t-2)	0.0063 [0.71]	0.0659*** [2.87]	-0.0014 [-0.15]
FanaticTone(t-3)	0.0200** [2.02]	0.0308** [2.57]	0.0128 [0.95]
RationalTone(t-3)	0.0230* [1.85]	0.0455*** [3.18]	0.0026 [0.13]
NaïveTone(t-3)	0.0226*** [2.72]	0.0951*** [4.71]	0.0123 [1.37]
Number of observations	245002	25554	219448
p-value of Granger causality test	RetailFlow(t)	RetailFlow(t)	RetailFlow(t)
Past FanaticTone	0.6%	0.6%	21.0%
Past RationalTone	0.0%	0.0%	72.1%
Past NaïveTone	0.7%	0.0%	31.7%

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)	0.0023*** [3.98]	0.0305** [2.44]	0.0016*** [3.26]
RetailFlow(t-2)	0.0006 [0.88]	0.0257* [1.65]	-0.0002 [-0.40]
RetailFlow(t-3)	0.0017** [2.37]	0.0273* [1.80]	0.0007 [1.63]
Number of observations	245002	25554	219448
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past RetailFlow	0.0%	4.6%	0.1%

Table 5. Social Media Activity Associated with Shorting Flows

This table presents results on shorting flows. Our sample period is Jan 2, 2020 to Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are 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). Influence measures each agent group's influence on investors, computed as the sum of the number of commentors of each agent group. Parameters are estimated using PVAR with GMM estimation with lag length L=3. 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)
FanaticTone(t-1)	0.0084 [1.29]	-0.0079 [-0.93]	0.0205** [2.42]
RationalTone(t-1)	-0.0033 [-0.36]	-0.0091 [-0.79]	0.0044 [0.37]
NaïveTone(t-1)	0.0223*** [2.84]	-0.023 [-1.52]	0.0287*** [3.48]
FanaticTone(t-2)	-0.0046 [-0.84]	-0.0107 [-1.37]	-0.0011 [-0.15]
RationalTone(t-2)	0.0003 [0.04]	-0.0162 [-1.36]	0.0134 [1.16]
NaïveTone(t-2)	0.0041 [0.99]	-0.0204 [-1.63]	0.0071 [1.63]
FanaticTone(t-3)	-0.0105* [-1.84]	-0.0145* [-1.73]	-0.0062 [-0.84]
RationalTone(t-3)	-0.0084 [-1.04]	-0.0177 [-1.60]	0.0005 [0.05]
NaïveTone(t-3)	0.0042 [1.00]	-0.0202 [-1.55]	0.0069 [1.60]
Number of observations	245002	25554	219448
p-value of Granger causality test	ShortFlow(t)	ShortFlow(t)	ShortFlow(t)
Past FanaticTone	10.2%	14.5%	7.9%
Past RationalTone	76.2%	15.8%	68.4%
Past NaïveTone	3.8%	21.9%	0.2%

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)	-0.0043*** [-2.74]	-0.0522*** [-3.13]	-0.0008 [-0.56]
ShortFlow(t-2)	0.0019 [1.08]	-0.0175 [-0.79]	0.0034** [2.47]
ShortFlow(t-3)	0.0023 [1.46]	0.0342* [1.92]	0.0011 [0.79]
Number of observations	245002	25554	219448
p-value of Granger causality test	Return(t)	Return(t)	Return(t)
Past ShortFlow	1.0%	0.6%	5.6%

Table 6. Robustness Check and Further Discussion

This table presents results on robustness check and further discussion. Our sample period is Jan 2, 2020 to Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Panel A reports the estimation results using alternative agent classifications and alternative measures for tones. In column I-II, we identify hardheaded agents using the submission and comment activity of the previous 10-day, rather than the previous 5-day; in column III-IV, 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 V-VI, 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 VII-VIII, 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. Panel B reports the estimation results using alternative proxies for social media activity, influence, and retail order flow. 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 network size weighted 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. Column VII presents the estimation results using the modified algorithm to identify retail trades following Barber et al. (2023). Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are 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: $(\text{number of positive words and emojis} - \text{number of negative words and emojis}) / (\text{number of words} + \text{number of emojis})$. Parameters are estimated using PVAR with GMM estimation with lag length $L=3$. 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. Alternative measures for agents and tone

