The Social Signal*

J. Anthony Cookson — *CU Boulder*
Runjing Lu — *Alberta*
William Mullins — *UC San Diego*
Marina Niessner — *Wharton*

October 8, 2022

Abstract

We examine social media attention and sentiment from three major platforms: Twitter, StockTwits, and Seeking Alpha. We find that attention is highly correlated across platforms, but sentiment is not: its first principal component explains little more variation than purely idiosyncratic sentiment. We attribute differences across platforms to differences in users (e.g., professionals vs. novices) and differences in platform design (e.g., character limits in posts). We also find that sentiment and attention are both positively related to retail trading imbalance, but contain different return-relevant information. Sentiment-induced retail trading imbalance predicts positive next-day returns, in contrast to attention-induced retail trading imbalance, which predicts strongly negative next-day returns. These results highlight the importance of distinguishing between social media sentiment and attention, and suggest caution when studying the social signal through the lens of a single platform.

Keywords: Social media, Retail trading, Social finance

JEL Codes: G14, G41, G12

*J. Anthony Cookson: University of Colorado at Boulder (tony.cookson@colorado.edu); Runjing Lu: University of Alberta (runjing1@ualberta.ca); William Mullins: UC San Diego (wmullins@ucsd.edu); Marina Niessner: University of Pennsylvania (marina.niessner@gmail.com). This draft has benefited from comments by Joey Engelberg and Brian Waters, as well as from seminar presentations at University of Colorado at Boulder and University of Texas—El Paso.

1. Introduction

Social media has grown exponentially over the past two decades. Americans spent 3.6 hours per day on some form of social media in 2020 (Forbes, 2021), and increasingly view social media as a primary source of news (Pew, 2021). Financial markets also reflect this: investors frequently post opinions about securities on social media, and firms use it to disclose information and interact with investors (Blankespoor et al., 2014). Despite these trends, investor social media was largely seen as a sideshow until recent social media-fueled trading frenzies, most prominently the 2021 GameStop phenomenon. These events raise questions about what role social platforms play for trading and information in financial markets (Pedersen, 2022), and an emerging line of research has organized around these important questions.

Analyses of investor social media have almost exclusively examined data from a single platform, and related papers often draw upon evidence from *different* investor social networks, typically StockTwits, Seeking Alpha, and Twitter.[1] While most of this work considers questions that are not specific to the particular investor social platform studied, these platforms differ in myriad ways.[2] Communication theory implies that social media platforms are unlikely to be interchangeable, because the characteristics of a communication medium affects both the content and impact of the message it carries (e.g., McLuhan, 1975). Differences between platforms – in user populations, incentives to post, and ability to engage – may lead to important differences in the information each platform attracts and aggregates.

To examine whether and how these platforms generate differing market-relevant information – the social media *signal* – this paper examines comprehensive firm-day level data (2012−2021) from the three most established investor social networks: StockTwits, Twitter, and Seeking Alpha. We first distinguish between social media *attention* (i.e., number of messages about a firm) and *sentiment* (i.e., bullish vs. bearish views about a firm), and

[1]For StockTwits see, for example, Giannini et al. (2019), Cookson et al. (2022a,b), Irvine et al. (2021). Similarly, for Twitter (Gu and Kurov, 2020, Chen et al., 2019), Seeking Alpha (Chen et al., 2014, Dim, 2020, Chen and Hwang, 2022, Farrell et al., 2022), and Reddit (Bradley et al., 2021).

[2]For example, Seeking Alpha articles are long-form and are lightly moderated; Twitter posts are limited to 160 characters but multiple posts can be threaded together for longer arguments; while StockTwits posts cannot be threaded, but since 2019 have a 1,000 character limit. These platforms also differ in their user base, recommendation algorithms, how individuals interact on the platform through messages and tagging, and many other characteristics. Figure 1 presents an example post for each platform in our study.

compare these across platforms. We find that over two-thirds of the firm-day attention signal is common across the major social platforms: on a given day, people on different platforms tend to talk about the same firms. By contrast, the common component of sentiment is weak, explaining only slightly more than it would if sentiment across the three platforms were orthogonal. Exploiting two events that could affect the social media information signal – a change in message character limits on StockTwits and the GameStop (GME) short squeeze – we find that differences across platforms (the character-limit experiment) *and* differences across users (the GME experiment) each contribute to differences in sentiment across investor platforms.

Turning to market outcomes, we see clear differences in the informativeness of retail trade that follows sentiment when compared to the retail trade following attention. Specifically, sentiment-informed trade predicts *positive* returns that quickly revert to baseline, whereas attention-informed trade predicts large *negative* returns. Our results suggest that it is important to draw a sharp distinction between questions about social investor sentiment and those about attention.

We now describe our findings in greater detail. We begin by decomposing the social signal generated by all three platforms using separate principal component analyses for attention and sentiment. The first principal component (PC) explains 70% of the variation in attention, but the first PC of sentiment only explains 38.8% of its variation across platforms. A similar PC analysis of investor subgroups on StockTwits shows that attention is similar for influencers, professionals, and novices, in contrast to the sentiment signals from these investor subgroups. We also observe that attention and sentiment from traditional financial news are only weakly correlated with social media attention and sentiment. These findings make clear that the information in investor social media differs from that in traditional media, and that differences in the social signal across platforms are likely linked to the users who inhabit each platform.

Next, we examine the connection between the social media signal and the *direction* of retail trading. Irrespective of social platform, sentiment exhibits a positive and significant relation to retail trading imbalance – that is, the tilt towards *buy* orders relative to sell orders from retail investors. Moreover, social media sentiment exhibits a much stronger relation to the direction of retail trading than does news sentiment. Specifically, a standard deviation increase in the first PC of social media sentiment is associated with 0.781% greater retail trading imbalance, but only 0.137% for news sentiment. Social media attention and sentiment also interact: when both are high, retail's tilt towards buying is even stronger.

Our results on the direction of retail trading also reveal substantial differences across social platforms. Though sentiment from all platforms is positively associated with retail buying, the coefficient estimate from the StockTwits signal is 3 to 4 times stronger than that from the Twitter or Seeking Alpha signal. However, including all signals in the same regression shows that Twitter and Seeking Alpha sentiment each have unique variation regarding the direction of retail trading: no single platform provides a complete picture.

After highlighting the connection between the direction of retail trading and the social signal, we next explore whether the retail trading that follows the social signal predicts abnormal returns. To do this, we first estimate an auxiliary regression in which we regress retail trading imbalance (RT imbalance) on sentiment and attention from the three social platforms. From this, we construct the RT imbalance predicted by all sentiment terms (we call this "RT imbalance from sentiment"), the RT imbalance predicted by attention terms ("RT imbalance from attention"), and the residual RT imbalance which is unrelated to the social signal by construction.[3] Consistent with Boehmer et al. (2021), we find that RT imbalance is positively associated with future returns. However, novel to our setting, we find that RT imbalance from social media sentiment predicts next-day returns twice as well as overall RT imbalance.

By contrast, retail trading from top-tercile attention strongly predicts *negative* next-day returns. Relative to bottom-tercile attention, top-tercile attention predicts a 21.7 basis points lower next-day return. These negative returns accumulate to an 87.4 basis point loss over the next 20 days, with the negative cumulative returns leveling off before 20 days. The specification controls for ten lags of sentiment and attention, indicators for 8-K and earnings announcement dates, and the amount and sentiment of news. Thus, the fact that RT imbalance on high-attention days strongly predicts negative returns is most naturally

[3] We also perform a similar exercise for interaction terms between sentiment and attention in this auxiliary regression.

explained by this component of retail trade being misinformed.

In the final part of our paper, we exploit two information experiments to provide a deeper insight into these patterns. First, we examine changes in the informativeness of the social signal across platforms around May 8, 2019, when StockTwits increased its character limit per message from 140 to 1,000 characters. We find that StockTwits sentiment became more predictive of next-day stock returns after this change. Moreover, this effect is driven by sentiment extracted from longer messages; the informativeness of shorter messages and attention were unchanged. Consistent with the fact that Twitter and Seeking Alpha were unaffected by StockTwits' format change, we find no change in the informativeness of signals from these platforms. These results indicate that a within-platform change to users' ability to communicate can affect the market-relevant information encoded in the social media signal. These findings also suggest that structural differences across investor social media platforms contribute to the differences we find in the social signal they generate.

Second, we examine how the informativeness of sentiment and attention changed around the January 2021 GME phenomenon. In 2020 the number of U.S. retail brokerage accounts increased rapidly and StockTwits saw an influx of new users, both likely a result of COVID stay-at-home orders coupled with the introduction of no-fee trading at many brokerages. Bradley et al. (2021) shows that Reddit retail trading based on "due diligence" reports became less informative in the wake of the GME short squeeze. We show that the informativeness of sentiment across all platforms deteriorated significantly after the GME short squeeze: both returns and retail trade imbalance became less sensitive to sentiment. Moreover, the drop in informativeness is concentrated among messages by new entrants, as the informativeness of the signal extracted from more established users (who joined before 2020) did not change after January 2021. These results suggest that the information environment became much noisier for retail traders after the GME short squeeze. Correspondingly, social media sentiment becomes a worse predictor of next-day returns, and retail investors are less likely to trade in the direction of social sentiment.

Related literature. Our paper makes several contributions to the literature on retail investors, sentiment, attention, and the informativeness of novel data sources in financial markets. Our core contribution is to quantify the information content, similarities, and differences across the three most-established investor social media platforms over the last decade. With the rising significance of social media platforms as a forum for communicating investor beliefs, a literature has emerged to study these platforms' information content. Investors discuss investment ideas on a plethora of forums, but analyses typically focus on a single platform, and employ different data (e.g., Chen et al., 2014, Cookson and Niessner, 2020, Gu and Kurov, 2020, Irvine et al., 2021).[4] Divergent findings may stem from examining different parts of the investor social media space. In this paper, we show that this concern is particularly important for sentiment and for the informativeness of retail trade, while cross-platform differences are less important for attention.

Our results also contribute to the literature on retail attention and sentiment (e.g., Sicherman et al., 2016, Gargano and Rossi, 2018). Existing work with investor social media either focuses on aspects of investor attention (e.g., Giannini et al., 2018, Cookson et al., 2022a, Irvine et al., 2021) or on sentiment and optimism (e.g., Renault, 2017, Cookson et al., 2020). Outside of social media, research on sentiment (e.g., Tetlock, 2007, Garcia, 2013) and attention (e.g., Barber and Odean, 2008, Da et al., 2011) has also typically focused on only one of the two. As a result, a seemingly conflicting body of evidence has emerged in which sentiment is typically informative of future returns, but retail attention appears strongly misinformed. The literature has partly resolved this tension by showing that different kinds of attention have different return implications (Ben-Rephael et al., 2017, Da et al., 2022, Barber et al., 2022). By examining them together across multiple platforms, we show that there is a striking difference in the informativeness of sentiment vs. attention.[5]

We also contribute to the literature on the role of new entrants to financial markets and their implications for markets (e.g., Bradley et al., 2021). In early work, Chen et al. (2014) shows that Seeking Alpha recommendations are informative. With the advent of new firm-day retail trading measures (Boehmer et al., 2021), the literature has examined how

[4]Recent work on earnings forecasts from Estimize has examined similar questions about information transmission and social influence – e.g., Da and Huang (2020) and Jame et al. (2016) study aspects of the wisdom of crowds, and Da et al. (2021) shows how Estimize analysts extrapolate their beliefs from past experience.

[5]Our results provide support for Barber et al. (2021), which shows that retail trades predict positive returns, but market weighted, generate losses for retail traders. While their evidence stems from portfolio sorts of stocks that are more or less heavily traded by retail traders, our decomposition shows that retail trading following attention drives negative returns, but this is not the case for sentiment-driven trade.

retail trading relates to social media activity, with a primary focus on Seeking Alpha (e.g., Farrell et al., 2022). This research has also shown that retail investor activity has important implications for market quality, particularly driven by new retail traders on Robinhood (e.g., Eaton et al., 2022). Relative to this literature, our results connect social media, retail trading, and market outcomes, and we show that new entrants lead to much of the decline in informativeness following the GME phenomenon.[6] This finding highlights how the content of previously informative signals can change upon the arrival of new participants, and this is a general phenomenon that is not just confined to one social network. More generally, we illustrate how features of different social media platforms (e.g., character limits and different user-bases) matter for retail trading informativeness.

2. Data and summary statistics

2.1 Social media sentiment and attention data

Our data come from three investor social media platforms: Twitter, Seeking Alpha, and StockTwits. We obtain firm-day data on financially-oriented Twitter posts (tweets) from Social Market Analytics (SMA), a firm that provides sentiment information to professional investors. Specifically, we use a daily 4pm snapshot of the number of tweets and average sentiment over the prior 24-hour period for each firm.

For Seeking Alpha we obtain article level sentiment from Ravenpack 1.0, keeping all articles with a relevance score above 75, which Ravenpack considers to be "significantly relevant." To measure sentiment we use the Event Sentiment Score (ESS) calculated by Ravenpack, which ranges between -1 and 1, with 0 indicating neutral sentiment, positive (negative) values indicating positive (negative) sentiment.

For the investor social platform StockTwits we obtain comprehensive message level data. Like Twitter, StockTwits allows users to publicly post short messages (henceforth "tweets") with a limited number of characters – 140 before May 8, 2019, and 1,000 thereafter. Unlike

[6] In connecting social media and retail trading, we also relate to the literature that studies motivations for retail trade (Liu et al., 2020). Prior work shows how peer interactions lead to excessive trading and exacerbate behavioral biases (Heimer, 2016), and emphasizes the role of overconfidence (Barber and Odean, 2001, Daniel and Hirshleifer, 2015), particularly about the precision of one's information (e.g., Daniel et al., 1998).

Twitter, StockTwits is primarily focused on financial markets. By including a "cashtag", a dollar sign ($) followed by a ticker symbol, StockTwits users can specify that their post refers to a specific firm or security. We limit our analysis to messages that mention exactly one company, to be able to accurately assign sentiment to the company. We have data on all single-firm tweets from 2010 through 2021: 150 million tweets from over 800,000 users. Similar to Cookson et al. (2022a), we drop users posting over 1,000 tweets in a day, and we restrict our sample to the top 1,500 firms by the number of tweets posted between 2010 and 2021.

StockTwits allows users to attach a sentiment tag to their tweet indicating if their tweet reflects "bullish" or "bearish" sentiment. We assign +1 to self-labeled "bullish" tweets and -1 to self-labeled "bearish" tweets. We also obtain a sentiment score for each tweet ranging from -1 (extremely bearish) to +1 (extremely bullish) which is calculated by StockTwits using a proprietary text classification algorithm called MarketLex.[7]

To aggregate sentiment at the firm-day level ($Sentiment_{i,t}$) for StockTwits and Seeking Alpha, we compute average sentiment across all tweets (or articles) about a firm i from 4:00 pm (close) on date $t-1$ to 4:00 pm on date t. These firm-day sentiment measures are thus comparable to the Twitter firm-day sentiment measure provided by SMA. Similarly, we compute firm-day message volume ($Messages_{i,t}$) for StockTwits and Seeking Alpha by counting the number of messages (tweets or articles) about each firm between 4:00 pm on date $t-1$ and 4:00 pm on date t, in order to match the timing of our Twitter tweet volume data. We then define a firm-day measure of attention, $Attention_{i,t}$, for each platform by dividing the firm-day number of messages by the total number of messages in a day:

$$\text{Attention}_{i,t} = \frac{Messages_{i,t}}{\sum_i Messages_{i,t}} \tag{1}$$

StockTwits users can voluntarily declare their level of experience using the options provided by StockTwits when filling out their user profile. StockTwits also provides information on how many followers each user has. Thus, for StockTwits, we can separate sentiment and

[7]According to StockTwits, this methodology uses lexical and semantic rules based on a custom-built lexicon for social finance, constructed from a combination of words and phrases from 4 million messages with user-provided bullish or bearish tags and manual human supervision.

attention into distinct series by user profile or follower base: Professionals, Intermediates, Novices, No experience classification, and Influencers ($>$ 99th percentile in followers). We also produce a separate series for self-classified sentiment (explicit bullish/bearish declarations), as opposed to StockTwits' own sentiment measure based on MarketLex.

2.2 Firm news data

In addition to social media sentiment and attention, we also control for firm news events. Specifically, we collect information on coverage and sentiment of traditional news media from the *Wall Street Journal* and the *Dow Jones Newswire*. These measures come from Ravenpack 1.0, which provides information on the number of articles by firm-day as well as article-level sentiment. We keep all articles with a relevance score above 75 and use the Ravenpack Event Sentiment Score, aggregating the article-level sentiment by averaging firm-specific sentiment across articles each day.

To capture other sources of news we collect information on 8-K filing dates (unscheduled firm-specific news) and earnings announcement dates. The 8-K filing dates are collected from the SEC Analytics Suite by WRDS, and the earnings announcement dates are from the I/B/E/S data set on WRDS.

