

---

# The Rise of Reddit:

## How Social Media Affects Belief Formation and Price Discovery

---



Danqi Hu

Charles M. Jones

Siyang Li (Presenter)

Valerie Zhang

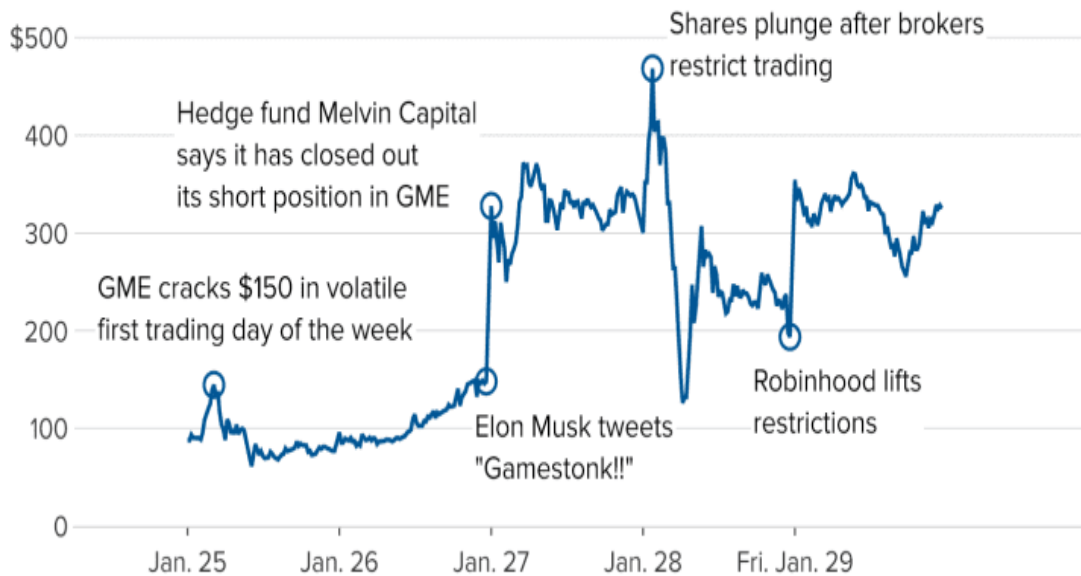
Xiaoyan Zhang

*ABFER 11<sup>th</sup> Annual Conference*

# The Story of GameStop and Reddit

## GameStop's wild week

The stock took off on a trading frenzy fueled by a Reddit message board

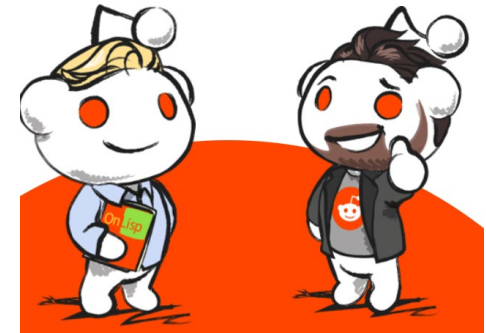


SOURCE: FactSet. Data as of market close on Jan. 29.



Steve Huffman  
Co-Founder/CEO  
u/spez

Alexis Ohanian  
Co-Founder  
u/kn0thing



---

# What We Do in This Study

- The big questions
    - Does social media significantly affect belief formation?
    - Does social media significantly affect price discovery?
    - How do different market participants react to this new dynamics from social media?
  - We collect information from the social media Reddit, as well as the capital market, and examine the predictions from Pedersen (2022) on our research questions.
    - Yes, social media significantly affects belief formation.
    - Yes, social media significantly affects price discovery.
    - Some investors choose to ride the bubble, while others choose to burst the bubble, all depending on the balance of benefits and costs.
-

---

# Previous Literature

- Social media: Antweiler and Frank (2004), Das and Chen (2007), Diangson and Jung (2021), Long, Lucey, and Yarovaya (2021), Bradley, Hanousek, James and Xiao (2023), Lyocsa, Baumohl and Vyrost (2022), Strych and Reschke (2022), and many others.
  - Retail investors: Kelley and Tetlock (2013), Boehmer, Jones, Zhang and Zhang (2021), Welch (2021), Eaton et al. (2022), Barber et al (2022) and many others.
  - Short sellers: Diamond and Verrecchia (1987), Asquith, Pathak, and Ritter (2005), Boehmer, Jones, and Zhang (2008), Engelberg, Reed, and Ringgenberg (2012), Allen, Nowak, Pirovano, and Tengulov (2022), Fusari, Jarrow and Lamichhane (2022) etc.
-