Alternative measures	I	II	III	IV	V	VI	VII	VIII
	Use past 10 days' information in the network NaïveTone(t)	Return(t)	Use P99 of number of posts as threshold of hardheaded NaïveTone(t)	Return(t)	Require stable tones for past 5 days for hardheaded NaïveTone(t)	Return(t)	Influence-weighted tone NaïveTone(t)	Return(t)
FanaticTone(t-1)	0.0260*** [5.53]	0.0074* [1.73]	0.0229*** [4.94]	0.0090* [1.94]	0.0233*** [5.10]	0.0081* [1.70]	0.0172*** [4.00]	0.0130** [2.05]
RationalTone(t-1)	0.0377*** [6.71]	-0.0025 [-0.44]	0.0388*** [6.51]	0.0010 [0.15]	0.0394*** [6.95]	-0.0128* [-1.73]	0.0673*** [10.00]	-0.0100 [-0.95]
NaïveTone(t-1)	0.0193*** [5.21]	0.0063 [1.51]	0.0202*** [5.53]	0.0061 [1.46]	0.0200*** [5.41]	0.0057 [1.34]	0.0320*** [8.43]	0.0057 [1.41]
FanaticTone(t-2)	0.0077* [1.90]	0.0030 [0.48]	0.0112*** [2.77]	0.0007 [0.15]	0.0106*** [2.67]	0.0048 [0.67]	0.0103*** [2.69]	0.0029 [0.50]
RationalTone(t-2)	0.0293*** [5.64]	0.0135*** [2.92]	0.0306*** [5.70]	0.0135*** [2.73]	0.0365*** [6.29]	0.0113** [1.97]	0.0379*** [5.29]	0.0182** [2.34]
NaïveTone(t-2)	0.0185*** [5.70]	0.0058** [2.32]	0.0180*** [5.56]	0.0061** [2.45]	0.0176*** [5.39]	0.0059** [2.36]	0.0276*** [7.02]	0.0077** [2.11]
FanaticTone(t-3)	0.0060 [1.49]	0.0038 [1.04]	0.0105** [2.55]	0.0005 [0.12]	0.0090** [2.26]	0.0013 [0.33]	0.0067* [1.79]	-0.0011 [-0.24]
RationalTone(t-3)	0.0312*** [5.77]	-0.0040 [-0.87]	0.0368*** [6.46]	-0.0036 [-0.79]	0.0329*** [5.26]	-0.0015 [-0.29]	0.0281*** [4.24]	-0.0039 [-0.73]
NaïveTone(t-3)	0.0216*** [7.12]	0.0033 [1.37]	0.0210*** [7.04]	0.0039* [1.67]	0.0211*** [7.08]	0.0041* [1.69]	0.0208*** [5.82]	0.0041 [1.27]
Number of observations	245002	245002	245002	245002	245002	245002	245002	245002

Panel B. Alternative proxies for social media activity, influence, and retail flows

Alternative proxies	I Traffic to proxy for social media activity			III Pagerank to proxy for influence, high influence		V Pagerank to proxy for influence, low influence		VII BHOS algorithm for retail order flow
	NaïveTraffic(t)	Return(t)		NaïveTone(t)	Return(t)	NaïveTone(t)	Return(t)	RetailFlow(t)
FanaticTraffic (t-1)	0.0715*** [12.55]	-0.0001 [-0.29]	FanaticTone (t-1)	0.0491*** [8.39]	0.0449*** [4.44]	0.0176*** [2.60]	-0.0071 [-1.58]	0.0169* [1.72]
RationalTraffic (t-1)	0.1288*** [20.10]	-0.0004 [-0.85]	RationalTone (t-1)	0.0830*** [10.08]	0.0426*** [2.87]	0.0278*** [2.81]	-0.0186*** [-2.89]	0.0162 [1.12]
NaïveTraffic (t-1)	0.3649*** [51.59]	-0.0002 [-0.62]	NaïveTone (t-1)	0.1855*** [13.01]	0.1775*** [6.02]	0.0028 [0.71]	-0.0121*** [-3.43]	0.0221* [1.80]
FanaticTraffic (t-2)	-0.0071 [-1.30]	0.0004 [0.99]	FanaticTone (t-2)	0.0301*** [6.01]	0.0295* [1.71]	0.0057 [1.06]	-0.0048 [-1.30]	0.0147* [1.71]
RationalTraffic (t-2)	0.0008 [0.13]	0.0006 [1.30]	RationalTone (t-2)	0.0485*** [6.85]	0.0379*** [3.54]	0.0316*** [3.72]	0.0034 [0.65]	0.0273** [2.31]
NaïveTraffic (t-2)	0.1435*** [26.02]	0.0002 [0.95]	NaïveTone (t-2)	0.0918*** [9.32]	0.0606*** [4.23]	0.0123*** [3.53]	0.0036 [1.45]	0.0043 [0.61]
FanaticTraffic (t-3)	-0.0079 [-1.57]	-0.0002 [-0.58]	FanaticTone (t-3)	0.0270*** [5.72]	0.0090 [1.06]	0.0051 [0.87]	-0.0035 [-0.92]	0.0163** [2.02]
RationalTraffic (t-3)	-0.0197*** [-3.47]	-0.0008** [-2.31]	RationalTone (t-3)	0.0406*** [6.25]	0.0183** [2.06]	0.0400*** [4.32]	-0.0113* [-1.80]	0.0086 [0.73]
NaïveTraffic (t-3)	0.1383*** [24.04]	0.0005* [1.78]	NaïveTone (t-3)	0.0838*** [9.06]	0.0695*** [5.87]	0.0153*** [4.76]	-0.0012 [-0.53]	0.0177*** [2.62]
Number of observations	245002	245002	Number of observations	25554	25554	219448	219448	245393