2.3 Retail trading and returns data

We identify retail trades using the methodology of Boehmer et al. (2021), which exploits two institutional features. First, most equity trades by retail investors take place off-exchange, either filled from the broker's own inventory or sold by the broker to wholesalers, which TAQ classifies with exchange code "D." Second, retail traders usually get a small fraction of a cent price improvement over the National Best Bid or Offer, while institutional orders tend to be executed at whole or half-cent increments. Therefore, we focus on trades executed on exchange code "D," and identify trades as retail purchases (sales) if the trade took place at a price just below (above) a round penny. We aggregate retail trades at the daily level. We follow Boehmer et al. (2021) and calculate retail order imbalance as the difference between retail buying and selling volume, divided by the total retail trading volume:

$$\text{RT Imbalance}_{i,t} = \frac{\text{Number of Buys}_{i,t} - \text{Number of Sells}_{i,t}}{\text{Number of Buys}_{i,t} + \text{Number of Sells}_{i,t}} \tag{2}$$

This proxy for retail trade imbalance is a signed measure of the direction of retail trading activity. A positive (negative) value means that there are more (less) retail purchases than retail sales. In addition, using CRSP data we compute daily abnormal returns by subtracting the value-weighted market return from the firm's daily return.

2.4 Sample Characteristics

To allow accurate measurement of the social signal, our sample focuses on the 1,500 firms with the most single-firm tweets about them on StockTwits between 2010 and 2021. Although this reduces the number of firms in our sample from more than 9,000 to 1,500, it only reduces the number of StockTwits messages by about 20% (from 150 million to 120 million). We also restrict attention to firm-days for which there are at least 10 single-firm tweets on StockTwits. Because Twitter and Seeking Alpha data are sparsely populated before 2012, we begin our analysis sample in 2012. After merging the social media data with Ravenpack for news media information, market data for return reactions, and TAQ data for the retail trading measures, we obtain a final sample of roughly 815,000 firm-day observations.

Summary statistics of our analysis sample are reported in Table 1. Panel A presents statistics on activity across the three platforms. For the average firm-day, the number of messages on StockTwits is a multiple of the number on Twitter or Seeking Alpha. Despite this substantial difference in message volume, the three platforms cover a similar number of firms (i.e., StockTwits mentions cover 1,497 firms in our final sample compared with 1,271 and 1,283 for Twitter and Seeking Alpha, respectively). Thus, even if individual messages on StockTwits contain less information than a Seeking Alpha post, the substantial volume of messages could contribute to an informative firm-day signal. In Panel B, we present the same statistics for subgroups of StockTwits investors. This decomposition highlights that there is significant activity among each subgroup (the average number of posts ranges from

5.87 to 13.37 across the categories with experience). Moreover, there is a significant volume of posts with self-labeled sentiment: the average firm-day observation in our data has 69.3 self-classified bullish or bearish posts.

Panel C shows how restricting our sample to firm-days with at least 10 StockTwits messages affects the observation count in our final sample. Our original firm-day sample falls from nearly 2.8 million to roughly 821,000 observations. Additional sample filters – e.g., requiring data on controls, returns, or retail trade imbalance classification from TAQ – have negligible impact on our observation count.

2.5 Platform features and users

Communication theory (e.g., McLuhan, 1975) holds that the characteristics of a communication medium affect both the content and impact of its message. Thus, differences between platforms – in user populations, incentives to post, and ability to engage – may lead to important differences in the information each platform attracts and aggregates. Figure 1 presents three messages about Apple Stock ($AAPL), one from each investor social media platform, in order to illustrate cross-platform differences.

The most immediate difference in Figure 1 is that Seeking Alpha content consists of long-form articles (the screenshot displays only the title and summary), in contrast to the short posts on StockTwits and Twitter. There are many other platform feature differences. For example, although StockTwits and Twitter both allow "cashtags," only on StockTwits can posts be flagged as bullish or bearish by the poster. Moreover, StockTwits is an investment-specific platform, while Twitter covers an unrestricted variety of topics. Other differences include the recommendation algorithms and the ability to thread tweets. Each of these can contribute to important discrepancies across platforms in both the social signal (sentiment and attention measures) as well as how the social signal relates to market outcomes. We exploit a change in one of these differences – when StockTwits increased their message character limit from 140 to 1,000 characters – to examine how platform features impact the information content on the platform.

Another major difference across social media platforms is that they attract different users. Seeking Alpha posters are much more selected than Twitter or StockTwits users, which can be anyone who signs up for an account. StockTwits has historically attracted users aiming to build reputation via their posts: deletion of past posts is not possible. Moreover, interest in these platforms has shifted and grown over time, as is clear from Appendix Figure A1. To explore the importance of user composition, we test whether newer StockTwits users provide a less informative signal around a notable market event: the GME short squeeze of early 2021. We also use our within-StockTwits decomposition of different user types to examine how each type contributes to the social signal.

3. Decomposing the social signal

This section describes the commonalities and differences in the signals drawn from social media investing platforms and across StockTwits investor types. In this discussion we distinguish between two key dimensions of the social signal: sentiment and attention.

3.1 Are social signals common across platforms?

We begin by examining how much overlap there is in the social signal across StockTwits, Twitter, and Seeking Alpha. First, Figure 2 and panel A of Table 2 present the bivariate correlations between StockTwits attention and sentiment, and the corresponding measures from Twitter and Seeking Alpha at the firm-day level. As a benchmark, we also present the correlations of StockTwits with traditional news coverage and sentiment from the Dow Jones Newswire (DJNW) and the Wall Street Journal (WSJ). The correlation between attention series on social platforms is relatively high at 0.595 between StockTwits and Twitter attention, and 0.398 between StockTwits and Seeking Alpha. By comparison, the correlations with coverage by traditional news media are much weaker: 0.163 for the WSJ and 0.144 for the DJNW. These correlations indicate that attention across social investing platforms contains a strong common component that is not well explained by news media coverage.

In contrast to the attention correlations, we observe much weaker correlations in sentiment series across different platforms. The correlation of StockTwits with Twitter sentiment is only 0.125, whereas the correlation of StockTwits with Seeking Alpha sentiment is 0.038. The correlation with news sentiment is also weak at 0.010 for the WSJ and 0.069 for the

DJNW. This suggests that sentiment is more idiosyncratic across social investing platforms, and as Figure 2 highlights, the difference in the magnitudes of correlations for attention and sentiment is striking.

In appendix Table A1, we regress StockTwits attention and sentiment on the matching signals from the other platforms and from news media. For attention we find that both Twitter and Seeking Alpha attention load strongly and positively, while traditional news media loads weakly and negatively. Similarly, for sentiment we observe that both Twitter and Seeking Alpha sentiment have explanatory power for StockTwits sentiment beyond each other, and in addition to the sentiment from traditional media. However, the explanatory power of cross-platform sentiment and news is substantially lower for sentiment than for attention. These results suggest that while there is a significant common component to the social media signals for both attention and sentiment, attention is more common across social platforms than is sentiment.[8]

Next, we describe the common variation between social media signals in principal component analyses (PCAs) for attention, and separately for sentiment. These PCAs are summarized in panels B and C of Table 2. Consistent with the view that most attention is common across investors on various social media platforms, the first principal component (PC) of attention explains 70% of the variation across platforms. Further, all three attention signals are given similar positive weights in this first PC, suggesting a natural interpretation as the common component of attention manifested in all three social media platforms. The second PC captures differences in attention across Seeking Alpha and StockTwits since it places positive weight on Seeking Alpha and negative weight on StockTwits (with roughly zero weight on Twitter). However, these differences in attention across platforms captured by the second PC only explain 18.9% of the variation in attention.

Interpreting the sentiment PCA in panel C, the first PC of sentiment only explains 38.8% of the variation across platforms. This is a weak common component, because purely idiosyncratic variation in three series would result in a first PC explaining 33.3%. Like the

[8]Table A1 also regresses StockTwits attention and sentiment signals from *subsets* of the user population on the main signals from the other two platforms. While displaying similar patterns across all subsets, Twitter and Seeking Alpha are better at explaining the sentiment of professionals (and the top 1% most followed users) than they are at explaining the sentiment of other subgroups.

attention PCA, the second PC of sentiment mostly highlights the difference between Seeking Alpha (positive weight, $w = 0.874$) and StockTwits (negative weight, $w = -0.464$) since the Twitter sentiment series has a much smaller weight ($w = -0.147$). The fact that the second PC explains 32.3% of the variation implies that differences across platforms in sentiment capture approximately as much variation across platforms as similarities.

3.2 Similarities and differences in the social signal across user types

In addition to allowing us to construct attention and sentiment signals at the firm-day level, the StockTwits data allow us to track these signals separately for investor subgroups. In this section, we disaggregate the StockTwits signal to separately consider the sentiment and attention of influencers (those in the top 1% by number of followers), professional users, intermediate users, novice users, and users who do not indicate an experience category (no experience). We also construct a separate sentiment and attention series by only focusing on posts that users themselves marked as being bullish or bearish on the stock (self-labeled).

Panel A of Table 3 presents the correlation of overall StockTwits attention and sentiment with the same measures for each subset of the data at the firm-day level. The correlations of attention across user groups on StockTwits range from 0.819 (influencers) to 0.987 (no-experience users). In contrast to these high attention correlations, the correlation of sentiment of one user group with another is typically quite weak. Except for the 0.783 correlation between no-experience users (which make up 88% of users) and overall StockTwits sentiment, the remaining subgroups exhibit correlations ranging from 0.232 (novices) to 0.387 (self-labeled posts). These relatively weak correlations in sentiment across user subgroups suggest that an important source of differences in social media sentiment is due to idiosyncratic differences across user *types*.

In panels B and C, we extend the PCA from the cross-platform table to include each of these alternative signals from StockTwits subgroups, in addition to Twitter and Seeking Alpha. For brevity, these panels only report the first four PCs. Statistics on the full PCAs are reported in appendix Table A2.

Consistent with our cross-platform evidence, we see that attention contains a strong common component (76.6% of the variation) while sentiment's common component is weaker at only 26.6%. Referring to Panel B, the second PC of attention captures differences in attention between StockTwits and the other platforms but does not explain much variation in attention (10.6%). In the attention PCA, components 3 onward explain very little variation in attention, while capturing idiosyncratic combinations of the underlying signals.

In the sentiment PCA in Panel C, the other three sentiment PCs explain a nontrivial fraction of variation. These sentiment PCs capture differences across platforms together with differences in sophistication. The second PC is a mixture of Seeking Alpha and Twitter sentiment, together with the more sophisticated parts of StockTwits (influencers and professionals), minus the much more numerous remaining groups. The third PC contrasts professionals and influencers with the other signals, while the fourth PC is mostly made up of the signal from novices. These differences suggest that the idiosyncratic nature of sentiment across groups is primarily driven by the underlying differences between investors' evaluations of the market.

3.3 Are social signals persistent?

We now examine the persistence of attention and sentiment over time by computing the partial autocorrelation function (PACF) for each platform's attention and sentiment signal. In Figure 3, we compute the PACF for each series out to 20 lags. Attention (dashed lines) tends to have high autocorrelations (around 0.8 at lag 1) that decay to near zero after lag 5. By contrast, sentiment has low autocorrelations (between 0.1 and 0.25) and decays more rapidly to zero.

This pattern constitutes another difference between attention and sentiment signals: attention exhibits a much greater and more persistent autocorrelation than does sentiment. We account for these underlying differences when we relate attention and sentiment to trading and returns.

4. The social signal and the direction of retail trading

In this section, we examine how the social signals (attention, sentiment, and their interaction) relate to the direction of retail trading. In particular, we find that the retail trading imbalance (RT Imbalance) – i.e., a tilt towards *buy* orders from retail investors – is positively related to attention, sentiment, and their interaction. We further show that this connection is unexplained by traditional news.

4.1 Retail trading imbalance

We estimate the following specification:

$$\begin{aligned} \textit{RT Imbalance}_{i,t} &= \beta_1 \textit{Attention}_{i,t} + \beta_2 \textit{Sentiment}_{i,t} + \beta_3 \textit{Sentiment}_{i,t} \times \textit{Attention}_{i,t} \\ &+ \beta_4 \textit{Coverage}_{DJNW,i,t} + \beta_5 \textit{RT Imbalance}_{i,t-1} + \mathbf{X_{i,t}} \times \mathbf{\Gamma} + \alpha_\mathbf{t} + \epsilon_\mathbf{it} \end{aligned} \tag{3}$$

where the dependent variable $\textit{RT Imbalance}_{i,t}$ is retail order imbalance from Boehmer et al. (2021) scaled by 100. $\textit{Attention}_{i,t}$ and $\textit{Sentiment}_{i,t}$ are firm-day measures from one of the platforms or the principal components constructed in the previous section. In addition, the controls ($X_{i,t}$) include the DJNW sentiment and attention, whether the day is an 8-K filing date or earnings announcement date, lagged RT imbalance (from $t-1$ to $t-5$ and from $t-6$ to $t-30$), lagged volatility ($t-1$ to $t-5$), lagged market returns (CAR $t-1$ to $t-5$ and $t-6$ to $t-30$), and a date fixed effect. These controls account for persistence in retail trade direction, news, and recent volatility and momentum effects, which comprise many alternative drivers of retail trading. We also control for lagged $\textit{Attention}_{i,\tau}$ and $\textit{Sentiment}_{i,\tau}$ (where $\tau = t-1$, $t-2$, ..., $t-10$) to account for the autocorrelation documented in the preceding section. Table 4 presents the results from estimating Equation (3), employing varying combinations of the social signal variables across table columns.

First, we find that the social signal variables from StockTwits, Twitter, and Seeking Alpha each exhibit a clear connection to the direction of retail trading at the firm-day level. Across platforms, more positive sentiment is associated with a more positive retail trade imbalance. A natural possibility is that users make investments consistent with their statements on these platforms. However, it is also possible that these statements instead proxy for the sentiment and attention of retail investors as a group, which in turn drives the trading activity.

Second, though each platform's sentiment and attention are strongly linked to retail trade imbalance, there are important differences across the signals. For example, the magnitude of the coefficients estimated from StockTwits sentiment is 3 to 4 times greater than the estimates from the other platforms (and explains additional variation – see the improvement to within-R^2 and RMSE). Additionally, when using the StockTwits signal, the interaction between sentiment and attention matters for retail trading, but this interaction is not significant for Twitter or Seeking Alpha. Third, when we include all signals in the same regression in column 5, we see that the connection to retail trade remains robust, holding constant the other signals.

In addition to these main specifications, which force the retail order imbalance relation to be linear in sentiment and attention, we allow them to be nonlinear by splitting attention and sentiment into terciles of their first PCs and interacting them in an analogous specification to Equation (3). Consistent with the linear specification, we observe a monotonic relation between sentiment and retail trade within each attention tercile (see panel A of Figure 4). As attention increases, retail trading imbalance increases for a given sentiment level. Thus the connection between sentiment, attention, and retail order imbalance does not depend on the particular specification we employ.

4.2 Informativeness of retail trades that follow the social signal

Next, we examine whether trades that follow the social signal predict returns. We begin with the observation that a regression of RT imbalance onto sentiment and attention *decomposes* overall RT imbalance into RT imbalance correlated with attention, RT imbalance correlated with sentiment, and residual RT imbalance that is orthogonal to these signals. We then examine how the components of RT imbalance following sentiment, attention, and the residual predict returns.

First, column 1 of Table 5 provides a baseline by presenting how next-day cumulative abnormal returns (CAR) are predicted by using RT imbalance, controlling for lagged RT imbalance, volatility and market returns, as well as and firm and date fixed effects (plus other controls). All else equal, a standard deviation increase in RT imbalance (i.e., a retail tilt towards buying) is associated with an increase of 3.5 bps in next-day abnormal return, consistent with the findings in (Boehmer et al., 2021).

We then decompose RT imbalance into (i) the components predicted by sentiment, attention, and/or their interactions, and (ii) a residual, which captures the RT imbalance informed by neither sentiment nor attention.[9] This decomposition results from an auxiliary regression of retail trading imbalance on all sentiment signals, terciles of the first principal component of attention, and/or the attention-sentiment interactions. The specification follows column 5 of Table 4 but also includes within-StockTwits signals (such as attention and sentiment from influencers, professionals, etc). We call the residual from this regression *residual imbalance*, and the component of the fitted regression that depends on social media sentiment terms *imbalance from sentiment*, and we extract similar terms for the terciles of attention and for sentiment-attention interactions.

Column 2 of Table 5 shows how well RT imbalance from sentiment and residual imbalance predict next day CAR (note that the auxiliary regression includes only sentiment signals, sentiment signals lagged 1-10 days, and all other controls listed in Table 5). RT imbalance from sentiment is associated with double the next-day return of residual RT imbalance. Specifically, a standard deviation increase in RT imbalance from sentiment is associated with a 6.2 bps greater next-day return vs. 3.4 bps for residual RT imbalance.

In column 3, we perform a similar analysis with terciles of attention, omitting the lowest tercile. In this case, greater RT imbalance from attention is associated with increasingly negative next-day returns. Relative to the baseline, a firm with attention in the middle tercile can expect a 9.1 bps lower next-day returns, and a firm in the top attention tercile can expect 20.8 bps lower next-day returns.

To examine the role of RT imbalance from the interaction of sentiment and attention, in column 4 we include sentiment and attention terms together (as well as in the auxiliary regression), and in column 5 we additionally include the interactions between sentiment and attention terms (as well as in the auxiliary regression). Including both sentiment and attention at the same time does not noticeably change the coefficients or their standard errors, and their interactions appear to have little predictive power. In short, the RT imbalance from sentiment and from attention appear to be separate, non-intersecting effects.

[9]As a complement to this main analysis, in Appendix Tables A4 and A5 we perform an alternative analysis that shows how the components of the social signal relate to returns directly.