---

# Our Contributions

- 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 examines concrete theoretical predictions.
  - Our study provides unique insights and timely answers to important questions, which are helpful for all market participants.
-

---

# Pedersen's Model

- Fundamentals and timelines:
    - Firm has a fundamental value of  $v$ .
    - Each person receives a random signal of  $v$  with noise at time 0.
    - All signals collectively reflect the true value of  $v$ .
  - Investors learn from the social network:
    - Rational investors: learn from everybody, have information of  $v$  after time 0, and don't change opinions afterwards.
    - Fanatic investors: they only learn from self and don't change their opinions.
    - Naïve investors: they learn from a subset of others investors, and constantly update their views.
  - Model predictions
    - Belief formation, price discovery and trading dynamics
-

---

# Data

- Reddit data from the subreddit r/wallstreetbets
  - All submissions from Jan. 2, 2020, to Feb. 15, 2021.
    - 9.76 million comments and submissions
  - We assign each submission or comment to one or multiple firms
- Return data from Bloomberg
- Shorting flow data from CBOE
- Retail order imbalance data from TAQ

---

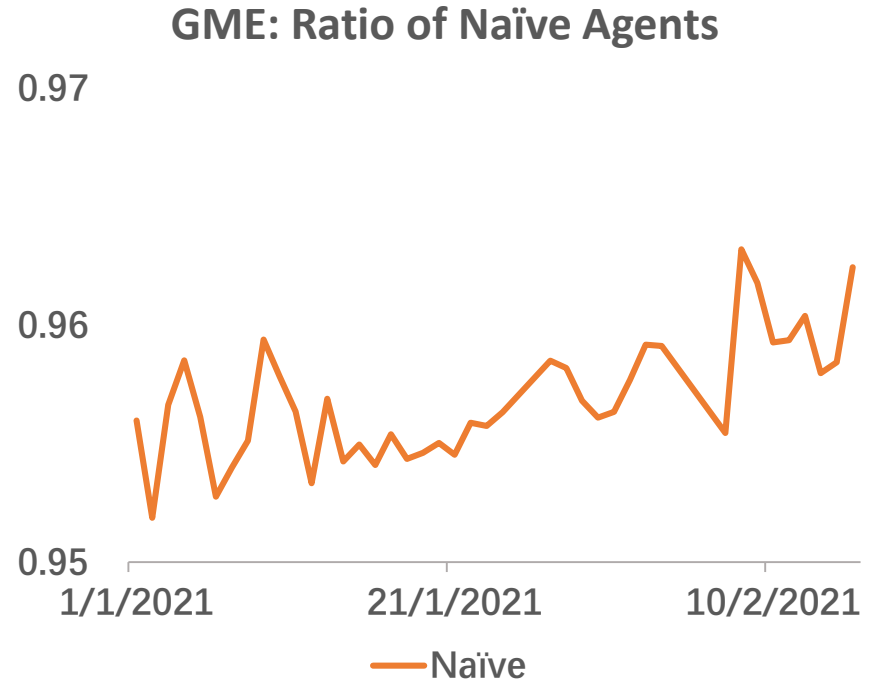
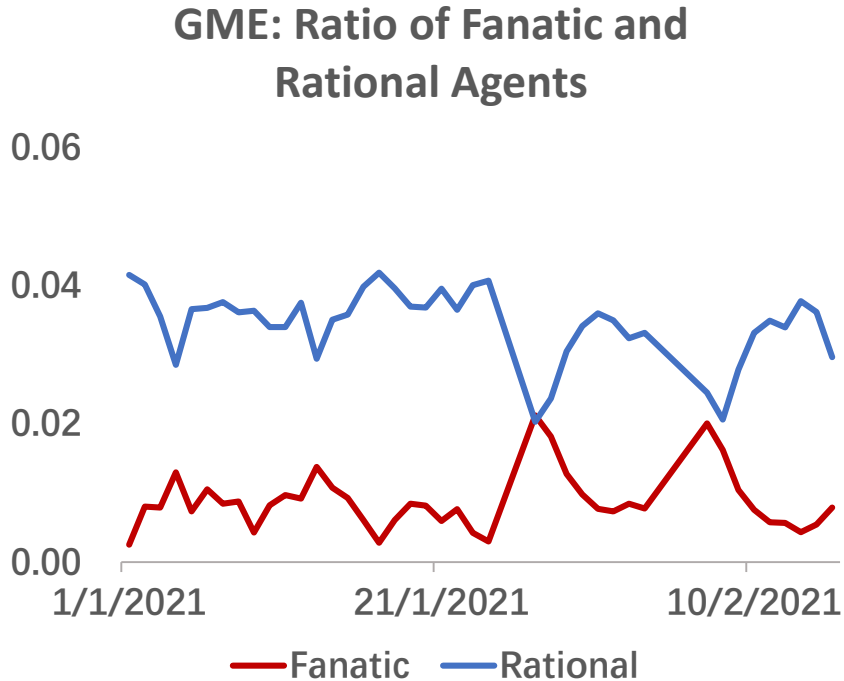
# Different Types of Agents