Figure 1. Distribution of Reddit Activities of Agent Group for GME from Jan.1 to Feb. 15, 2021

These graphs present the distribution of Reddit activities of three agent groups for GME from Jan.1 to Feb. 15, 2021. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents.

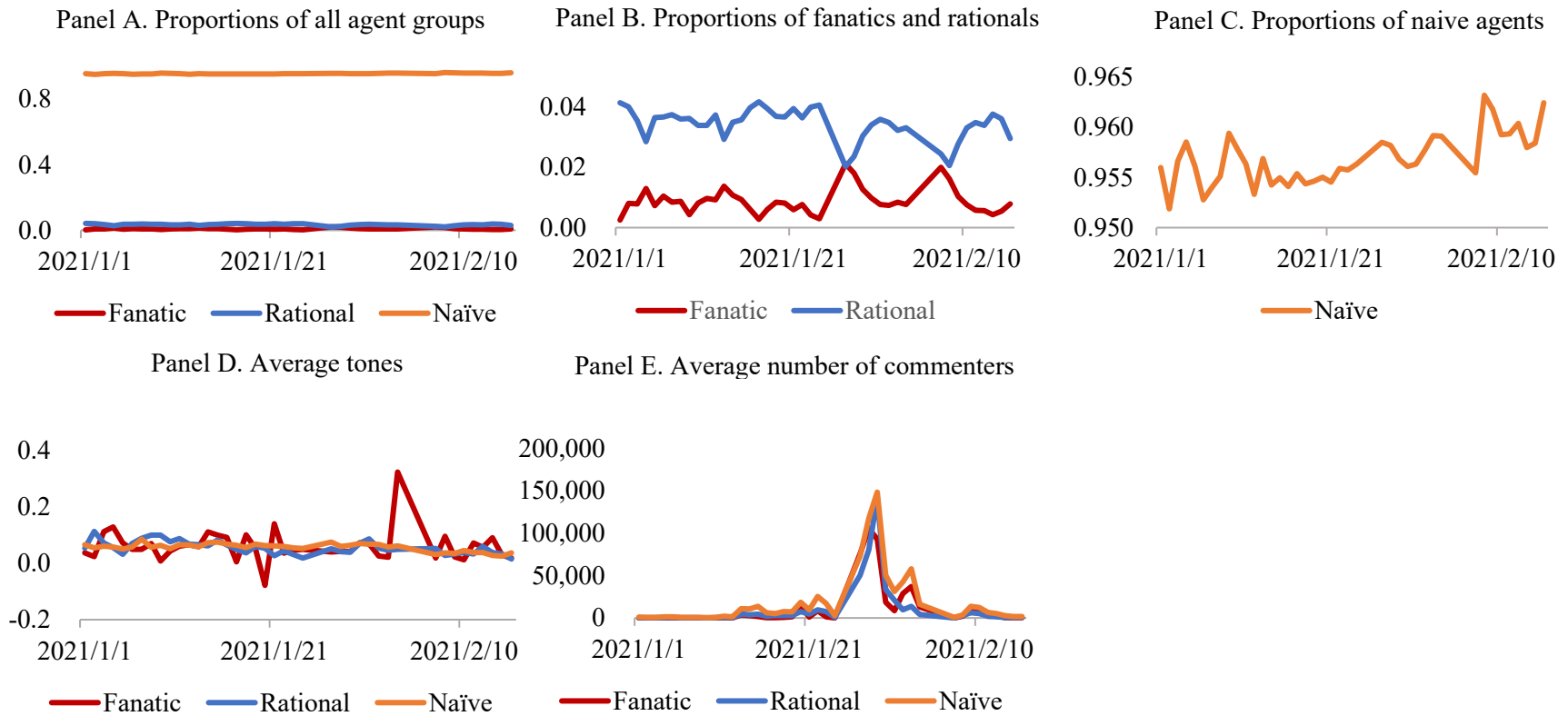


Figure 2. Impulse Responses for Agents' Tones

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.