4.3 Dynamics of return effects

Next, we consider the predictive power of the RT imbalance that follows the social signal for returns further into the future. Table 6 presents the results for returns from $t+1$ to $t+20$. RT imbalance informed by social media sentiment predicts persistently positive returns for 20-day CARs. However, it is not much more informative than residual RT imbalance. For instance, a standard deviation increase in RT imbalance from sentiment is associated with a 10.1 bps higher return over the 20-day window, vs. the 8.6 bps higher return from a similar increase in residual RT imbalance. To understand how quickly the informativeness of retail trade following sentiment reverts to baseline, panel (a) of Figure 5 presents the estimated sentiment-informed and residual RT imbalance coefficients from a series of regressions using CARs over varying windows as the outcome. The informativeness of RT imbalance from sentiment significantly exceeds the informativeness of the residual RT imbalance on day 1. On on subsequent days, it remains slightly positive, but its informativeness for returns is no greater than that of residual RT imbalance, at around 10 or 11 bp.

In contrast to our findings for sentiment, the RT imbalance from attention terms do not quickly revert to the residual RT imbalance baseline. In fact, from column 4, a standard deviation increase in RT imbalance for a stock with middle tercile attention can expect to earn 31 bps less than one in the bottom tercile by day 20. For a top tercile stock this difference is an *87 bps* lower return. Panel (b) of Figure 5 presents the daily coefficients on the RT-imbalance predicted by mid and high terciles of attention, as well as the residual imbalance. Imbalance from attention's informativeness for future returns accumulates until flattening out around day 15. The figure suggests that the impact of attention – via RT imbalance – on returns seems to be mostly impounded into prices by day 20.[10]

[10] In Appendix Figure A4, we show our results are robust to changing the sampling frame. Namely, we show results when (i) we also require that there be at least 4 Twitter messages for that firm-day (the same firm-day percentile as 10 StockTwits messages), and/or (ii) focusing on a different set of stocks (i.e., the most-discussed stocks on StockTwits during 2018-2021 rather than 2010-2021). The results are similar, with a slightly more pronounced sentiment effect, and a slightly weaker negative reaction to social media attention.

5. Information experiments

In this section, we present the results from two market events that change the information impounded in the social signal, illustrating how the social signal differs across platforms and in response to market events.

5.1 StockTwits character limit change

On May 8, 2019, StockTwits changed the limits on its posts from 140 characters to 1,000 characters. We explore whether this change affected the informativeness of sentiment and attention originating from StockTwits in comparison to the informativeness in unaffected signals from Twitter and Seeking Alpha. To focus on the StockTwits format change, we analyze the period from one year before to one year after May 8, 2019.

Figure 6 shows how the distribution of characters per message (panel a) and firm-day level average number of characters (panel b) changed during this event window. Consistent with this format change only affecting the content of longer messages, we see that only messages in the top percentiles of characters per message are made longer by the format change. Similarly, the impact of the character limit expansion is also larger for the firm-day observations in higher percentiles of the number of characters.

To focus more cleanly on the impact of the StockTwits character limit increase, we present a set of platform-by-platform regressions of the form:

$$\begin{aligned}\text{Outcome}_{i,t} = \beta_1 Attention_{i,t} + \beta_2 Sentiment_{i,t} + \beta_3 Sentiment_{i,t} \times Post_t \\ + \beta_4 Attention_{i,t} \times Post_t + \mathbf{X_{i,t}} \times \mathbf{\Gamma} + \alpha_{\mathbf{t}} + \alpha_{\mathbf{i}} + \epsilon_{\mathbf{it}}\end{aligned} \tag{4}$$

where $Outcome_{i,t}$ is either $RT\ Imbalance_{i,t}$ or next-day return $CAR_{i,t+1}$. This specification includes controls and fixed effects as in the baseline specifications for retail trading and next-day returns while also including firm fixed effects. Relative to Equation 3, the novel terms are sentiment and attention interacted with a $Post_t$ indicator for whether the day is after May 8, 2019. The coefficients of interest are these interactions with $Post_t$, which capture the change in the informativeness of the social signal around the character limit increase.

Table 7 presents the results from estimating this specification separately for StockTwits, Twitter, and Seeking Alpha. Consistent with sentiment becoming more informative after the character limit increase, we estimate that the coefficient on sentiment for next-day returns increases by 6.6 bp (column 1). Although this estimate is only statistically significant at the 10% level, its magnitude is nearly twice the main effect of sentiment (3 bp, row 3 of the table). We use the StockTwits sentiment signal extracted exclusively from posts in the top quartile by length in column 2, and find an even stronger increase in informativeness of 13.7 bp. By contrast, we see no change for Twitter or Seeking Alpha in columns 3 and 4, indicating that the change in informativeness is specific to StockTwits.

Turning to the connection to retail trading (columns 5 through 8), we see little change in the sensitivity of retail trading to sentiment after StockTwits character limit increase, irrespective of the social media platform.

To more formally estimate the impact of the character limit increase, we perform analysis akin to a difference-in-differences design in which we define "treated" observations as firms with average character length in the top quartile in a day, and "control" as those in the bottom quartile. Using this definition, we extend the specification in Equation 4 to one that also contains interactions with the $Treated$ indicator.

When we estimate such a specification, the triple interaction term $Post \times Treated \times Sentiment(z)$ is 0.186 when the dependent variable is next-day return (see Appendix Table A6). This estimate implies that, after the character limit increase, sentiment becomes more informative for next-day returns, especially for long messages affected by the increase. Specifically, a standard deviation increase in sentiment predicts an 18.6 bps greater return for long versus short messages on StockTwits. We do not see a significant change in sentiment's relationship with retail order imbalance.

Figure 7 presents the quarterly estimates of the triple interaction in leads and lags around the character limit increase. These plots indicate that the effect is not driven by any obvious trends in informativeness of sentiment over time.

If we replace sentiment and attention with the RT imbalance predicted by each of them – as in the preceding section, examining CAR – we similarly find that retail trading predicted by sentiment becomes more informative after the character limit increase, and particularly so for longer messages. Figure 8 panel (a) plots the estimates for the change in informativeness of sentiment before versus after the character limit increase between firms with longer versus shorter messages.

5.2 Changes around the GameStop short squeeze

In this section, we analyze a second market event that likely influenced the informativeness of social media signals: the GameStop Short Squeeze event (GME event) in late January 2021. Bradley et al. (2021) study a class of posts on Reddit's forum Wall Street Bets (WSB) called "due diligence reports" around this event, and find that these reports were informative for future returns before the event, but their informativeness fell markedly afterwards. We perform a similar analysis for the informativeness of sentiment and attention around the GME event using the first principal components of attention and sentiment constructed in Section 3.1.

We look at 11 months prior and post the GME event since we have data until the end of 2021 (only 11 months). We also exclude January 2021 to have a cleaner pre-post comparison. The specification is exactly analogous to Equation (4).

Table 8 presents the findings on the informativeness of the social signal for next-day returns and retail trading. We find that returns' sensitivity to sentiment drops substantially following the GME event, and the informativeness of retail trading imbalance for future returns declines. Specifically, a standard deviation in sentiment (the first PC) is associated with a 9.8 bps lower return after the GME event (column 1). This completely offsets the pre-GME informativeness of social media sentiment (9.5 bps). For retail trading imbalance (column 5), we estimate that a standard deviation change in sentiment leads to 0.267 percentage points lower retail order imbalance in column 3, a reduction of more than one third of the baseline level (0.766). In column 2 and 6, we additionally include the second and third PCs of sentiment (capturing cross-platform differences in sentiment) but the coefficients are essentially unchanged.

Appendix Table A8 performs a similar analysis for the informativeness of retail trading imbalance and its attention and sentiment components. Column 1 provides the baseline result that retail order imbalance does not, on average, change its informativeness after the GME event. However, once we decompose retail trading into its social signal components in column 2, we see that retail trading associated with social media *sentiment* becomes substantially less predictive of return after the GME short squeeze.

To better understand the mechanism behind this decline in informativeness, we use message-level data from StockTwits. StockTwits saw an influx of new users and increase in posts starting in 2020, likely the result of stay-at-home orders coupled with the introduction of no-fee trading at many brokerages in late 2019. Using when users joined the platform, we split the sample into tweets by those who joined prior to January 2020 (*established users*) and tweets by users who joined more recently (*new users*). From each subsample of tweets, we construct separate measures of attention and sentiment.

New users displayed a stronger interest in "short squeeze" strategies after the GME event. In Figure 9, we document a persistent uptick in mentions of short squeezes on StockTwits from an average of roughly 6,200 mentions per month in the year before the GME event to an average of nearly 13,000 afterwards. This spike in posts mentioning "short squeeze" is primarily driven by new users (who registered on StockTwits after January 1, 2020) with an increase from around 4,300 to over 17,000 posts per month; in contrast, short squeeze posts from established users only see a moderate uptick (from 8,040 to 8,180 per month).

Consistent with new users' stronger interest in "short squeeze," we find that the informativeness of the new users' signal declines significantly by 10.2 basis points for a standard deviation increase in sentiment after the GME event (Table 8 column 4), whereas the informativeness of the established users' signal does not change (column 3). In summary, this evidence shows how differences in user bases on an investor social media platform can have first order effects on the informativeness of the signal it generates. Indeed, during the post-GME period, nearly all the predictability of retail trading following sentiment is eroded (10.7 bps on the interaction relative to 10.9 bps as baseline). However, the decline in informativeness is due to new users.[11]

To explore the "short squeeze" mechanism further, we split the sample of stocks based on exposure to short squeeze strategies – either by mentions of short squeeze-related terms in

[11] As a complement to the sample split evidence in the main text, Appendix Tables A9 and A10 present evidence from specifications that contrast the informativeness of the social signal for new users vs. established users, pre vs. post the event. These tests reveal that the difference between the change in signal informativeness between new and established users is statistically significant.

their StockTwits posts, or by their short sale utilization, which we measure using daily short selling data from Markit. Specifically, we define stocks as highly exposed to the post-GME "short squeeze" phenomenon if their past month's mentions of "short squeeze" terms (or short selling utilization) are above-median (H) and little exposed if below-median (L).[12]

Based on these indicators for high vs. low short selling exposure, we estimate the $Post \times SentimentPC1(z)$ and $Post \times AttentionPC1(z)$ terms separately for above- vs. below-median subsamples. Table 9 presents these results. Consistent with the finding that the social signal became less informative, we find that the decline in the informativeness of sentiment is concentrated in stocks with more short-selling mentions and in heavily shorted stocks – with estimated coefficients on $Post \times SentimentPC1(z)$ of -0.154 and -0.225 respectively. Outside of these high short selling stocks, in columns 2 and 4 we see no significant impact of the GME event on the informativeness of the social signal. We see a similar pattern for retail trading imbalance, with the post-GME decline in the sensitivity to sentiment concentrated among the stocks that are most exposed to short selling.

Overall, these findings from the GME event highlight how new users and emergent retail investor coordination (i.e., retail "short squeeze" strategies) can influence the informativeness of the social media signal. This evidence specifically shows how different user bases contribute to the information in the social media signal.

6. Conclusion

In this paper, we explore the similarities and differences in the social signals generated from StockTwits, Twitter, and Seeking Alpha, as well as their information content for retail trading. Our analysis reveals differences across social investing platforms that are much more pronounced for sentiment than for attention. We attribute these differences to differences in types of investors (e.g., influencers, professionals, and novices) *and* differences across platform features (e.g., character limits on posts).

We also find striking differences in the informativeness of sentiment vs. attention. Retail trade imbalance following sentiment is relatively well-informed – in that it positively predicts

[12] Our "short squeeze" dictionary contains the words "squeeze," "short interest," "short seller," and "short volume," similar to Bradley et al. (2021)'s price pressure list

next-day returns – while retail trade following high attention predicts strongly negative returns. This finding bridges two largely disconnected research agendas on sentiment and attention and underlines an important economic difference between social media attention and sentiment.

Investor social media has increased steadily in popularity over the past two decades, and has grown rapidly in recent years. Online investment forums (like StockTwits, Twitter, and Seeking Alpha) attract hundreds of thousands of daily users who engage in intense debate about individual securities. Given the differences across platforms, particularly new entrants that rely on other kinds of media (e.g., Discord and TikTok), it is natural to expect the informativeness of future social signals and retail trading to evolve as well (Chang and Peng, 2021, Pyun, 2021). Will these new technologies enhance or weaken the information environment? We expect ample opportunities for future work to examine the consequences of these emerging technologies.

References

Barber, B. M., Huang, X., Odean, T., and Schwarz, C. (2022). Attention induced trading and returns: Evidence from robinhood users. *Journal of Finance, forthcoming.*

Barber, B. M., Lin, S., and Odean, T. (2021). Resolving a paradox: Retail trades positively predict returns but are not profitable. *Working paper.*

Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1):261–292.

Barber, B. M. and Odean, T. (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.

Ben-Rephael, A., Da, Z., and Israelsen, R. D. (2017). It depends on where you search: A comparison of institutional and retail attention. *Review of Financial Studies*, 30(9):3009–3047.

Blankespoor, E., Miller, G. S., and White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter. *The Accounting Review*, 89(1):79–112.

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

Bradley, D., Hanousek Jr, J., Jame, R., and Xiao, Z. (2021). Place your bets? the market consequences of investment research on reddit's wallstreetbets. *Working paper.*

Chang, R. and Peng, L. (2021). Tiktok sentiment, emotional contagion, and global stock returns. *Working paper.*

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27(5):1367–1403.

Chen, H. and Hwang, B.-H. (2022). Listening in on investors' thoughts and conversations. *Journal of Financial Economics*, 145(2):426–444.

Chen, H., Hwang, B.-H., and Liu, B. (2019). The emergence of'social executives' and its consequences for financial markets. *Working paper.*

Cookson, J. A., Engelberg, J., and Mullins, W. (2022a). Echo chambers. *Review of Financial Studies, Forthcoming.*

Cookson, J. A., Engelberg, J. E., and Mullins, W. (2020). Does partisanship shape investor beliefs? Evidence from the covid-19 pandemic. *Review of Asset Pricing Studies*, 10(4):863–893.

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

Cookson, J. A., Niessner, M., and Schiller, C. (2022b). Can social media inform corporate decisions? evidence from merger withdrawals. *Working paper*.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of finance*, 66(5):1461–1499.

Da, Z., Hua, J., Hung, C.-C., and Peng, L. (2022). Market returns and a tale of two types of attention. *Available at SSRN 3551662*.

Da, Z. and Huang, X. (2020). Harnessing the wisdom of crowds. *Management Science*, 66(5):1847–1867.

Da, Z., Huang, X., and Jin, L. J. (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics*, 140(1):175–196.

Daniel, K. and Hirshleifer, D. (2015). Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives*, 29(4):61–88.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *Journal of Finance*, 53(6):1839–1885.

Dim, C. (2020). Should retail investors listen to social media analysts? evidence from text-implied beliefs. *Working paper*.

Eaton, G. W., Green, T. C., Roseman, B. S., and Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, 146(2):502–528.

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

Forbes (2021). Americans spent on average more than 1,300 hours on social media last year. *Forbes*. https://www.forbes.com/sites/petersuciu/2021/06/24/americans-spent-more-than-1300-hours-on-social-media/?sh=352920522547 (accessed August 29, 2022).

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, 68(3):1267–1300.

Gargano, A. and Rossi, A. G. (2018). Does it pay to pay attention? *The Review of Financial Studies*, 31(12):4595–4649.

Giannini, R., Irvine, P., and Shu, T. (2018). Nonlocal disadvantage: An examination of social media sentiment. *The Review of Asset Pricing Studies*, 8(2):293–336.

Giannini, R., Irvine, P., and Shu, T. (2019). The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. *Journal of Financial Markets*, 42:94–120.

Gu, C. and Kurov, A. (2020). Informational role of social media: Evidence from twitter sentiment. *Journal of Banking & Finance*, 121:105969.

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

Irvine, P. J., Shen, S., and Shu, T. (2021). Aggregate attention. *Working paper.*

Jame, R., Johnston, R., Markov, S., and Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, 54(4):1077–1110.

Liu, H., Peng, C., Xiong, W. A., and Xiong, W. (2020). Resolving the excessive trading puzzle: An integrated approach based on surveys and transactions. *NBER working paper.*

McLuhan, M. (1975). *Understanding media: The extensions of man.* Routledge & Kegan Paul.

Pedersen, L. H. (2022). Game on: Social networks and markets. *Journal of Financial Economics.*

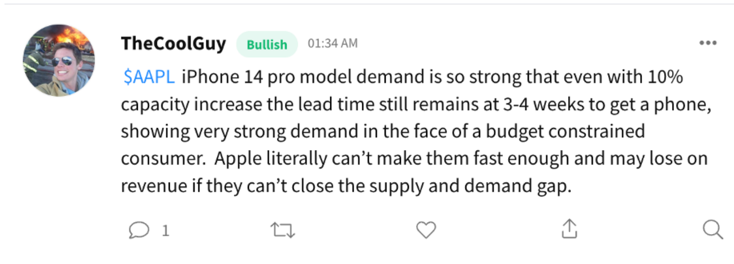
Pew (2021). News use across social media platforms in 2020. *Pew Research Center.*

Pyun, C. (2021). Social media group investing. *Working paper.*

Renault, T. (2017). Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance*, 84:25–40.