- Hardheaded agents on Reddit: users who express more opinions (so they can be heard) and whose tone remains stable over time.
    - agent  $k$  posts more submissions and comments about firm  $i$  than 95% of all other agents over the past 5-day window
    - posts have stable tones, with either non-positive or non-negative tones for at least 75% of their posts during the past 5 days
  - Rational: hardheaded agents who write at least one post that includes a value-relevant word during the 5-day window.
  - Fanatic: hardheaded agents whose posts contain no value-relevant words are identified as fanatics.
  - Naive: Reddit users who are not hardheaded are classified as naïve investors in our sample.
-



# Distribution of Agents

|                | mean  | std   | median |
|----------------|-------|-------|--------|
| Fanatic agent  | 0.079 | 0.146 | 0.000  |
| Rational agent | 0.044 | 0.108 | 0.000  |
| Naïve agent    | 0.877 | 0.161 | 0.947  |



---

# Social Media Measures

- Beliefs/opinions: tone of the submission and comments, + or -
    - Use word count method with Loughran and McDonald (LM) dictionary, add Reddit lingo (e.g., emojis, slang, jokes, and special meaning words).
  - Influence: number of direct commenters for each agent type's submissions
    - For instance, if one agent has more direct commenters, then she attracts more people's attention and have higher influence.
    - We identify firms with high-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.
-

# Empirical Method

- Panel vector autoregressions (PVAR) can capture the joint dynamics of belief formation, price discovery and trading behaviors in the social network

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

$$y_{i,t} = (FanaticTone_{i,t}, RationalTone_{i,t}, NaiveTone_{i,t}, Return_{i,t}, RetailFlow_{i,t}, ShortFlow_{i,t})'$$

- Matrix  $A_l$  is a coefficient matrix for lag  $l$ , and  $l = 1, \dots, L$ , is lag length. We choose the optimal lag length  $L=3$  using the Bayesian information criterion (BIC).
- Impulse response functions (IRF) describe how a variable responds to a one-time shock in another variable within the system.

---

# Hypothesis I: Belief Formation

- Prop. 1: Network belief spillover and convergence
    - In the limit of iterated communications, every naïve agent's view is a convex combination of views of fanatics and rational agents.
    - **Empirical hypothesis H1:** Naïve investors' views can be predicted by views from fanatic and rational investors.
  
  - Prop. 2: Influencers and thought leaders:
    - An agent has more influences on the belief formation when they have greater thought leadership or influencer value.
    - **Empirical hypothesis H2:** The more influences fanatic and rational agents have, the greater impact they have on the views of the naive agents.
-