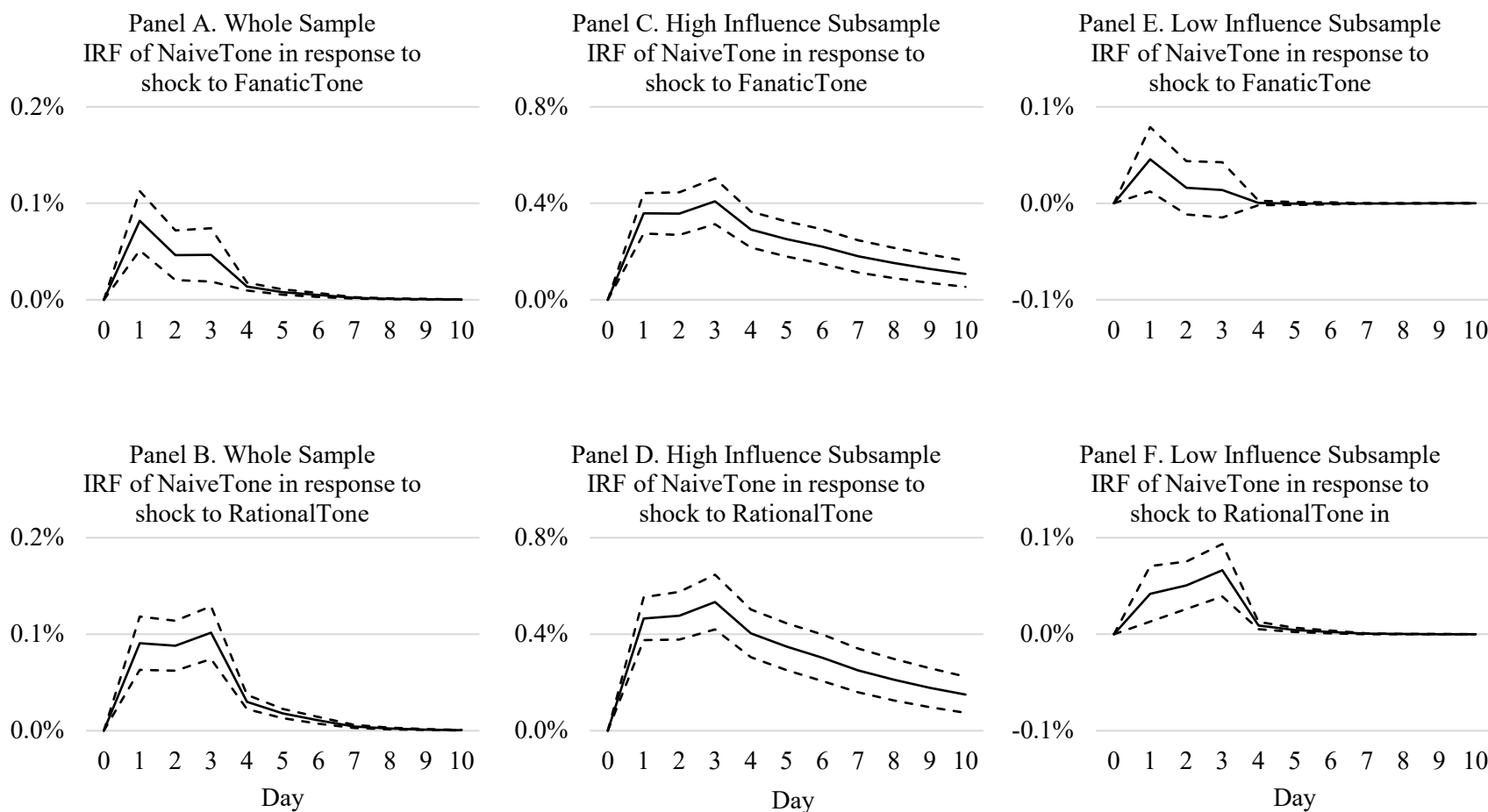


Figure 3. Impulse Responses for Agents' Tones and Returns

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.