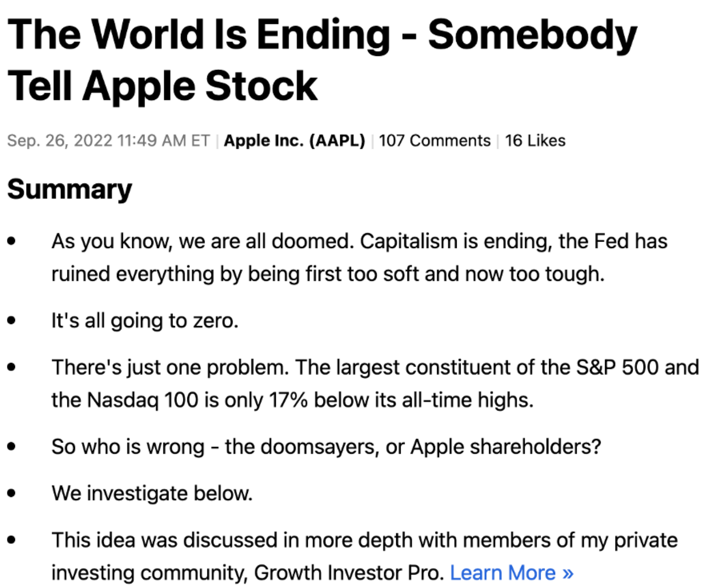
Sicherman, N., Loewenstein, G., Seppi, D. J., and Utkus, S. P. (2016). Financial attention. *The Review of Financial Studies*, 29(4):863–897.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3):1139–1168.



(a) StockTwits

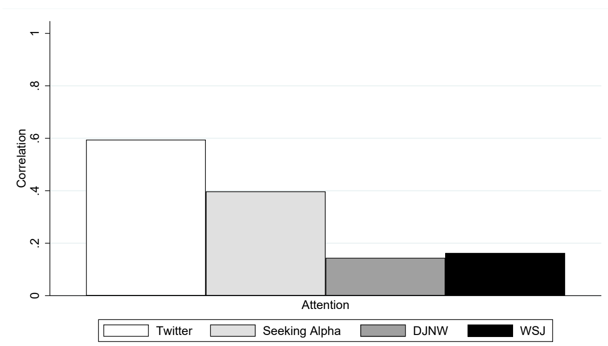
(b) Twitter



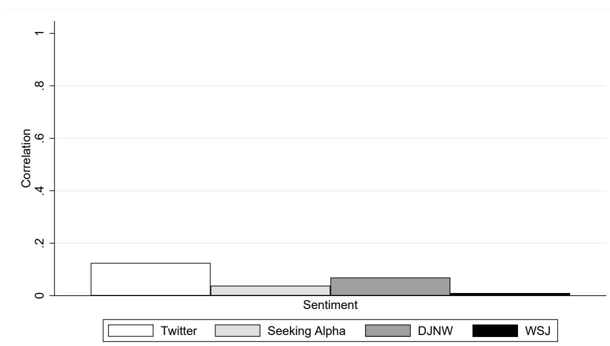
(c) Seeking Alpha

Figure 1: Examples of Posts Across Three Social Media Platforms

Note: This figure presents example posts and tweets separately by social media platform for StockTwits, Twitter, and Seeking Alpha. For ease of direct comparison, all three example posts are about Apple stock (AAPL) on the same day (September 28, 2022).



(a) Attention



(b) Sentiment

Figure 2: Cross-platform Correlation for Social Signals

Note: This figure reports the bivariate correlations of attention and sentiment between StockTwits and another platform at the firm-day level. Attention is measured by the fraction of messages, reports, or articles about a firm across all firms on a platform in a day. Sentiment is measured by the average sentiment of all messages about a firm on a platform in a day. Sample consists of firm-day observations with at least 10 messages on StockTwits.

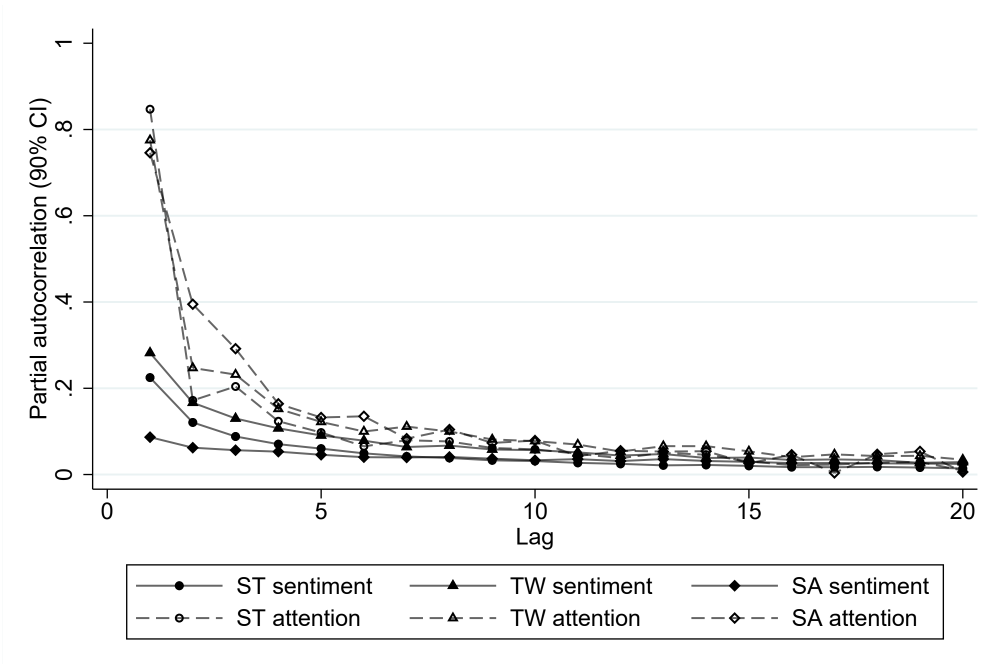
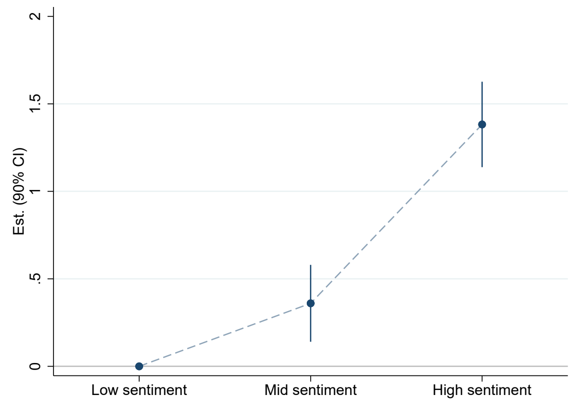
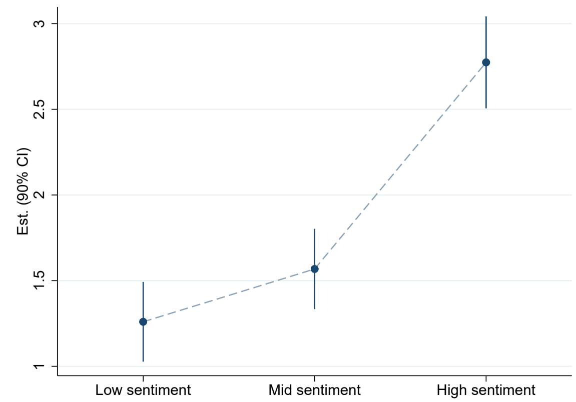


Figure 3: Partial Auto-correlation Function for Social Signals

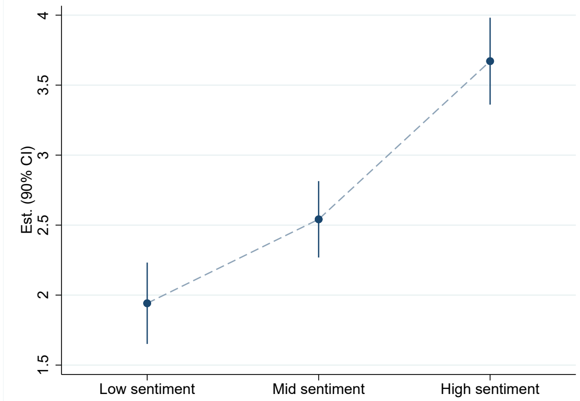
Note: This figure reports the partial auto-correlation for attention and sentiment on StockTwits (ST), Twitter (TW), and Seeking Alpha (SA). Sample consists of firm-day observations with at least 10 messages on StockTwits.



A1. Low attention

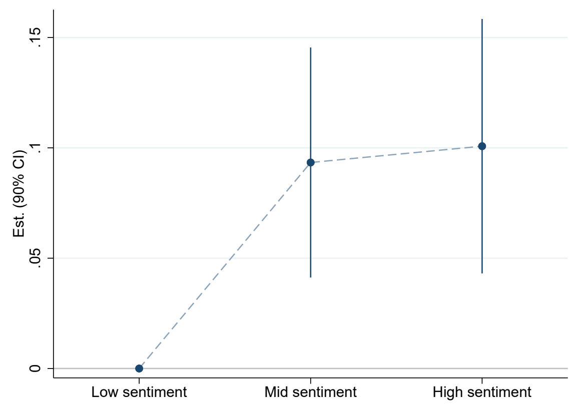


A2. Mid attention

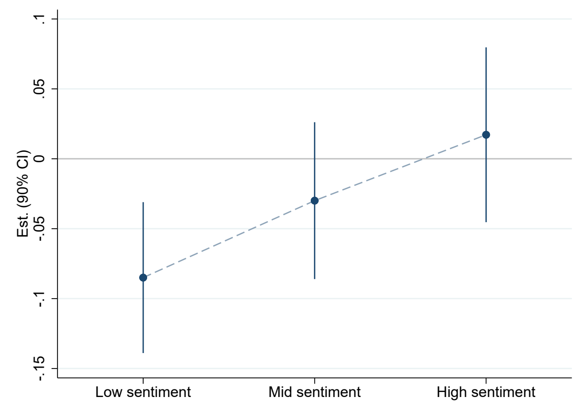


A3. High attention

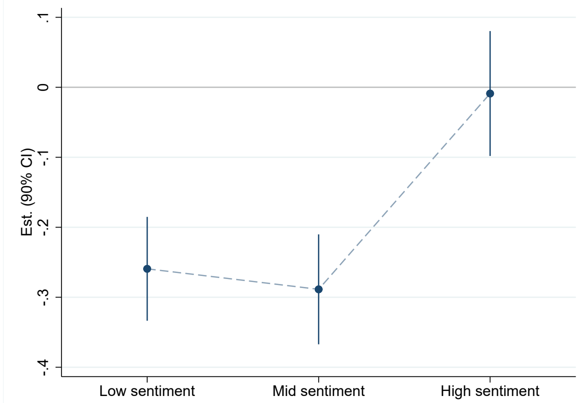
Panel A. Dep. variable: Retail trading imbalance



B1. Low attention



B2. Mid attention

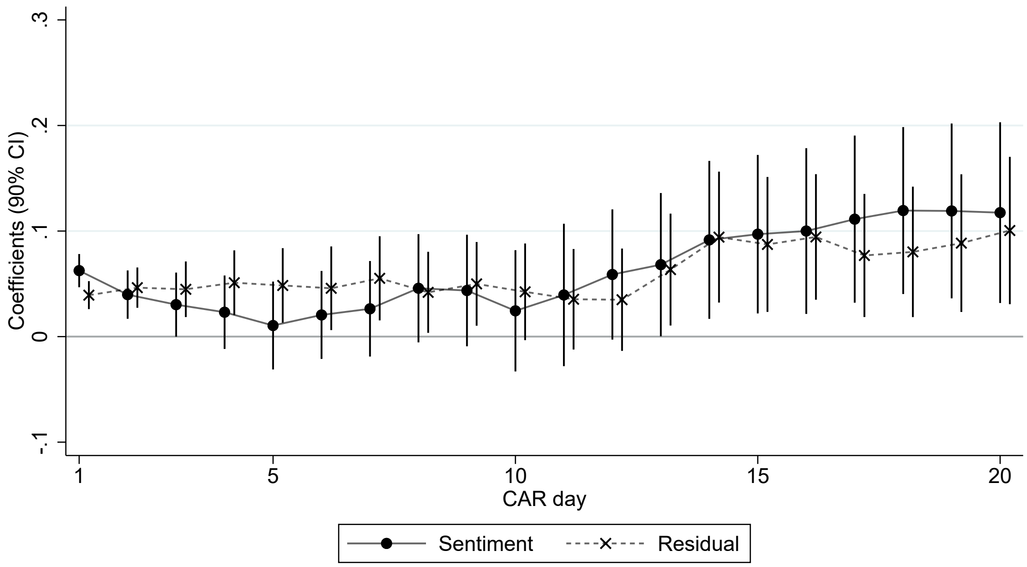


B3. High attention

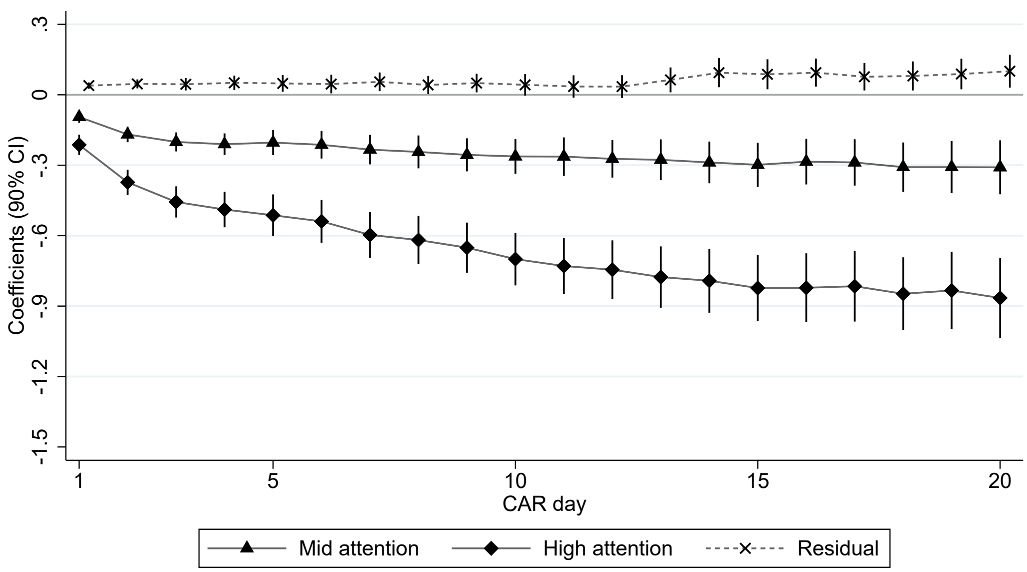
Panel B. Dep. variable: CAR t+1

Figure 4: Return and Trading Prediction of Social Signals

Note: This figure presents how sentiment-and-attention bins predict CAR t+1 and retail trading imbalance relative to the low-attention-and-low-sentiment bin. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome variable is retail trading imbalance scaled by 100 in panel A and CAR t+1 scaled by 100 in panel B. Panel A plots the coefficients and 90 percent confidence intervals for the sentiment-and-attention bins (i.e., "Mid sentiment × Low attention", ..., "High sentiment × High attention") in Table A3 column (2); panel B plots those in Table A3 column (1). Standard errors are clustered by firm and by date.



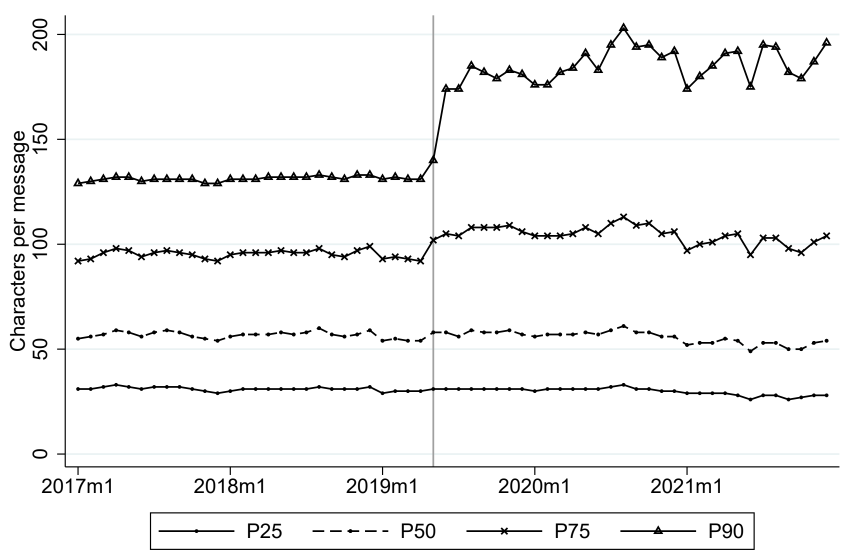
(a) Imbalance from sentiment v.s. from residual



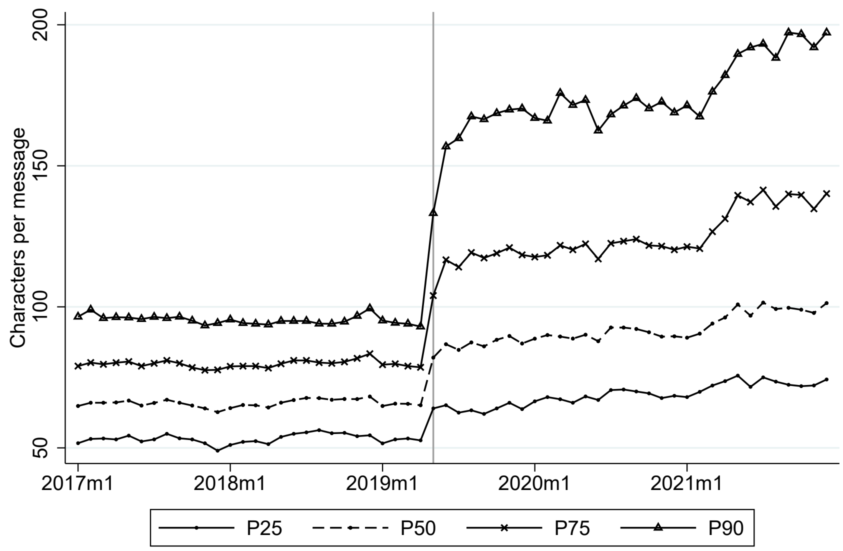
(b) Imbalance from attention v.s. from residual

Figure 5: Informativeness of Components of Retail Trading Imbalance: CAR t+1 through CAR t+1∼t+20

Note: This figure reports how the part of retail trading imbalance predicted by social signal and its residual relate to returns. The plotted coefficients and 90 percent confidence intervals are for "Imbalance from sentiment (z)" (Sentiment), "Imbalance from mid attention" (Mid attention), "Imbalance from high attention" (High attention), and "Residual imbalance (z)" (Residual) from Table 5 column (4) using CAR t+1, CAR t+1∼t+2, ..., CAR t+1∼t+20 as outcomes. Sample consists of firm-day observations with at least 10 messages on StockTwits. Standard errors are clustered by firm and by date.