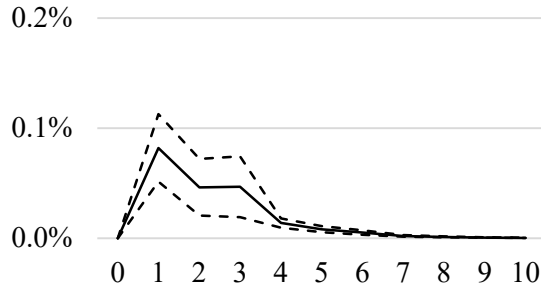
# Empirical Results on Belief Formation

- We only present the coefficients for the first lag for simplicity

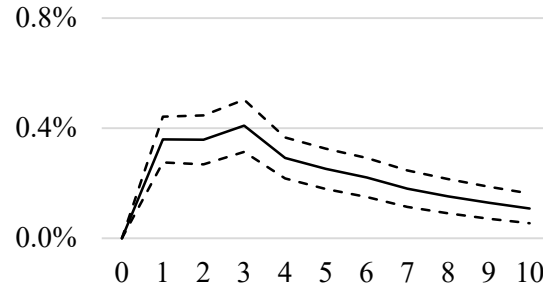
|                                   | I. Whole sample     | II. High influence subsample | III. Low influence subsample |
|-----------------------------------|---------------------|------------------------------|------------------------------|
|                                   | NaïveTone(t)        | NaïveTone(t)                 | NaïveTone(t)                 |
| FanaticTone(t-1)                  | 0.0241***<br>[5.43] | 0.0491***<br>[8.39]          | 0.0176***<br>[2.60]          |
| RationalTone(t-1)                 | 0.0394***<br>[6.63] | 0.0830***<br>[10.08]         | 0.0278***<br>[2.81]          |
| NaïveTone(t-1)                    | 0.0198***<br>[5.42] | 0.1855***<br>[13.01]         | 0.0028<br>[0.71]             |
| Number of observations            | 245002              | 25554                        | 219448                       |
| p-value of Granger causality test | Naïve<br>Tone(t)    | Naïve<br>Tone(t)             | Naïve<br>Tone(t)             |
| Past FanaticTone                  | 0.0%                | 0.0%                         | 3.4%                         |
| Past RationalTone                 | 0.0%                | 0.0%                         | 0.0%                         |

# Empirical Results on Belief Formation

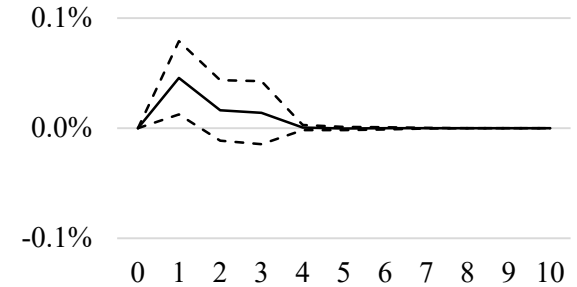
Panel A. Whole Sample  
IRF of NaiveTone in response to  
shock to FanaticTone



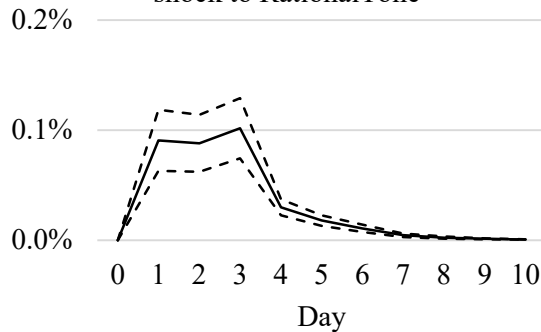
Panel C. High Influence Subsample  
IRF of NaiveTone in response to  
shock to FanaticTone



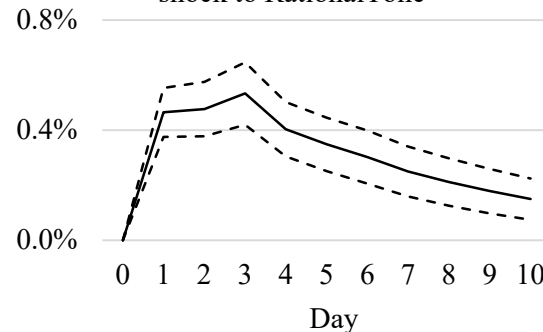
Panel E. Low Influence Subsample  
IRF of NaiveTone in response to  
shock to FanaticTone



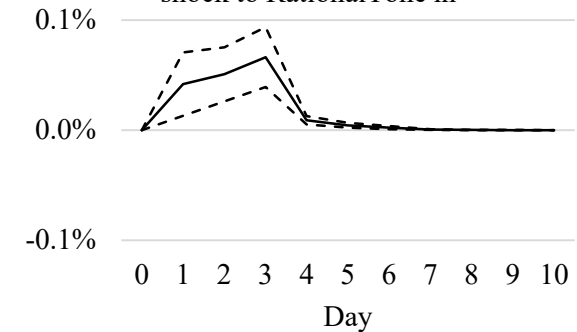
Panel B. Whole Sample  
IRF of NaiveTone in response to  
shock to RationalTone