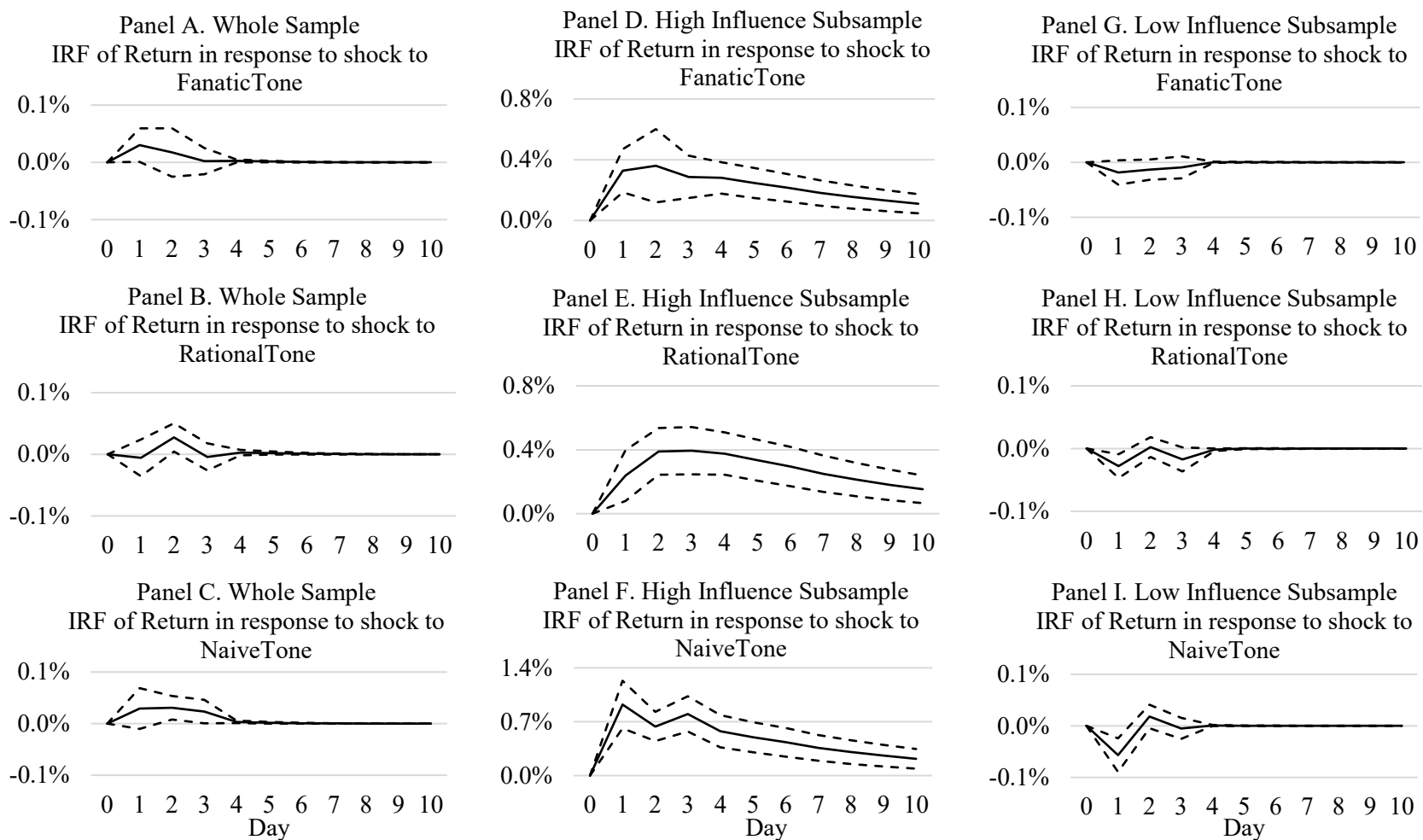


Figure 4. Impulse Responses for Agents' Tones and Retail Flows

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.