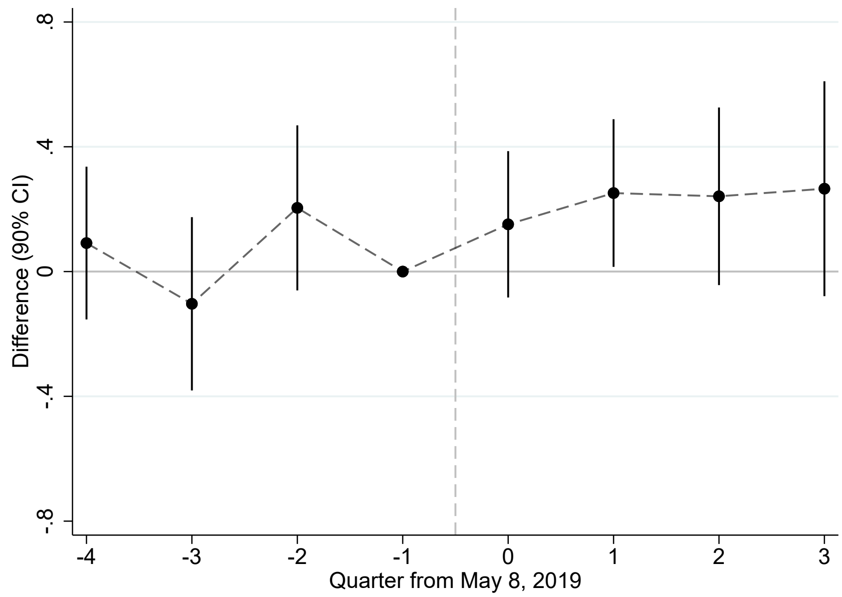
(a) Number of characters per message



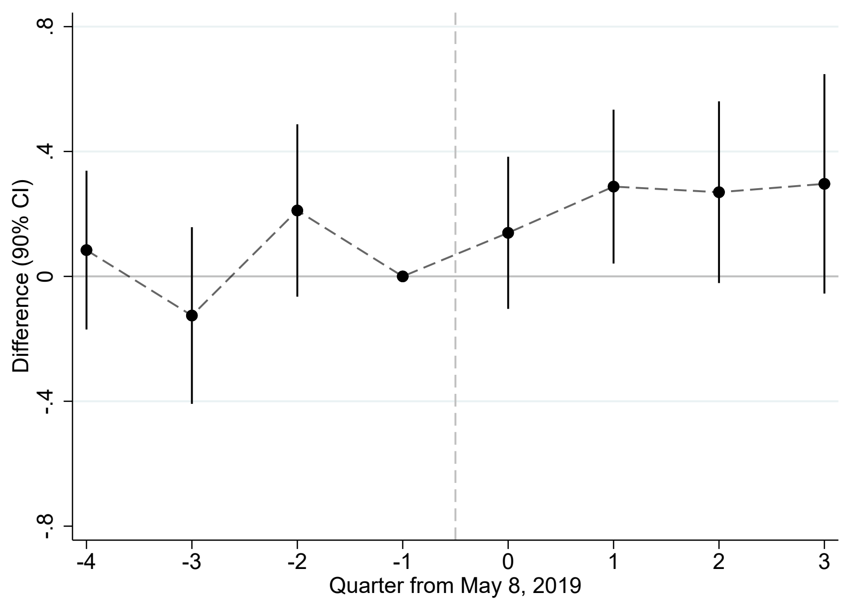
(b) Firm-day level average number of characters per message

Figure 6: Monthly Quartile of Number of Characters per Message

Note: This figure plots the monthly quartile of number of characters per message (panel A) and the monthly quartile of the firm-day level average number of characters per message (panel B). The vertical line represents May 8, 2019, the date when StockTwits increased its character limit from 140 to 1,000.



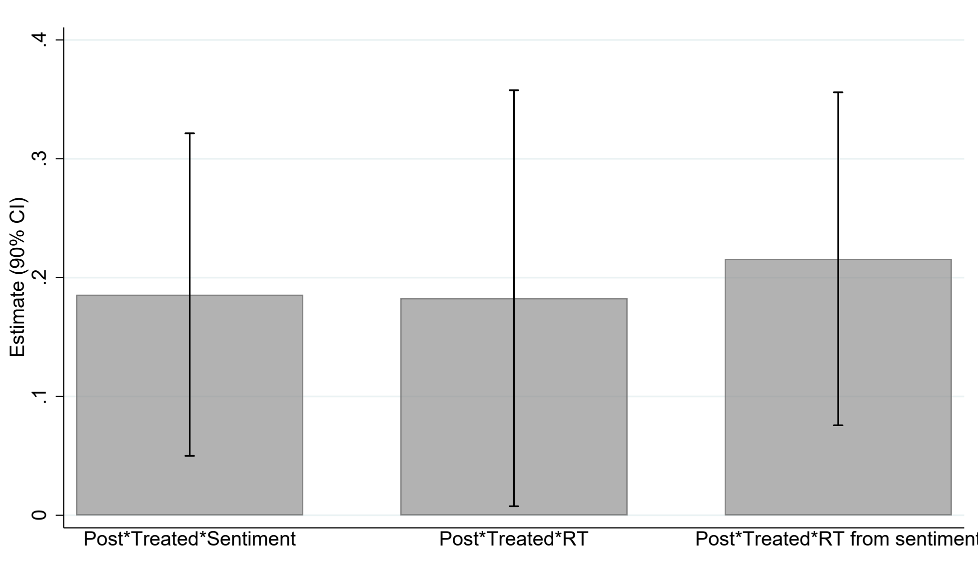
(a) How sentiment relates to CAR t+1



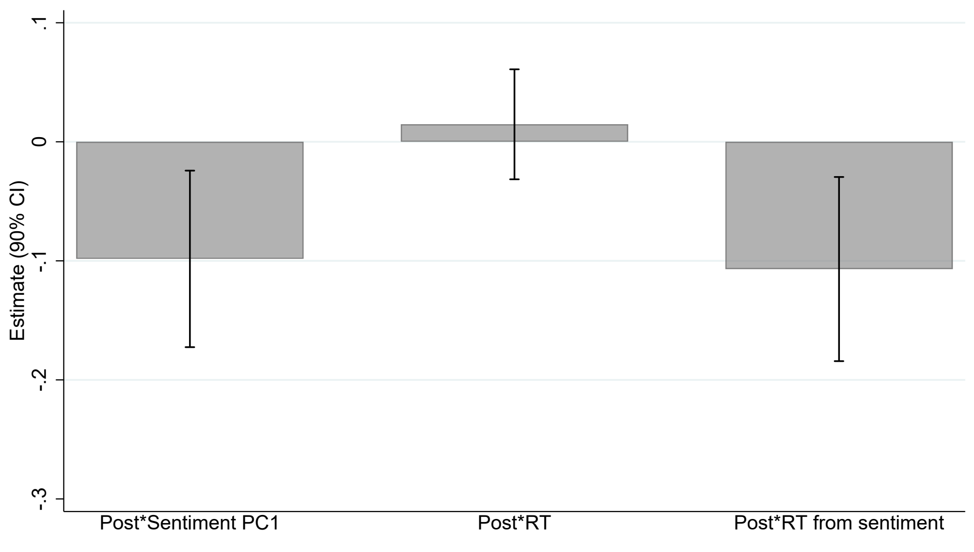
(b) How retail trade predicted by sentiment relates to CAR t+1

Figure 7: How Did the Informativeness of Sentiment and Retail Trade Predicted by Sentiment for Next-day Returns Change Around StockTwits Character Limit Increase?

Note: This figure compares the change in how StockTwits sentiment (panel A) and retail trading imbalance predicted by StockTwits sentiment relate to CAR t+1 in the quarters around StockTwits character limit increase on May 8, 2019. The treated group is stocks whose daily average number of characters per message is in the top quartilel; the comparison group is the stocks whose daily average number of characters per message is in the bottom quartile. Sample consists of firm-day observations with at least 10 messages on StockTwits between May 8, 2018 and May 8, 2020. Event time 0 represents the three months following May 8, 2019. The omitted period is -1. Coefficients and 90 percent confidence intervals are plotted. Specifications and variable definitions in panels A and B mirror those in Tables A6 column (1) and A7 column (2), respectively.



(a) Experiment: StockTwits character limit increase



(b) Experiment: GameStop event

Figure 8: How Did the Informativeness of Sentiment and Components of Retail Trade for Next-day Returns Change Around StockTwits character limit increase and the GameStop Event?

Note: This figure plots the change in how sentiment, retail trading imbalance, and retail trading imbalance predicted by sentiment relate to CAR t+1 before vs after two information experiments: StockTwits character limit increase (panel A) and GameStop event (panel B). Coefficients and 90 percent confidence intervals in panel A are from Table A6 column (1) and Table A7 columns (1) and (2). Coefficients and 90 percent confidence intervals in panel B are from Table 8 column (1) and Table A8 columns (1) and (2).

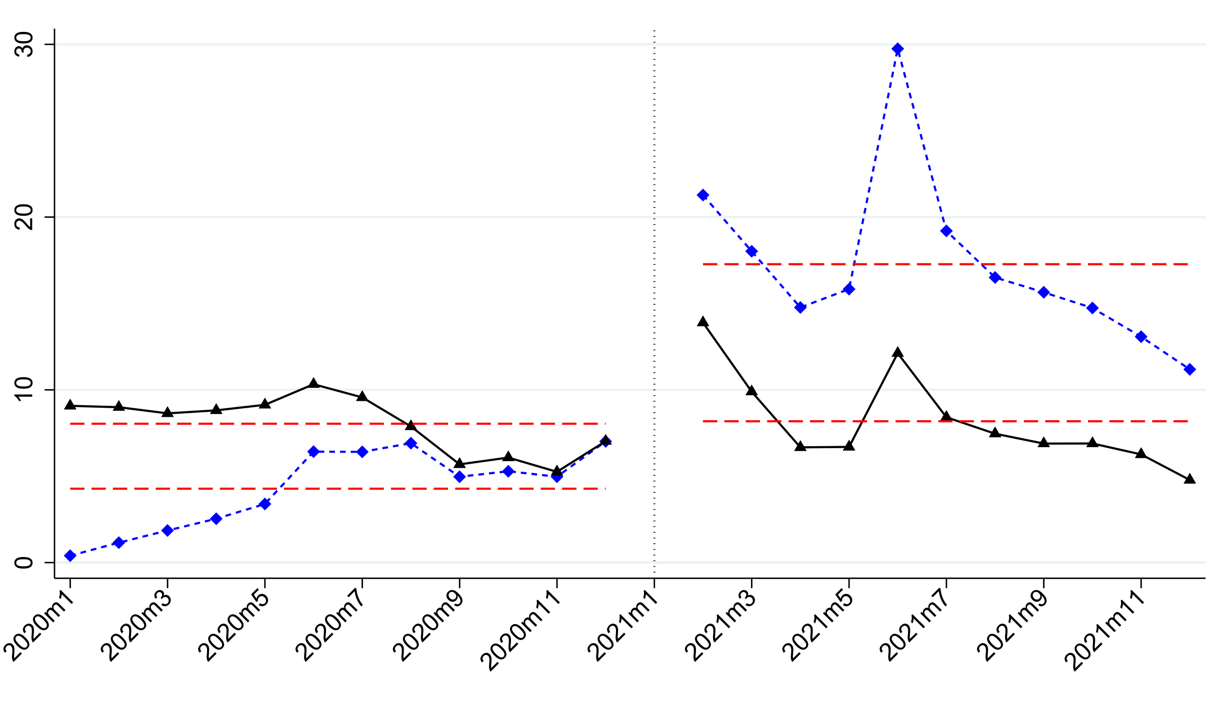


Figure 9: StockTwits Mentions of "Short Squeeze" Around the GameStop Event
Old versus New Users

Note: This figure presents evidence on the changing user composition of StockTwits in the months around the GME short squeeze event. Specifically, the figure plots mentions of "short squeeze" from new users (blue diamonds) and old users (black triangles) by month in the GME event window. Old users are those who joined StockTwits before January 2020 and new users are those joined in or after January 2020.

Table 1: Summary Statistics

Panel A: Across social media platforms

	Daily sentiment				# messages (daily)				# of firms		Firm-day observations	
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Ever mentioned	All	Mentioned	All
StockTwits	0.10	0.14	-0.97	0.97	132.40	734.97	10	138,280	1,497	1,500	815,980	815,980
Twitter	0.02	0.06	-0.80	0.94	18.84	62.69	0	7,160	1,271	1,500	522,284	815,980
Seeking Alpha	0.02	0.12	-1	1	0.46	1.75	0	150	1,283	1,500	137,018	815,980

Panel B: Across user types on StockTwits

	Daily sentiment				# messages (daily)				Users		Firm-day observations	
	Mean	DS.	Min	Max	Mean	DS.	Min	Max	#	Share	Non-zero	All
Top 1%	0.07	0.29	-1	1	5.97	34.04	0	4,212	7,173	0.01	512,549	815,980
Professional	0.09	0.30	-1	1	7.89	28.92	0	2,405	20,073	0.02	591,383	815,980
Intermediate	0.09	0.29	-1	1	13.37	50.88	0	5,439	45,156	0.05	687,993	815,980
Novice	0.07	0.29	-1	1	5.87	26.18	0	3,645	34,118	0.04	514,773	815,980
No experience	0.10	0.18	-0.99	0.99	105.27	658.60	0	127,243	730,164	0.88	810,614	815,980
Self-labeled	0.66	0.46	-1	1	69.30	494.75	0	100,680	649,542	0.78	792,386	815,980

Panel C: Firm-day observations satisfying sample restriction

Sample Restriction	# obs.	# dropped obs.
Full sample	2,795,852	-
At least 10 StockTwits messages	821,534	1,974,318
Non-missing controls data	815,980	5,554
Non-missing controls and return	814,646	1,334
Non-missing controls and retail trade	810,484	5,496

Note: Panel A provides statistics on the firm-day level sentiment and attention by social media platform for all firm-day observations with at least 10 StockTwits messages. The sample time frame is Jan. 1, 2012 to Dec. 31, 2021. "# of firms - Ever mentioned" refers to the # of firms ever mentioned on a platform during our sample period; "# of firms - All" refers to the # firms included in our analysis sample (with the sentiment of firms not mentioned replaced by zeros). "Firm-day observations - Mentioned" refers to the # of firm-day observations with non-zero attention; "firm-day observations - All" refers to the # of firm-day observations in our analysis sample. Panel B provides similar statistics on the firm-day level sentiment and attention by StockTwits user type. "Users - # (or Share)" refers to the # (or share) of StockTwits users of a certain type; self-labeled refers to the # of users who posted at least one message with self-labeled sentiment on StockTwits. Panel C shows how sample restrictions reduce the # of firm-day observations to arrive at our analysis sample.

Table 2: How Common is the Social Signal across Platforms?

Panel A: Correlations with the StockTwits Signal

	Twitter	Seeking Alpha	DJNW	WSJ
StockTwits attention	0.595	0.398	0.220	0.163
StockTwits sentiment	0.125	0.038	0.032	0.010

Panel B: PCA of Platform-Level Attention Signals

	Comp1	Comp2	Comp3
StockTwits	0.565	-0.665	0.489
Twitter	0.614	-0.057	-0.787
Seeking Alpha	0.551	0.745	0.376
Fraction of variation	70.0%	18.9%	11.1%

Panel C: PCA of Platform-Level Sentiment Signals

	Comp1	Comp2	Comp3
StockTwits	0.611	-0.464	0.642
Twitter	0.662	-0.147	-0.735
Seeking Alpha	0.435	0.874	0.217
Fraction of variation	38.8%	32.3%	29.0%

Note: This table reports the correlations and principal component analyses of social signals across platforms. Panel A reports the bivariate correlations of attention and sentiment between StockTwits and another platform. Panels B and C reports the principal components for attention and sentiment across platforms, respectively. Sample consists of firm-day observations with at least 10 messages on StockTwits.