Panel D. High Influence Subsample  
IRF of NaiveTone in response to  
shock to RationalTone



Panel F. Low Influence Subsample  
IRF of NaiveTone in response to  
shock to RationalTone in



---

# Hypothesis II: Price Discovery

- Prop. 4: Network effects on price
    - The equilibrium price for the social network is the sum of rational price plus a network price
      - Network dynamics, and agents' views affect equilibrium price
    - **Empirical hypothesis H3:** Agent views from social media network predict next day stock returns.
  
  - Prop. 5 and 6: Influencer effect on price
    - Each agent view's impact on price depends on their thought leadership or influencer value.
    - **Empirical hypothesis H4:** Agents with higher influences have larger impacts on stock returns.
-

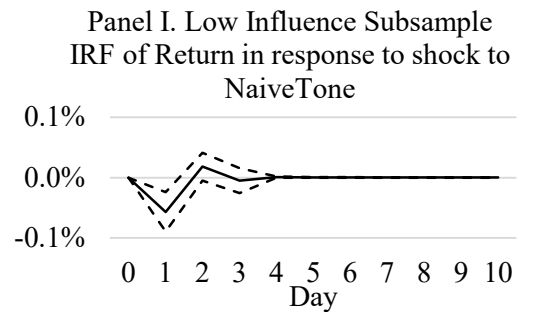
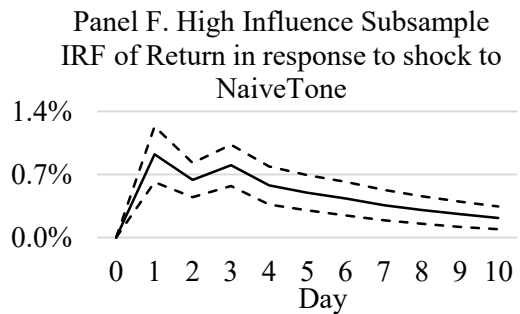
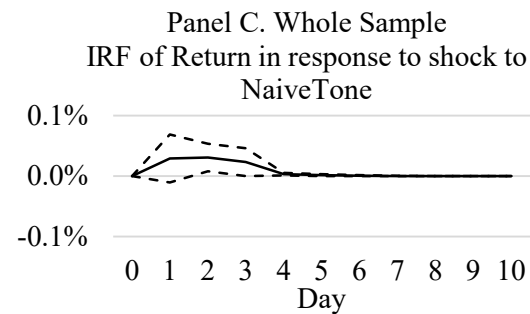
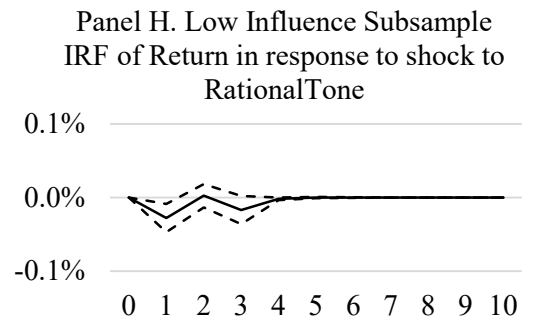
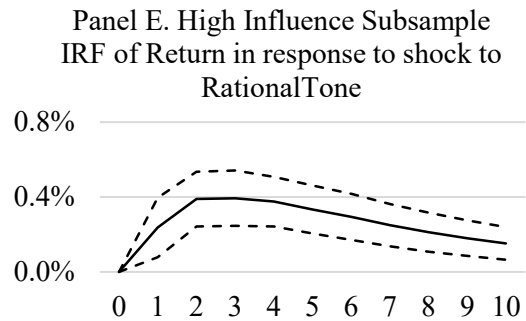
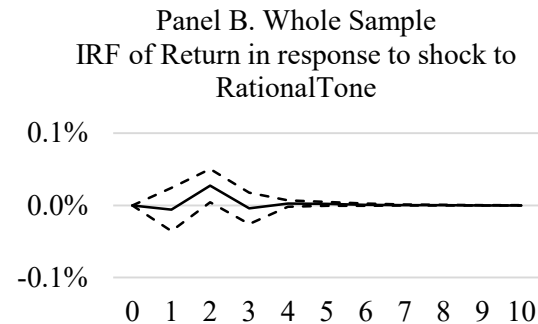
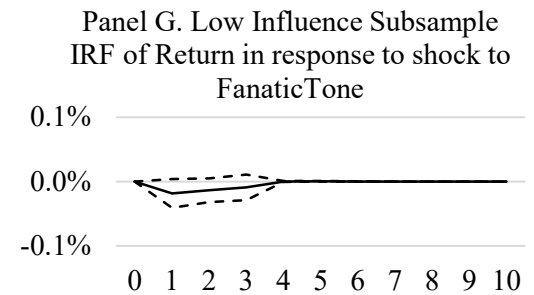
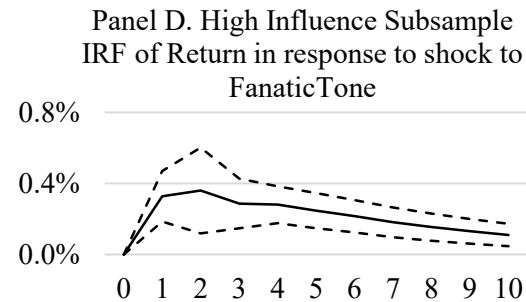
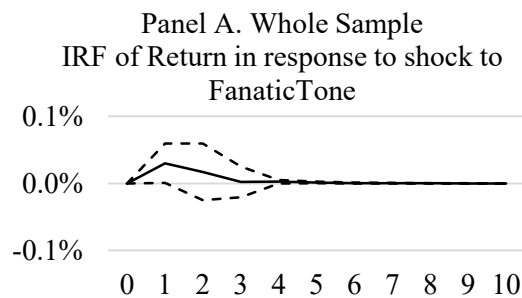
# Empirical Results on Price Discovery