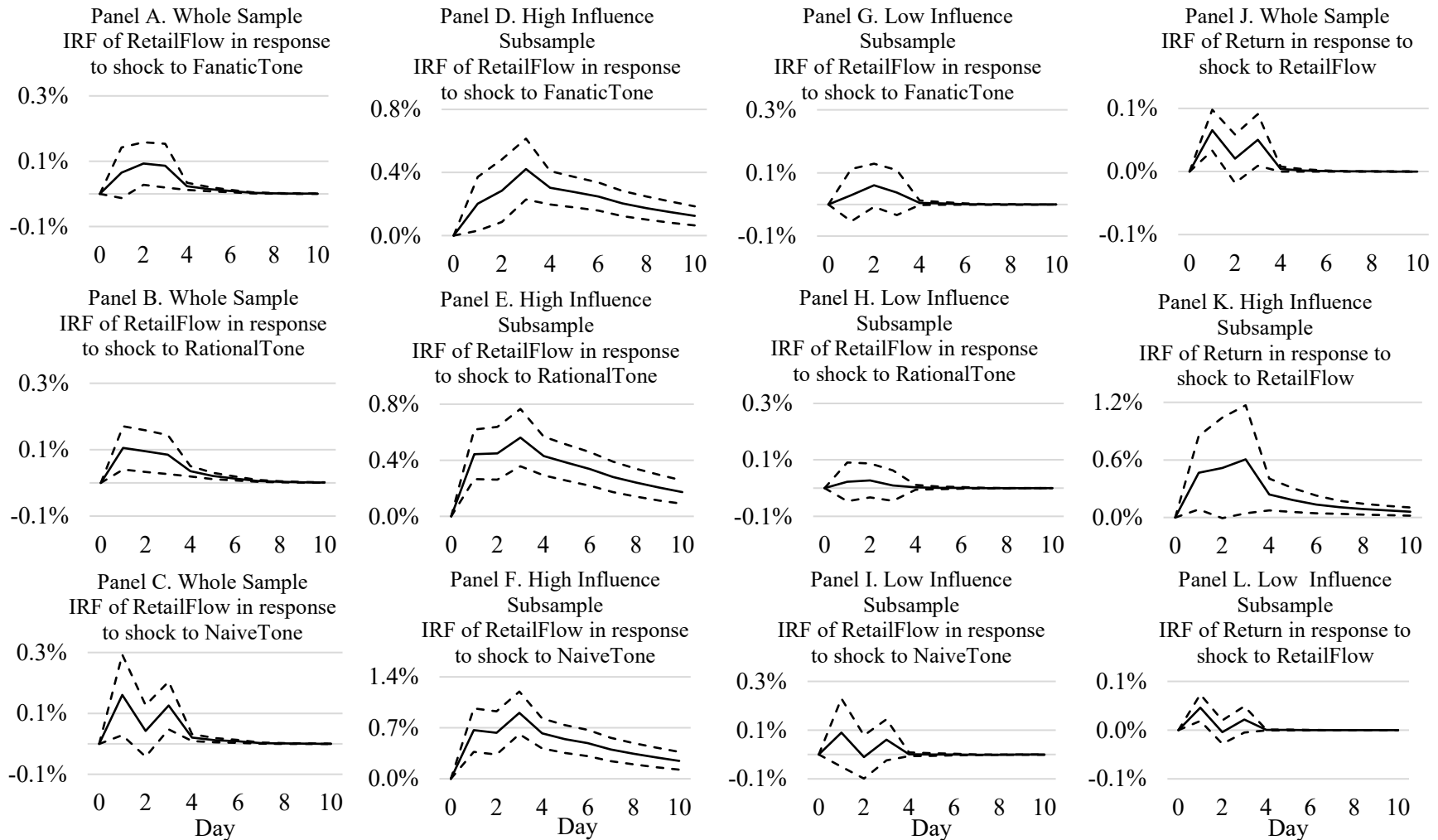
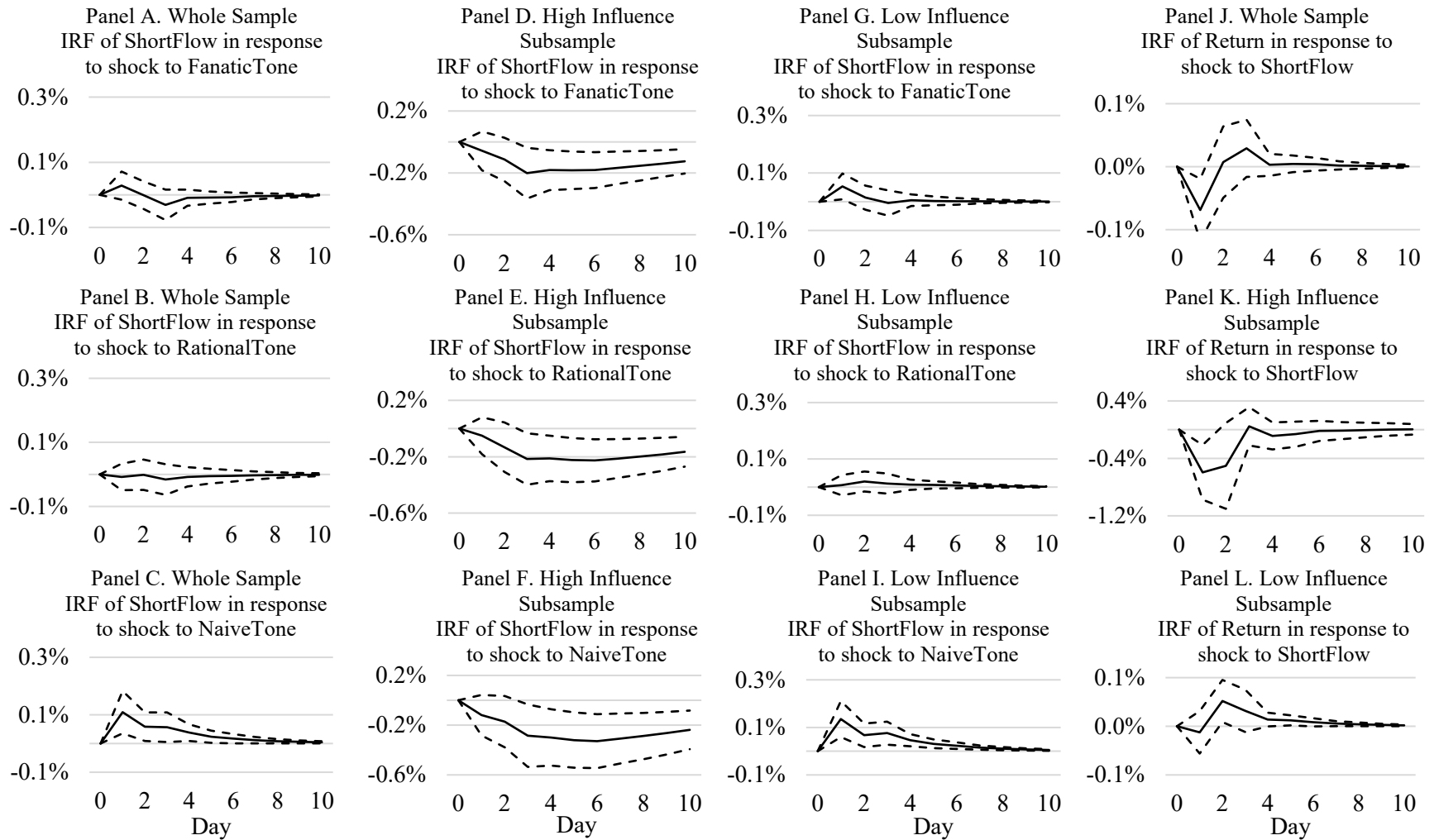