Table 3: How Common is the Social Signal across User Types on StockTwits?

Panel A: Correlations with the StockTwits Signal

	Top 1%	Professional	Intermediate	Novice	No experience	Self-labeled
StockTwits attention	0.819	0.884	0.966	0.929	0.987	0.931
StockTwits sentiment	0.232	0.313	0.348	0.222	0.783	0.387

Panel B: PCA of Attention Signals

	Comp1	Comp2	Comp3	Comp4
StockTwits	0.374	-0.141	0.108	-0.093
Top 1%	0.329	-0.086	-0.362	0.723
Professional	0.353	0.009	-0.312	0.216
Intermediate	0.368	-0.095	0.037	-0.084
Novice	0.347	-0.210	0.220	-0.163
No experience	0.366	-0.144	0.193	-0.130
Self-labeled	0.348	-0.172	0.237	-0.178
Twitter	0.273	0.514	-0.586	-0.511
Seeking Alpha	0.203	0.778	0.523	0.280
Fraction of variation	76.6%	10.6%	4.7%	3.4%

Panel C: PCA of Sentiment Signals

	Comp1	Comp2	Comp3	Comp4
StockTwits	0.595	-0.121	0.063	-0.048
Top 1%	0.255	0.327	-0.542	0.075
Professional	0.262	0.383	-0.520	-0.042
Intermediate	0.270	-0.043	-0.045	-0.160
Novice	0.184	0.024	0.140	0.956
No experience	0.502	-0.239	0.212	-0.171
Self-labeled	0.362	-0.248	0.122	-0.044
Twitter	0.144	0.514	0.330	-0.027
Seeking Alpha	0.052	0.589	0.491	-0.139
Fraction of variation	26.6%	12.0%	11.1%	10.7%

Note: This table reports the correlations and principal component analyses of social signals across different user types on StockTwits and other platforms. Panel A reports the bivariate correlations of attention and sentiment between StockTwits total signal and signal by user type. Panels B and C report the first four principal components for attention and sentiment across all user types and platforms, respectively (see Table A2 for all principal components). Sample consists of firm-day observations with at least 10 messages on StockTwits.

Table 4: How Does Retail Trading Imbalance Relate to Social Signals?

	Dep. variable: retail trading imbalance t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ST sentiment (z)		0.719***			0.695***		
		(0.047)			(0.048)		
ST attention (z)		1.285***			1.469***		
		(0.186)			(0.168)		
ST sentiment (z) × ST attention (z)		0.568***			0.527***		
		(0.117)			(0.128)		
Twitter sentiment (z)			0.219***		0.140***		
			(0.029)		(0.028)		
Twitter attention (z)			0.295***		-0.317***		
			(0.080)		(0.070)		
Twitter sentiment (z) × Twitter attention (z)			0.031		-0.067*		
			(0.036)		(0.041)		
SA sentiment (z)				0.147***	0.119***		
				(0.024)	(0.023)		
SA attention (z)				0.109**	0.085**		
				(0.051)	(0.043)		
SA sentiment (z) × SA attention (z)				-0.028	-0.029*		
				(0.017)	(0.016)		
Sentiment PC1 (z)						0.781***	0.784***
						(0.048)	(0.048)
Sentiment PC2 (z)							0.121***
							(0.028)
Sentiment PC3 (z)							0.007
							(0.026)
Attention PC1 (z)						1.226***	1.275***
						(0.209)	(0.201)
Sentiment PC1 × Attention PC1 (z)						0.300***	0.299***
						(0.068)	(0.060)
Sentiment PC2 × Attention PC1 (z)							-0.094**
							(0.043)
Sentiment PC3 × Attention PC1 (z)							-0.031
							(0.048)
DJNW sentiment (z)	0.210***	0.157***	0.192***	0.183***	0.136***	0.137***	0.116***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
DJNW attention (z)	0.080	0.022	0.009	0.062	0.031	-0.024	-0.049
	(0.056)	(0.058)	(0.059)	(0.058)	(0.059)	(0.058)	(0.057)
8-K report date	-0.780***	-0.879***	-0.912***	-0.808***	-0.749***	-0.977***	-1.046***
	(0.139)	(0.141)	(0.145)	(0.140)	(0.147)	(0.146)	(0.145)
EA date	2.050***	1.912***	2.137***	1.954***	1.855***	1.837***	1.864***
	(0.221)	(0.226)	(0.225)	(0.222)	(0.224)	(0.230)	(0.229)
Retail trade imbalance t-1∼ t-5	0.155***	0.154***	0.155***	0.155***	0.154***	0.154***	0.154***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Retail trade imbalance t-6∼ t-30	0.182***	0.182***	0.182***	0.182***	0.182***	0.184***	0.183***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Volatility t-1∼ t-5	-2.509***	-2.774***	-2.142***	-2.403***	-2.355***	-2.740***	-2.250***
	(0.549)	(0.569)	(0.561)	(0.551)	(0.592)	(0.563)	(0.561)
CAR t-1∼ t-5	0.001	-0.001	0.000	0.001	-0.003	-0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CAR t-6∼ t-30	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***	-0.002***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Lagged signals	N	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	-0.326	-0.329	-0.329	-0.329	-0.329	-0.329	-0.329
Outcome SD	23.173	23.174	23.174	23.174	23.174	23.174	23.174
Observations	811,157	810,484	810,484	810,484	810,484	810,484	810,484
R^2	0.0202	0.0216	0.0205	0.0203	0.0219	0.0219	0.0220
Within R^2	0.0110	0.0124	0.0112	0.0111	0.0126	0.0127	0.0128
Root MSE	22.9728	22.9575	22.9713	22.9731	22.9555	22.9546	22.9530

Note: This table reports how retail trading imbalance relates to social signals. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome is retail trading imbalance scaled by 100. All regressions control for date fixed effects and lagged 1-10 sentiment and attention. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table 5: Informativeness of Components of Retail Trading Imbalance (CAR t+1)

	Dep. variable: CAR t+1				
	(1)	(2)	(3)	(4)	(5)
Retail trading imbalance (z)	0.035***				
	(0.008)				
Imbalance from sentiment (z)		0.062***		0.063***	0.071***
		(0.010)		(0.010)	(0.009)
Imbalance from mid attention (z)			-0.091***	-0.095***	-0.095***
			(0.015)	(0.015)	(0.015)
Imbalance from high attention (z)			-0.208***	-0.213***	-0.217***
			(0.027)	(0.026)	(0.026)
Imbalance from sentiment × Mid attention (z)					0.008
					(0.006)
Imbalance from sentiment × High attention (z)					-0.021**
					(0.009)
Residual imbalance (z)		0.034***	0.041***	0.039***	0.039***
		(0.008)	(0.008)	(0.008)	(0.008)
DJNW sentiment (z)	0.081***	0.076***	0.088***	0.084***	0.081***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
DJNW attention (z)	-0.011	-0.009	-0.001	0.001	-0.001
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
8-K report date	0.034	0.036	0.123***	0.128***	0.126***
	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)
EA date	-0.553***	-0.558***	-0.434***	-0.435***	-0.438***
	(0.091)	(0.092)	(0.094)	(0.094)	(0.093)
Retail trade imbalance t-1∼ t-5	0.001	0.000	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Retail trade imbalance t-6∼ t-30	0.005***	0.005***	0.006***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Volatility t-1∼ t-5	-0.683**	-0.656**	-0.532	-0.508	-0.501
	(0.323)	(0.324)	(0.335)	(0.336)	(0.337)
CAR t-1∼ t-5	-0.005***	-0.005***	-0.005**	-0.005**	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CAR t-6∼ t-30	-0.001**	-0.001**	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lagged signals	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
Outcome Mean	-0.046	-0.047	-0.047	-0.047	-0.047
Outcome SD	7.115	7.117	7.117	7.117	7.117
Observations	810,517	809,844	809,844	809,844	809,844
R^2	0.032	0.032	0.032	0.032	0.032

Note: This table reports how the components of retail trading imbalance predicted by sentiment, attention, and/or their interactions, and a residual relate to CAR t+1. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome is CAR t+1 scaled by 100. The components come from an auxiliary regression of retail trading imbalance on all social media sentiment signals, terciles of the first PC of attention, and/or the interactions between sentiment and attention signals. The specification follows column 5 of Table 4 but also includes within-StockTwits signals (such as attention and sentiment from influencers, professionals, etc). "Residual imbalance" is the residual from this regression, and "Imbalance from sentiment" refers to the component of the fitted regression that depends on social media sentiment signals; "Imbalance from mid (or high) attention" and "Imbalance from sentiment × mid (or high) attention"are similarly defined. All regressions control for firm fixed effects, date fixed effects, and lagged 1-10 sentiment and attention. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table 6: Informativeness of Components of Retail Trade Imbalance (CAR t+1∼t+20)

	Dep. variable: CAR t+1∼t+20				
	(1)	(2)	(3)	(4)	(5)
Retail trading imbalance (z)	0.091**				
	(0.043)				
Imbalance from sentiment (z)		0.101**		0.117**	0.111**
		(0.051)		(0.052)	(0.049)
Imbalance from mid attention (z)			-0.306***	-0.309***	-0.304***
			(0.070)	(0.070)	(0.069)
Imbalance from high attention (z)			-0.861***	-0.865***	-0.874***
			(0.105)	(0.104)	(0.104)
Imbalance from sentiment × Mid attention (z)					0.032
					(0.034)
Imbalance from sentiment × High attention (z)					-0.124**
					(0.050)
Residual imbalance (z)		0.086**	0.103**	0.100**	0.102**
		(0.042)	(0.043)	(0.042)	(0.042)
DJNW sentiment (z)	0.099***	0.089***	0.117***	0.108***	0.103***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
DJNW attention (z)	-0.061*	-0.060*	-0.027	-0.021	-0.033
	(0.033)	(0.033)	(0.034)	(0.034)	(0.034)
8-K report date	0.088	0.083	0.341*	0.357**	0.340*
	(0.166)	(0.167)	(0.174)	(0.174)	(0.175)
EA date	-0.174	-0.197	0.060	0.061	0.048
	(0.214)	(0.213)	(0.221)	(0.221)	(0.220)
Retail trade imbalance t-1∼ t-5	0.009	0.009	0.011	0.011	0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Retail trade imbalance t-6∼ t-30	0.058***	0.058***	0.069***	0.068***	0.068***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Volatility t-1∼ t-5	-2.506	-2.476	-0.470	-0.260	-0.244
	(2.080)	(2.094)	(2.030)	(2.047)	(2.047)
CAR t-1∼ t-5	-0.039***	-0.039***	-0.040***	-0.040***	-0.040***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
CAR t-6∼ t-30	-0.015***	-0.015***	-0.013**	-0.013**	-0.013**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Lagged signals	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y
Outcome Mean	-0.549	-0.554	-0.554	-0.554	-0.554
Outcome SD	30.141	30.151	30.151	30.151	30.151
Observations	796,798	796,125	796,125	796,125	796,125
R^2	0.096	0.096	0.097	0.097	0.097

Note: This table reports how the components of retail trading imbalance predicted by sentiment, attention, and/or their interactions, and a residual relate to CAR t+1∼t+20. The outcome is CAR t+1∼t+20 scaled by 100. The sample, specification, and variable definitions mirror those in Table 5.
*** 1%, ** 5%, * 10% significance level

Table 7: How Did the Informativeness of Different Social Signals for Next-Day Returns and Retail Trade Change around StockTwits Character Limit Increase?

	Dep. variable: CAR t+1				Dep. variable: Retail imbalance t			
	(1) StockTwits	(2) StockTwits top quartile	(3) Twitter	(4) Seeking Alpha	(5) StockTwits	(6) StockTwits top quartile	(7) Twitter	(8) Seeking Alpha
Post × Sentiment (z)	0.066*	0.137**	-0.010	-0.004	-0.161	0.036	0.068	-0.008
	(0.034)	(0.055)	(0.043)	(0.034)	(0.121)	(0.237)	(0.099)	(0.071)
Post × Attention (z)	0.153*	-0.325	-0.007	-0.016	-0.101	2.114**	0.017	-0.163*
	(0.085)	(0.222)	(0.027)	(0.032)	(0.243)	(0.897)	(0.050)	(0.084)
Sentiment (z)	0.030	0.002	-0.003	0.079***	0.696***	0.376**	0.172***	0.126**
	(0.023)	(0.037)	(0.018)	(0.024)	(0.091)	(0.167)	(0.057)	(0.055)
Attention (z)	-0.355***	-0.249	-0.029	-0.016	0.747***	0.236	0.146**	0.097*
	(0.110)	(0.185)	(0.027)	(0.023)	(0.256)	(0.778)	(0.070)	(0.054)
DJNW sentiment (z)	0.097***	0.092***	0.101***	0.086***	0.198***	0.321***	0.233***	0.224***
	(0.016)	(0.026)	(0.016)	(0.016)	(0.041)	(0.087)	(0.041)	(0.040)
DJNW attention (z)	0.026	-0.011	-0.003	-0.009	0.066	0.463*	0.102	0.125**
	(0.027)	(0.072)	(0.029)	(0.026)	(0.069)	(0.241)	(0.067)	(0.063)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	-0.093	0.004	-0.093	-0.093	-0.291	-1.123	-0.291	-0.291
Outcome SD	7.823	6.451	7.823	7.823	24.264	26.028	24.264	24.264
Observations	214,260	53,447	214,260	214,260	212,796	53,090	212,796	212,796
R^2	0.027	0.066	0.026	0.026	0.030	0.059	0.029	0.029

Note: This table compares how social signals from different platforms and/or user types changed their predictive power for CAR t+1 and retail trading imbalance before versus after StockTwits character limit increase on May 8, 2019. Sample consists of firm-day observations with at least 10 messages on StockTwits between May 8, 2018 and May 8, 2020. The outcome is CAR t+1 scaled by 100 in columns (1)-(4) and retail trading imbalance scaled by 100 in columns (5)-(8). *Post* is one if a day is on or after May 8, 2019. Social signals in columns (1) and (5), (2) and (6), (3) and (7), and (4) and (8) are based on messages from all users on StockTwits (StockTwits), with a number of characters in the top daily quartile on StockTwits (StockTwits top quartile), from Twitter (Twitter), and from Seeking Alpha (Seeking Alpha), respectively. Controls are 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.

*** 1%, ** 5%, * 10% significance level

Table 8: How Did the Informativeness of Different Social Signals for Next-Day Returns and Retail Trade Change around GameStop Event?

	Dep. variable: CAR t+1				Dep. variable: Retail imbalance t			
	(1) PC signal	(2) PC signal	(3) StockTwits old	(4) StockTwits new	(5) PC signal	(6) PC signal	(7) StockTwits old	(8) StockTwits new
Post × Sentiment (z)	-0.098**	-0.098**	0.001	-0.102**	-0.267**	-0.268**	-0.039	-0.049
	(0.045)	(0.045)	(0.034)	(0.043)	(0.106)	(0.106)	(0.107)	(0.106)
Post × Attention (z)	0.006	0.006	-0.017	0.016	-0.228	-0.227	-0.125	-0.364*
	(0.097)	(0.097)	(0.091)	(0.109)	(0.276)	(0.275)	(0.254)	(0.189)
Sentiment (z)	0.095**	0.095**	0.035	0.095**	0.766***	0.761***	0.498***	0.256***
	(0.039)	(0.039)	(0.028)	(0.039)	(0.081)	(0.081)	(0.074)	(0.090)
Attention (z)	-0.077	-0.078	-0.070	-0.069	0.506**	0.504**	0.501*	0.569***
	(0.058)	(0.058)	(0.056)	(0.067)	(0.247)	(0.246)	(0.259)	(0.102)
Post × Sentiment PC2 (z)		-0.009				-0.015		
		(0.031)				(0.082)		
Post × Sentiment PC3 (z)		0.022				-0.077		
		(0.030)				(0.076)		
Sentiment PC2 (z)		0.011				0.137**		
		(0.025)				(0.060)		
Sentiment PC3 (z)		0.016				0.046		
		(0.026)				(0.060)		
DJNW sentiment (z)	0.086***	0.082***	0.087***	0.088***	0.192***	0.176***	0.210***	0.229***
	(0.016)	(0.016)	(0.016)	(0.017)	(0.034)	(0.035)	(0.034)	(0.034)
DJNW attention (z)	-0.065**	-0.066**	-0.066**	-0.070**	0.192***	0.182***	0.183***	0.210***
	(0.031)	(0.032)	(0.031)	(0.031)	(0.056)	(0.056)	(0.055)	(0.051)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	-0.004	-0.004	-0.004	-0.004	-1.390	-1.390	-1.390	-1.390
Outcome SD	7.867	7.867	7.867	7.867	22.388	22.388	22.388	22.388
Observations	287,833	287,833	287,833	287,833	286,349	286,349	286,349	286,349
R^2	0.049	0.049	0.049	0.049	0.029	0.029	0.029	0.028

Note: This table compares how social signals from different platforms and/or user types changed their predictive power for CAR t+1 and retail trading imbalance before versus after the GameStop event on January 28, 2021. Sample consists of firm-day observations with at least 10 messages on StockTwits between February 1,2020 and December,31,2021, excluding January 2021. The outcome is CAR t+1 scaled by 100 in columns (1)-(4) and retail trading imbalance scaled by 100 in columns (5)-(8). *Post* is one if a day is on or after February 1, 2021. Social signals in columns (1)-(2) and (5)-(6), (3) and (7), and (4) and (8) are based on the first PC (standardized) of all attention or all sentiment signals in Table A2 panels A and B (PC signal), messages from users who joined StockTwits before 2020 (StockTwits old), and messages from users who joined StockTwits in or after 2020 (StockTwits new), respectively. Controls are 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table 9: How Did the Informativeness of Social Signals for Next-Day Returns and Retail Trade Change around GameStop Event?