- We only present the coefficients for the first lag for simplicity

|                                   | I. Whole sample    | II. High influence subsample | III. Low influence subsample |
|-----------------------------------|--------------------|------------------------------|------------------------------|
|                                   | Return(t)          | Return(t)                    | Return(t)                    |
| FanaticTone(t-1)                  | 0.0089**<br>[1.99] | 0.0449***<br>[4.44]          | -0.0071<br>[-1.58]           |
| RationalTone(t-1)                 | -0.0025<br>[-0.38] | 0.0426***<br>[2.87]          | -0.0186***<br>[-2.89]        |
| NaïveTone(t-1)                    | 0.0060<br>[1.43]   | 0.1775***<br>[6.02]          | -0.0121***<br>[-3.43]        |
| 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%                         |



# Empirical Results on Price Discovery



---

# Hypothesis III: Trading Dynamics

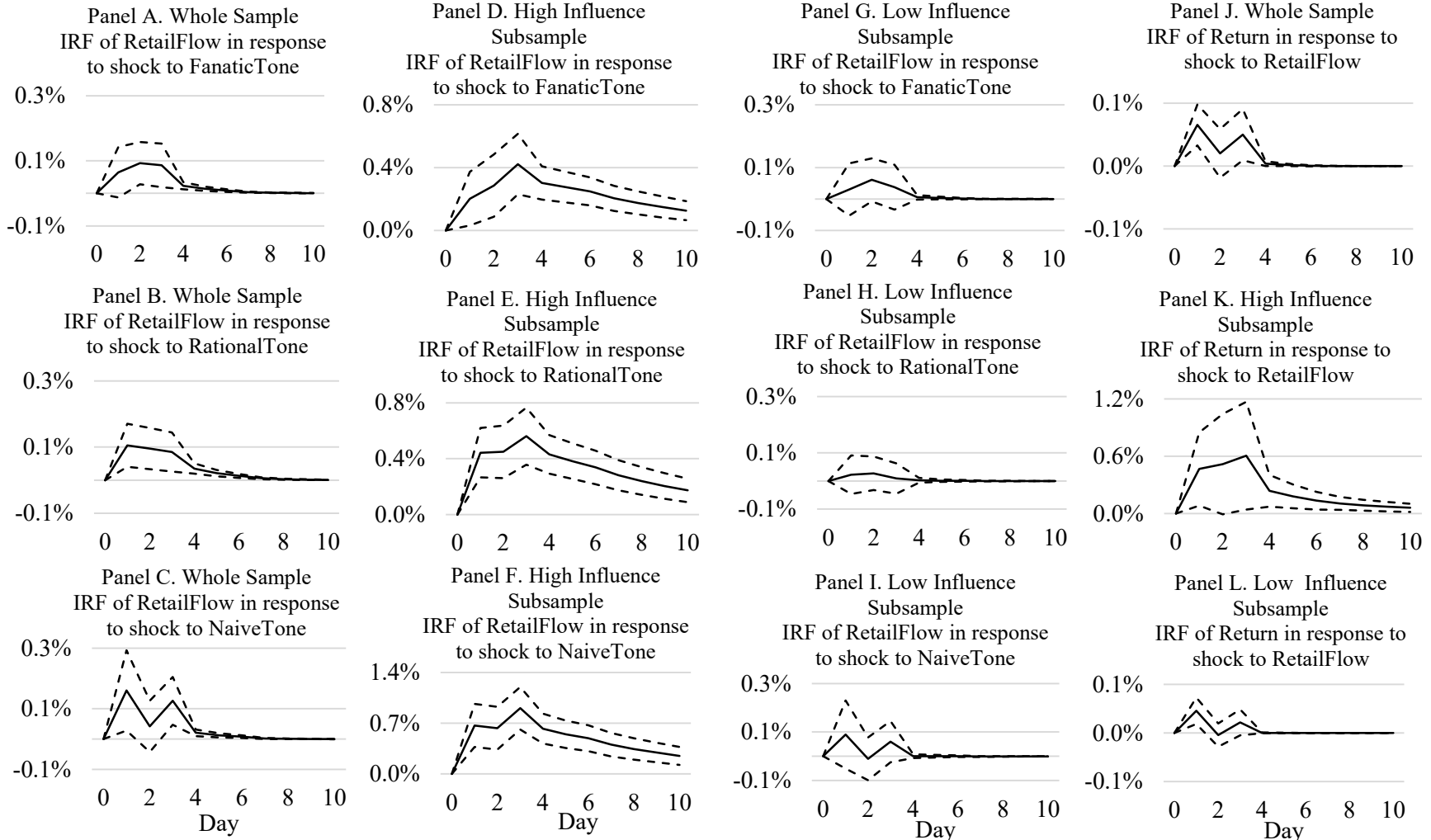
- Prop. 7 and 8: Value and momentum trading
    - If positive bubble, short-term investors initially buy a rising undervalued asset, continue to hold when the asset becomes over-valued (momentum, riding bubble), and finally shorts when the over-valuation is large enough (value on the short side, bursting bubble).
  - **Empirical Hypotheses:**
    - 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.
-

# Empirical Results on Retail Flows

- We only present the coefficients for the first lag for simplicity

|                                   | I. Whole sample     | II. High influence subsample | III. Low influence subsample |
|-----------------------------------|---------------------|------------------------------|------------------------------|
|                                   | RetailFlow(t)       | RetailFlow(t)                | RetailFlow(t)                |
| FanaticTone(t-1)                  | 0.0191*<br>[1.67]   | 0.0276**<br>[2.28]           | 0.0114<br>[0.70]             |
| RationalTone(t-1)                 | 0.0458***<br>[3.11] | 0.0790***<br>[4.56]          | 0.0148<br>[0.64]             |
| NaïveTone(t-1)                    | 0.0328**<br>[2.32]  | 0.1284***<br>[4.40]          | 0.0191<br>[1.27]             |
| 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%                        |