Figure 5. Impulse Responses for Agents' Tones and Shorting Flows

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., 2021). 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 notice that there are 3 popular emojis. We include them in our sentiment dictionary as well, and report them 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	hand	fun	hold	love	rocket	super	asshol	dont	kill	sold
awesom	call	get	hope	million	pump	sure	bear	dumb	piss	sorri
beat	certain	glad	invest	moon	purchas	tendi	bitch	fake	put	stupid
big	correct	go	join	nice	rich	thank	boomer	fall	restrict	suck
bless	crush	gold	jump	power	right	trust	broke	fomo	rip	tank
bought	decent	got	leap	pump	rise	upvot	bullshit	fucker	scare	wtf
break	diamond	grow	legend	purchas	rocket	well	crash	hate	sell	
bull	energi	growth	legit	rich	safe		dead	hit	shit	
bullish	fine	high	like	right	smart		delet	idiot	shitpost	
buy	free	higher	long	rise	solid		die	issu	shitti	

Panel B. Words in value dictionary

dd	product	revenu	liquid	industri	grow	merger	illeg
earn	data	info	growth	fundament	loan	asset	oper
news	cap	store	debt	cut	suppli	dividend	manufactur
valu	announc	releas	demand	ipo	analysi	valuat	guidanc
report	target	research	ceo	quarter	bankrupt	undervalu	