Heterogeneity by Short Mentions and Utilization of Stocks

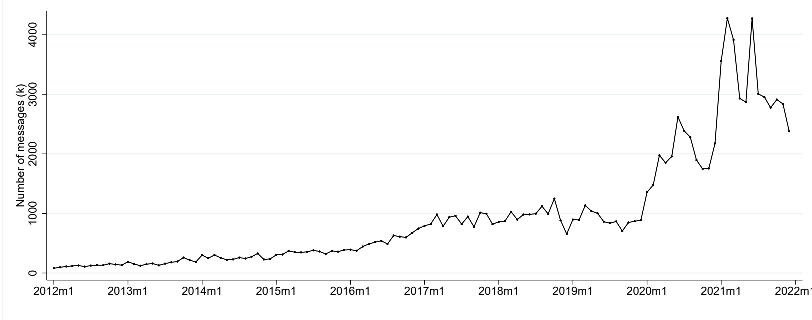
	Dep. variable: CAR t+1				Dep. variable: Retail imbalance t			
	(1) H mentions	(2) L mentions	(3) H utilization	(4) L utilization	(5) H mentions	(6) L mentions	(7) H utilization	(8) L utilization
Post × Sentiment PC1 (z)	-0.154**	-0.047	-0.225***	0.036	-0.314**	-0.172	-0.311*	-0.186
	(0.072)	(0.033)	(0.060)	(0.046)	(0.158)	(0.130)	(0.158)	(0.135)
Post × Attention PC1 (z)	0.033	1.088***	-0.032	-0.163**	-0.234	0.099	-0.636**	0.078
	(0.096)	(0.376)	(0.143)	(0.067)	(0.269)	(1.287)	(0.316)	(0.107)
Sentiment PC1 (z)	0.134**	0.053*	0.174***	-0.012	1.008***	0.493***	0.847***	0.654***
	(0.063)	(0.029)	(0.051)	(0.039)	(0.120)	(0.097)	(0.127)	(0.106)
Attention PC1 (z)	-0.089	-1.527***	0.000	-0.186*	0.462*	6.004***	0.849***	0.343
	(0.060)	(0.309)	(0.110)	(0.096)	(0.245)	(0.994)	(0.257)	(0.321)
DJNW sentiment (z)	0.100***	0.065***	0.134***	0.044***	0.236***	0.133**	0.253***	0.128***
	(0.028)	(0.014)	(0.028)	(0.013)	(0.044)	(0.052)	(0.050)	(0.044)
DJNW attention (z)	-0.119**	0.015	-0.349***	0.039	0.292***	-0.117	0.457***	0.102
	(0.055)	(0.019)	(0.067)	(0.036)	(0.081)	(0.080)	(0.109)	(0.066)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	0.076	-0.133	0.018	-0.030	-1.568	-1.104	-1.689	-1.034
Outcome SD	9.348	4.558	8.847	6.511	22.935	21.475	22.807	21.873
Observations	177,635	110,198	156,429	131,404	176,598	109,751	155,725	130,624
R^2	0.056	0.081	0.064	0.045	0.030	0.047	0.032	0.037

Note: This table compares how social signals among stocks with high vs. low interest from short sellers changed their predictive power for CAR t+1 and retail trading imbalance before versus after the GameStop event on January 28, 2021. Social signals are the first PC (standardized) of all attention or all sentiment signals in Table A2 panels A and B. Samples in columns (1) and (5) (or columns 2 and 6) consist of stocks with an above-median (or below-median) mentions of “squeeze,” “short interest,” “short seller” or “short volume” on StockTwits in a month. Samples in columns (3) and (7) (or columns 4 and 8) consist of stocks with an above-median (or below-median) utilization in a month. All other specifications and variable definitions mirror those in Table 8 columns (1) and (5). Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

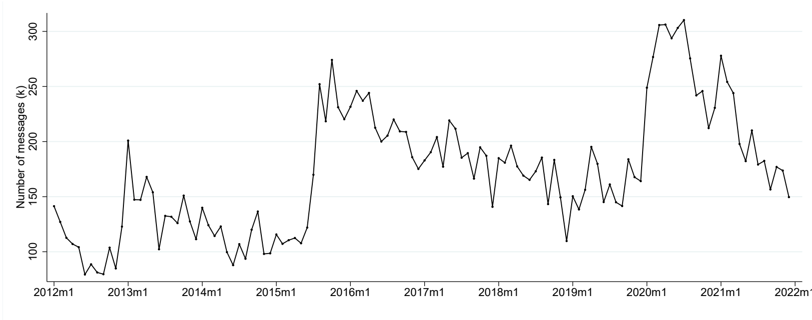
Online Appendix

The Social Signal

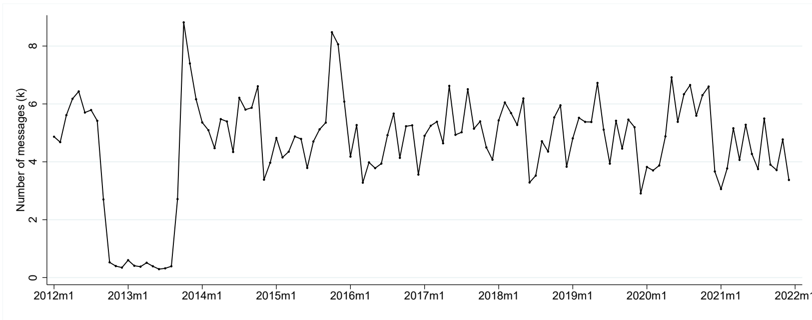
J. Anthony Cookson, Runjing Lu, William Mullins, and Marina Niessner



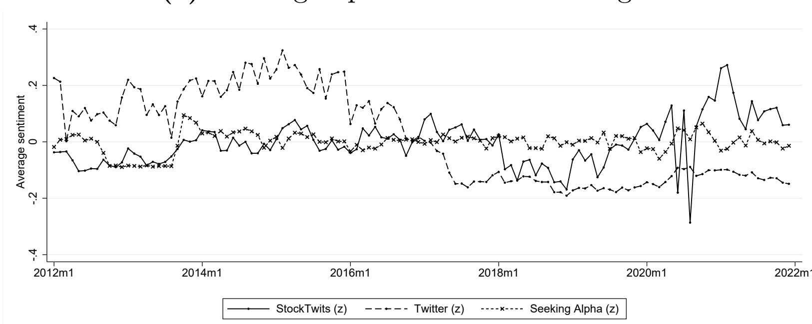
(a) StockTwits number of messages



(b) Twitter number of messages



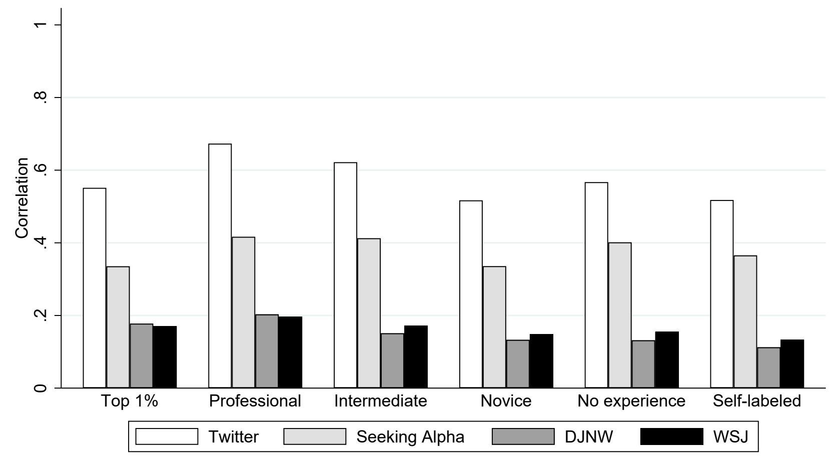
(c) Seeking Alpha number of messages



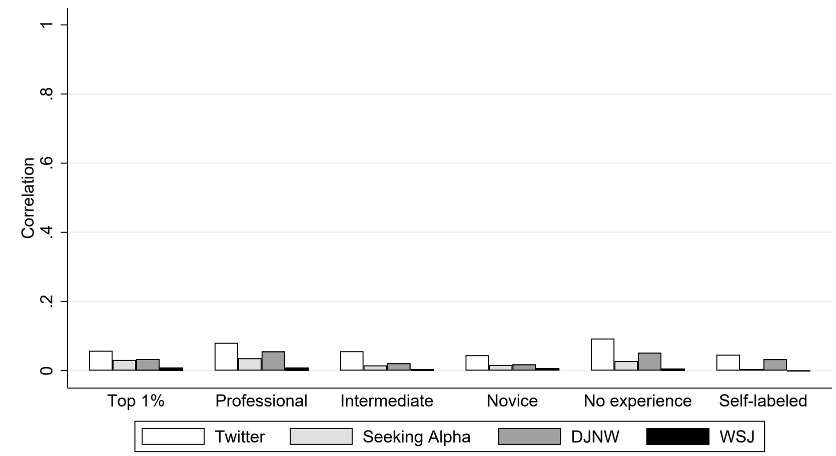
(d) Sentiment by platform

Figure A1: Monthly Number of Messages and Sentiment Across Platforms

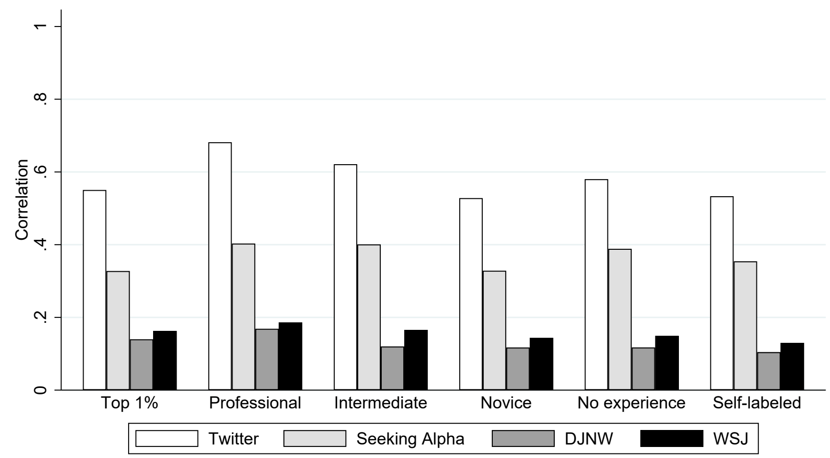
Note: This figure plots the monthly number of messages on StockTwits in panel A, Twitter in panel B, and Seeking Alpha in panel C, as well as monthly average standardized sentiment on each of the three platforms in panel D. Units are in thousands of messages in panels A-C and of one in panel D.



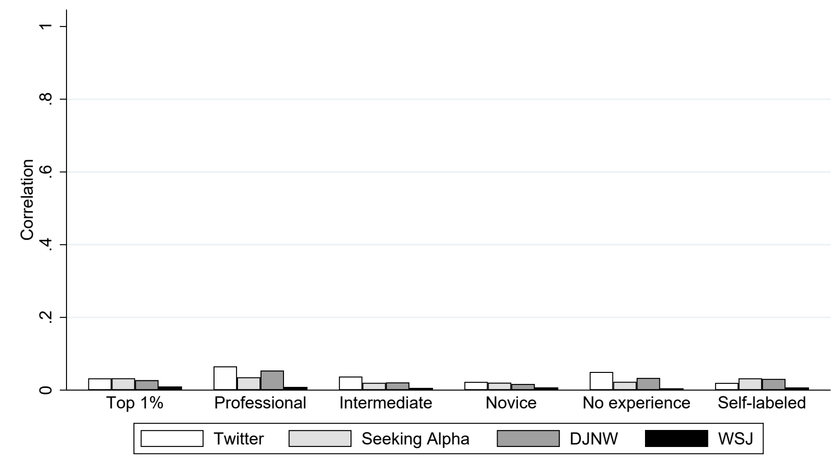
(a) Attention, firm-day with ≥ 10 Stocktwits messages



(b) Sentiment, firm-day with ≥ 10 Stocktwits messages



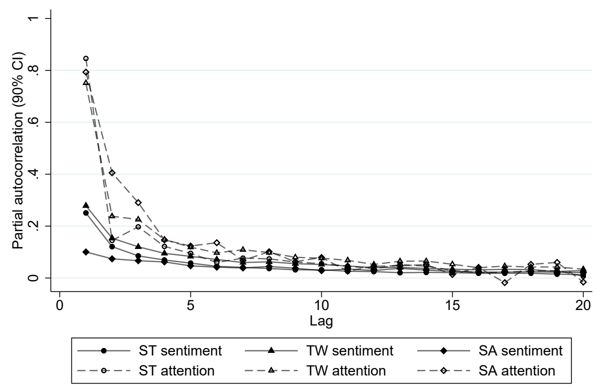
(c) Attention, all firm-day



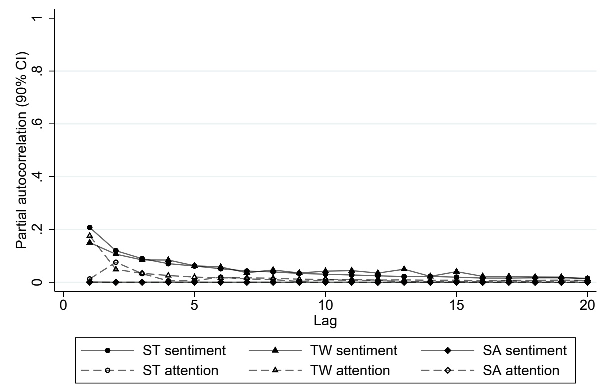
(d) Sentiment, all firm-day

Figure A2: Cross-platform Correlation for Attention and Sentiment
by Stocktwits User Type

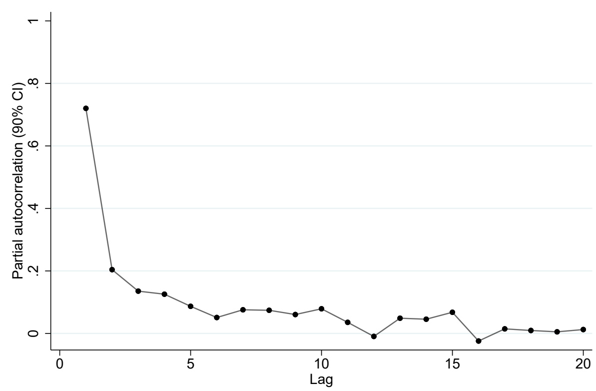
Note: This figure reports the bivariate correlations of attention and sentiment between various types of Stocktwits users and another platform at the firm-day level. Attention is measured by the fraction of messages, reports, or articles about a firm across all firms on a platform in a day. Sentiment is measured by the average sentiment of all messages about a firm on a platform in a day. Sample in panels A and B consists of firm-day observations with at least 10 messages on Stocktwits; sample in panels C and D consists of all firm-day observations.



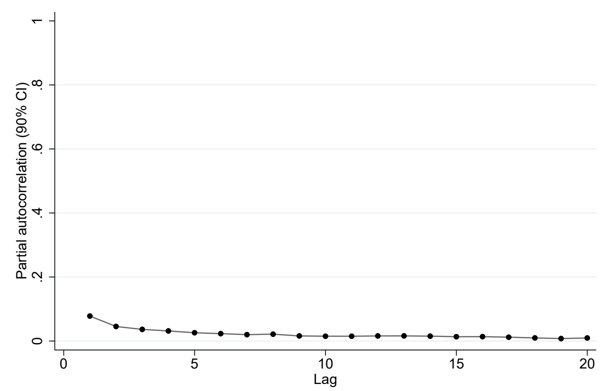
(a) Social signals: high attention on day 0



(b) Social signals: low attention on day 0



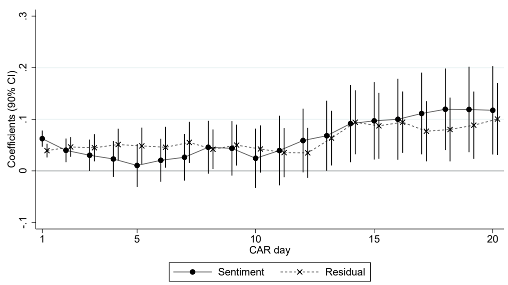
(c) Retail trading volume



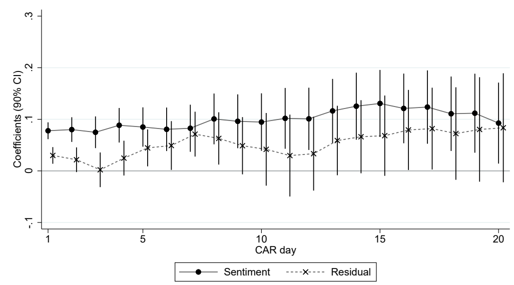
(d) Retail trading imbalance

Figure A3: Partial Auto-correlation Function for Social Signals and Retail Trading *by Level of Initial Attention*

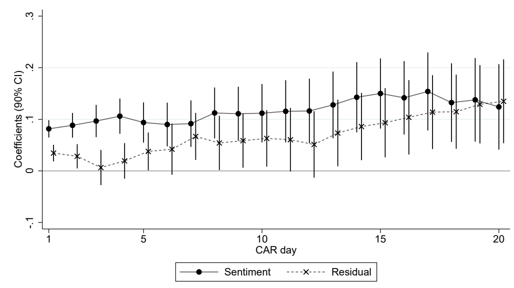
Note: This figure reports the partial auto-correlation for attention and sentiment on StockTwits (ST), Twitter (TW), and Seeking Alpha (SA) by whether attention on day 0 is high (panel A) or low (panel B). Panels C and D report the partial auto-correlation for retail trading volume and retail trading imbalance, respectively. Sample consists of firm-day observations with at least 10 messages on StockTwits.