# Empirical Results on Retail Flows

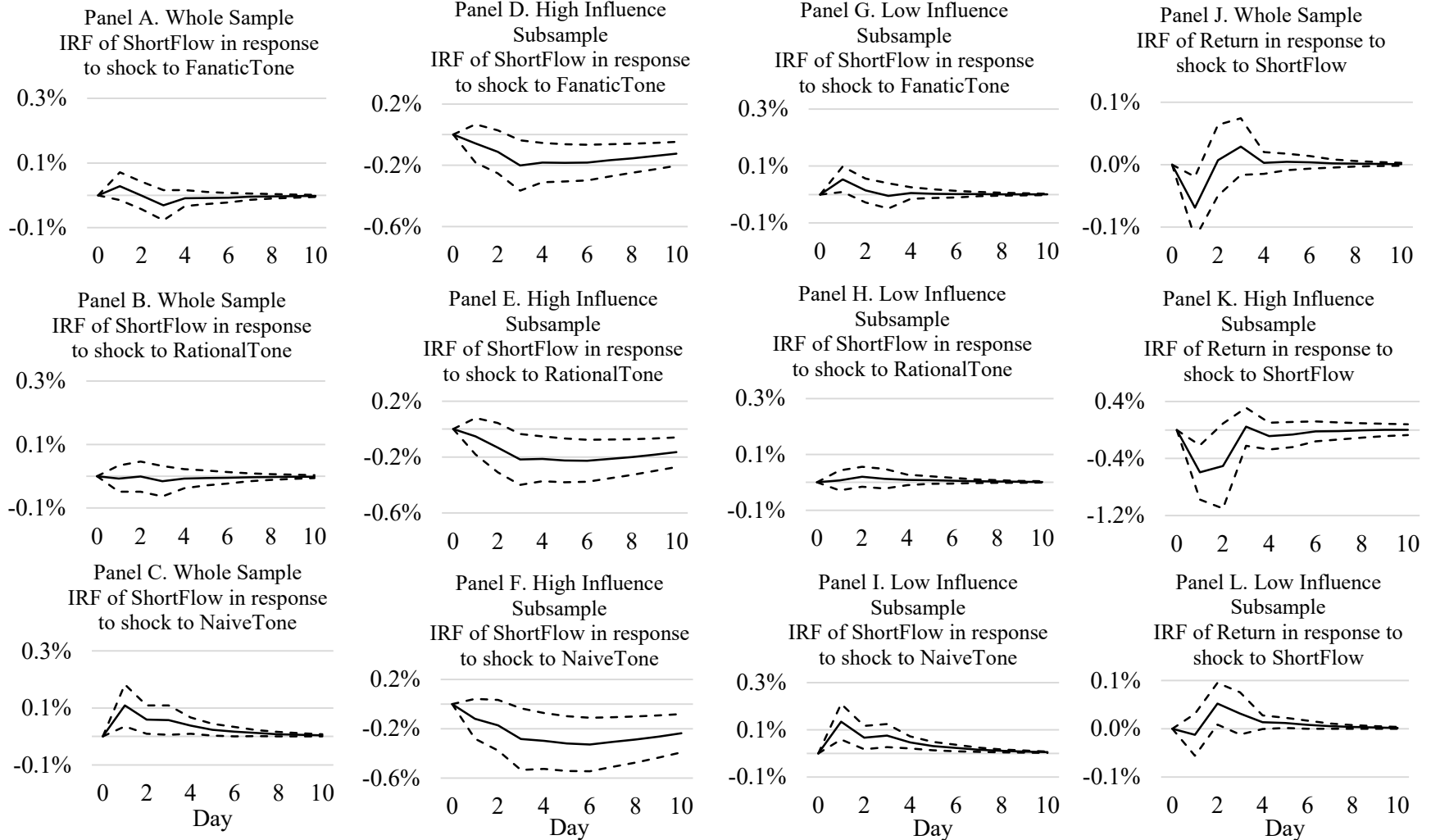


# Empirical Results on Shorting Flows

- We only present the coefficients for the first lag for simplicity

|                                   | I. Whole sample     | II. High influence subsample | III. Low influence subsample |
|-----------------------------------|---------------------|------------------------------|------------------------------|
|                                   | ShortFlow(t)        | ShortFlow(t)                 | ShortFlow(t)                 |
| FanaticTone(t-1)                  | 0.0084<br>[1.29]    | -0.0079<br>[-0.93]           | 0.0205**<br>[2.42]           |
| RationalTone(t-1)                 | -0.0033<br>[-0.36]  | -0.0091<br>[-0.79]           | 0.0044<br>[0.37]             |
| NaïveTone(t-1)                    | 0.0223***<br>[2.84] | -0.023<br>[-1.52]            | 0.0287***<br>[3.48]          |
| 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%                         |

# Empirical Results on Shorting Flows



---

# Conclusion

- Using 2020-2021 data from social media platform Reddit, we examine theoretical predictions from Pedersen (2022), and our findings generally support Pedersen's predictions.
    - Opinions of hard-headed investors (fanatic or rational) significantly predict future opinions of naïve investors, especially when these investors have larger influences.
    - Social media tones positively and significantly predict future returns, and more so when agents' influences are higher.
    - 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.
-