Panel C. Most used emojis

Diamond		Hand		Rocketship	
---------	---	------	---	------------	---

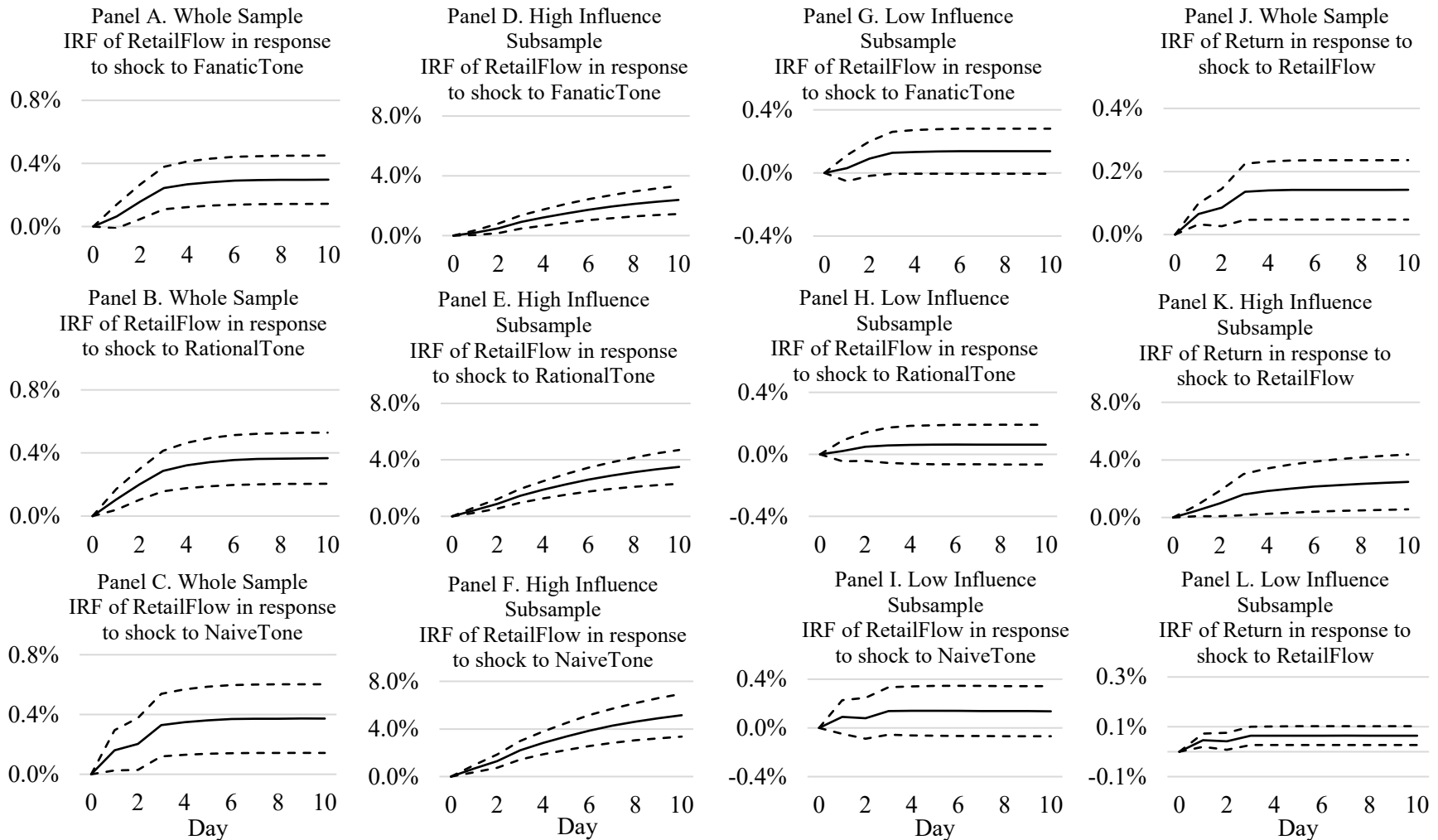
Appendix B Estimation Results Using PVAR with One Lag

This table presents results using PVAR with one lag. Our sample period is Jan 2, 2020 to Feb 15, 2021, and our sample firms are common stocks listed on NYSE, NYSE MKT, or Nasdaq. Fanatic agents are hard-headed agents with stable views that are not related to firm fundamental values. Rational agents are hard-headed agents with stable views that are related to firm fundamental values. Naïve agents have fluid views and are not hard-headed agents. The tone of each investor 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 in specification (8) 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%). Bold numbers denote the Granger-causal relations (p-value < 0.05).

	I	II	III	IV	V	VI
	FanaticTone(t)	RationalTone(t)	NaïveTone(t)	Return(t)	RetailFlow(t)	ShortFlow(t)
FanaticTone(t-1)	0.0738*** [11.75]	0.0133*** [5.18]	0.0260*** [6.04]	0.0106** [2.36]	0.0221* [1.94]	0.0040 [0.53]
RationalTone(t-1)	0.0157*** [2.63]	0.1656*** [18.69]	0.0502*** [8.10]	0.0025 [0.38]	0.0563*** [3.74]	-0.0046 [-0.42]
NaïveTone(t-1)	0.0168*** [6.01]	0.0080*** [4.70]	0.0199*** [5.50]	0.0071* [1.80]	0.0263* [1.87]	0.0237** [2.49]
Return(t-1)	0.0061*** [4.08]	0.0049*** [4.90]	0.0125*** [6.06]	0.0037 [0.33]	0.0476*** [7.08]	0.0616*** [7.11]
RetailFlow(t-1)	0.0003* [1.83]	0.0002* [1.95]	0.0007** [2.45]	0.0021*** [4.02]	0.0367*** [10.70]	-0.0003 [-0.28]
ShortFlow(t-1)	-0.0015** [-2.47]	-0.0007* [-1.91]	-0.0012 [-1.47]	-0.0039*** [-2.62]	-0.0300*** [-4.71]	0.3714*** [58.77]
Number of observations	269455	269455	269455	269455	269455	269455

Appendix C Cumulative Impulse Responses for Agents' Tones and Retail Flows

The figure reports the cumulative 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.



Appendix D Cumulative Impulse Responses for Agents' Tones and Shorting Flows

The figure reports the cumulative 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.