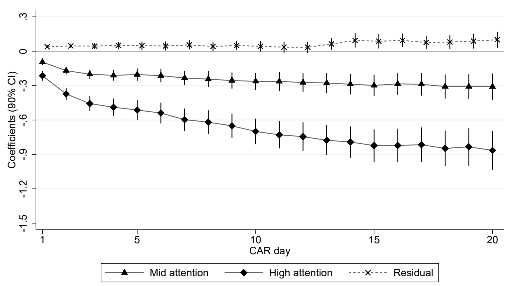
(a) Imbalance from sentiment, original



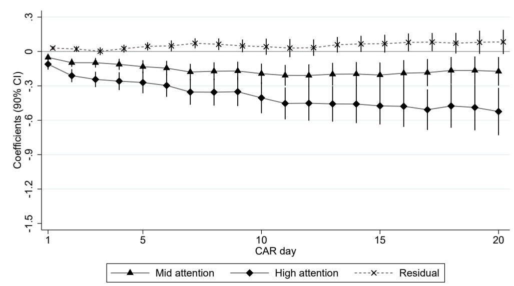
(b) New message selection



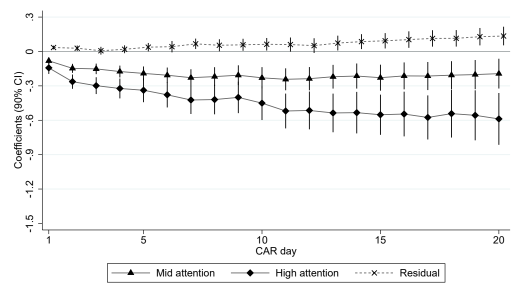
(c) New stock & message selection



(d) Imbalance from attention, original



(e) New message selection



(f) New stock & message selection

Figure A4: Informativeness of Components of Retail Trading Imbalance: CAR t+1 through CAR t+1∼t+20
Robustness Sample Selection

Note: This figure reports robustness checks for Figure 5 by varying sample selections based on (ii) total mentions on StockTwits (stock selection) and (ii) number of messages per day on a platform (message selection). Column (1) displays the same figures in Figure 5; column (2) presents the version for stock-day observations with at least 10 messages on StockTwits *and* at least 4 messages on Twitter; column (3) presents the version for 1,500 stocks with the highest mentions on Stocks during 2018-2021 and those satifying the new message selection. All other specifications and variable definitions mirror those in Figure 5.

Table A1: Partial Correlation of the Social Signal across Platforms

	(1) StockTwits	(2) Top 1%	(3) Professional	(4) Intermediate	(5) Novice	(6) No experience	(7) Self-labeled
Panel A: Attention							
Twitter	0.585***	0.548***	0.669***	0.617***	0.512***	0.545***	0.500***
	(0.085)	(0.057)	(0.063)	(0.098)	(0.075)	(0.086)	(0.046)
Seeking Alpha	0.136***	0.068***	0.096***	0.138***	0.102***	0.164***	0.149***
	(0.025)	(0.013)	(0.018)	(0.020)	(0.019)	(0.031)	(0.042)
DJNW	-0.142***	-0.052***	-0.085***	-0.149***	-0.110***	-0.152***	-0.148***
	(0.029)	(0.013)	(0.019)	(0.030)	(0.023)	(0.033)	(0.025)
WSJ	-0.031***	-0.014***	-0.028***	-0.030***	-0.019***	-0.029***	-0.035***
	(0.007)	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	821,534	821,534	821,534	821,534	821,534	821,534	821,534
R^2	0.373	0.308	0.462	0.407	0.278	0.347	0.291
Panel B: Sentiment							
Twitter	0.104***	0.118***	0.141***	0.101***	0.097***	0.098***	0.060***
	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Seeking Alpha	0.008***	0.022***	0.019***	0.007***	0.011***	0.007***	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
DJNW	0.030***	0.032***	0.050***	0.018***	0.017***	0.028***	0.025***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
WSJ	-0.000	0.003***	0.001	0.001	0.004***	-0.001	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	821,534	821,534	821,534	821,534	821,534	821,534	821,534
R^2	0.020	0.005	0.010	0.004	0.002	0.011	0.003

Note: This table reports the partial correlation of attention and sentiment across different platforms at the firm-day level. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcomes in panel A (or B) in columns (1) through (7) are daily attention (or sentiment) for each firm based on messages from all StockTwits users, influencers who have top 1% followers, professional users, intermediate users, novice users, users with no experience classification, and messages with self-labeled sentiment, respectively.
*** 1%, ** 5%, * 10% significance level

Table A2: How Common is the Social Signal across User Types on StockTwits?
Full List of Principal Components

Panel A: PCA of Attention Signals

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9
StockTwits	0.374	-0.141	0.108	-0.093	0.008	-0.077	-0.202	-0.202	-0.856
Top 1%	0.329	-0.086	-0.362	0.723	-0.002	0.472	-0.081	0.021	0.005
Professional	0.353	0.009	-0.312	0.216	-0.066	-0.745	0.389	-0.110	0.091
Intermediate	0.368	-0.095	0.037	-0.084	0.136	-0.243	-0.558	0.650	0.191
Novice	0.347	-0.210	0.220	-0.163	0.663	0.223	0.506	0.096	0.076
No experience	0.366	-0.144	0.193	-0.130	-0.032	0.030	-0.350	-0.674	0.461
Self-labeled	0.348	-0.172	0.237	-0.178	-0.732	0.223	0.334	0.245	0.066
Twitter	0.273	0.514	-0.586	-0.511	0.010	0.237	0.003	-0.023	-0.001
Seeking Alpha	0.203	0.778	0.523	0.280	0.023	-0.024	0.026	0.016	-0.012
Fraction of variation	76.6%	10.6%	4.7%	3.4%	1.8%	1.4%	0.9%	0.6%	0.1%

Panel B: PCA of Sentiment Signals

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9
StockTwits	0.595	-0.121	0.063	-0.048	-0.056	0.007	0.001	-0.262	-0.744
Top 1%	0.255	0.327	-0.542	0.075	0.064	0.148	-0.668	0.236	0.013
Professional	0.262	0.383	-0.520	-0.042	-0.270	-0.051	0.627	-0.122	0.170
Intermediate	0.270	-0.043	-0.045	-0.160	0.889	0.151	0.188	-0.098	0.199
Novice	0.184	0.024	0.140	0.956	0.063	0.074	0.078	-0.064	0.112
No experience	0.502	-0.239	0.212	-0.171	-0.293	-0.040	-0.254	-0.320	0.602
Self-labeled	0.362	-0.248	0.122	-0.044	-0.091	0.035	0.215	0.856	0.049
Twitter	0.144	0.514	0.330	-0.027	0.148	-0.752	-0.080	0.106	0.005
Seeking Alpha	0.052	0.589	0.491	-0.139	-0.095	0.615	0.031	0.028	-0.000
Fraction of variation	26.6%	12.0%	11.1%	10.7%	10.3%	10.2%	9%	8.4%	1.5%

Note: This table reports the principal component analyses of social signals across different user types on StockTwits and other platforms. Panels A and B focus on attention and sentiment signals, respectively. Sample consists of firm-day observations with at least 10 messages on StockTwits.

Table A3: How Do Next-Day Returns and Retail Trade Relate to Social Signal Bins?

	(1) CAR t+1	(2) Retail trade imbalance t
Mid sentiment × Low attention	0.090***	0.534***
	(0.031)	(0.133)
High sentiment × Low attention	0.094***	1.633***
	(0.035)	(0.150)
Low sentiment × Mid attention	-0.080**	1.243***
	(0.033)	(0.140)
Mid sentiment × Mid attention	-0.027	1.727***
	(0.034)	(0.143)
High sentiment × Mid attention	0.016	3.024***
	(0.038)	(0.165)
Low sentiment × High attention	-0.255***	1.913***
	(0.045)	(0.174)
Mid sentiment × High attention	-0.282***	2.702***
	(0.048)	(0.168)
High sentiment × High attention	-0.001	3.930***
	(0.055)	(0.193)
DJNW sentiment (z)	0.081***	0.121***
	(0.008)	(0.023)
DJNW attention (z)	0.023**	0.008
	(0.009)	(0.056)
Lagged signals	Y	Y
Controls	Y	Y
Date FE	Y	Y
Outcome Mean	-0.048	-0.329
Outcome SD	7.129	23.174
Observations	814,646	810,484
R^2	0.029	0.022
Root MSE	7.035	22.949

Note: This table reports how next-day returns and retail trading imbalance relate to social signal bins defined by terciles of sentiment first PC from Table 3 panel B and terciles of attention first PC from Table 3 panel A. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome is CAR t+1 scaled by 100 in column (1) and retail trading imbalance scaled by 100 in column (2). Controls are 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for date fixed effects and lagged 1-10 sentiment and attention. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table A4: How Do Returns Relate to Social Signals? (CAR t+1)

	Dep. variable: CAR t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ST sentiment (z)		0.047***			0.047***		
		(0.012)			(0.014)		
ST attention (z)		-0.143***			-0.182***		
		(0.052)			(0.060)		
ST sentiment (z) × ST attention (z)		0.019			0.039		
		(0.050)			(0.062)		
Twitter sentiment (z)			0.027***		0.015**		
			(0.007)		(0.007)		
Twitter attention (z)			-0.017		0.107***		
			(0.020)		(0.033)		
Twitter sentiment (z) × Twitter attention (z)			-0.006		-0.020		
			(0.008)		(0.014)		
SA sentiment (z)				0.084***	0.079***		
				(0.010)	(0.010)		
SA attention (z)				-0.028**	-0.022*		
				(0.012)	(0.012)		
SA sentiment (z) × SA attention (z)				0.001	0.002		
				(0.004)	(0.004)		
Sentiment PC1 (z)						0.051***	0.051***
						(0.010)	(0.010)
Sentiment PC2 (z)							0.071***
							(0.009)
Sentiment PC3 (z)							0.030***
							(0.008)
Attention PC1 (z)						-0.139***	-0.150***
						(0.050)	(0.053)
Sentiment PC1 × Attention PC1 (z)						-0.013	-0.009
						(0.027)	(0.029)
Sentiment PC2 × Attention PC1 (z)							0.067***
							(0.021)
Sentiment PC3 × Attention PC1 (z)							-0.078***
							(0.028)
DJNW sentiment (z)	0.082***	0.079***	0.080***	0.068***	0.065***	0.079***	0.068***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
DJNW attention (z)	0.005	0.022**	0.003	0.001	-0.010	0.025**	0.016
	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
8-K report date	0.101**	0.119***	0.108**	0.105**	0.081*	0.130***	0.118***
	(0.041)	(0.042)	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)
EA date	-0.575***	-0.585***	-0.562***	-0.588***	-0.561***	-0.579***	-0.590***
	(0.090)	(0.091)	(0.091)	(0.091)	(0.091)	(0.091)	(0.090)
Retail trade imbalance t-1∼ t-5	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Retail trade imbalance t-6∼ t-30	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Volatility t-1∼ t-5	-1.173***	-1.179***	-1.112***	-1.139***	-1.064***	-1.182***	-1.084***
	(0.314)	(0.317)	(0.315)	(0.314)	(0.320)	(0.316)	(0.317)
CAR t-1∼ t-5	-0.004*	-0.003*	-0.004**	-0.004**	-0.003*	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CAR t-6∼ t-30	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lagged signals	N	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	-0.048	-0.048	-0.048	-0.048	-0.048	-0.048	-0.048
Outcome SD	7.127	7.129	7.129	7.129	7.129	7.129	7.129
Observations	815,322	814,646	814,646	814,646	814,646	814,646	814,646
R^2	0.0291	0.0293	0.0292	0.0293	0.0296	0.0293	0.0294
Within R^2	0.0007	0.0009	0.0008	0.0009	0.0012	0.0009	0.0011
Root MSE	7.0333	7.0350	7.0356	7.0352	7.0344	7.0351	7.0347

Note: This table reports how next-day returns relate to social signals. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome is CAR t+1 scaled by 100. All regressions control for date fixed effects and lagged 1-10 sentiment and attention. Standard errors are clustered by firm and by date. *** 1%, ** 5%, * 10% significance level

Table A5: How Do Returns Relate to Social Signals? (CAR t+1∼t+20)

	Dep. variable: CAR t+1∼t+20						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ST sentiment (z)		0.087			0.094		
		(0.100)			(0.118)		
ST attention (z)		-0.868***			-0.923***		
		(0.291)			(0.298)		
ST sentiment (z) × ST attention (z)		-0.149			-0.067		
		(0.510)			(0.611)		
Twitter sentiment (z)			0.124***		0.083***		
			(0.031)		(0.031)		
Twitter attention (z)			-0.225**		0.282**		
			(0.088)		(0.138)		
Twitter sentiment (z) × Twitter attention (z)			-0.083*		-0.084		
			(0.046)		(0.112)		
SA sentiment (z)				0.114***	0.068**		
				(0.028)	(0.030)		
SA attention (z)				-0.159***	-0.074**		
				(0.042)	(0.035)		
SA sentiment (z) × SA attention (z)				-0.018	-0.005		
				(0.013)	(0.013)		
Sentiment PC1 (z)						0.078	0.058
						(0.059)	(0.064)
Sentiment PC2 (z)							0.129***
							(0.038)
Sentiment PC3 (z)							0.035
							(0.039)
Attention PC1 (z)						-0.890***	-0.996***
						(0.267)	(0.316)
Sentiment PC1 × Attention PC1 (z)						-0.215	-0.260
						(0.290)	(0.313)
Sentiment PC2 × Attention PC1 (z)							0.393**
							(0.181)
Sentiment PC3 × Attention PC1 (z)							-0.213**
							(0.107)
DJNW sentiment (z)	0.134***	0.126***	0.128***	0.108***	0.107***	0.133***	0.098***
	(0.033)	(0.034)	(0.033)	(0.032)	(0.033)	(0.035)	(0.033)
DJNW attention (z)	0.197***	0.294***	0.131***	0.163***	0.062	0.315***	0.235***
	(0.043)	(0.058)	(0.043)	(0.042)	(0.050)	(0.062)	(0.052)
8-K report date	0.788***	0.847***	0.935***	0.830***	0.782***	0.934***	0.817***
	(0.212)	(0.222)	(0.220)	(0.211)	(0.227)	(0.219)	(0.217)
EA date	-0.578***	-0.644***	-0.326	-0.375*	-0.133	-0.571**	-0.491**
	(0.217)	(0.232)	(0.222)	(0.218)	(0.225)	(0.230)	(0.224)
Retail trade imbalance t-1∼ t-5	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Retail trade imbalance t-6∼ t-30	0.044**	0.044**	0.043**	0.043**	0.045**	0.043**	0.044**
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Volatility t-1∼ t-5	-11.212***	-11.030***	-10.357***	-10.877***	-9.417***	-11.121***	-10.135***
	(2.365)	(2.387)	(2.356)	(2.357)	(2.379)	(2.388)	(2.335)
CAR t-1∼ t-5	-0.009	-0.009	-0.011	-0.010	-0.012	-0.009	-0.010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
CAR t-6∼ t-30	0.003	0.004	0.003	0.003	0.004	0.003	0.003
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Lagged signals	N	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	-0.564	-0.569	-0.569	-0.569	-0.569	-0.569	-0.569
Outcome SD	30.167	30.177	30.177	30.177	30.177	30.177	30.177
Observations	801,538	800,862	800,862	800,862	800,862	800,862	800,862
R^2	0.0671	0.0674	0.0674	0.0672	0.0680	0.0674	0.0676
Within R^2	0.0013	0.0016	0.0016	0.0014	0.0022	0.0016	0.0018
Root MSE	29.1831	29.1878	29.1885	29.1917	29.1795	29.1887	29.1852

Note: This table reports how 20-day cumulative returns relate to social signals. Sample consists of firm-day observations with at least 10 messages on StockTwits. The outcome is CAR t+1×t+20 scaled by 100. All regressions control for date fixed effects and lagged 1-10 sentiment and attention. Standard errors are clustered by firm and by date. *** 1%, ** 5%, * 10% significance level

Table A6: How Did Informativeness of StockTwits Signals Change around StockTwits Character Limit Increase?
Firms with Long versus Short Messages

	(1) CAR t+1	(2) Retail imbalance
Post × Treated × Sentiment (z)	0.186**	0.069
	(0.082)	(0.320)
Post × Treated × Attention (z)	-0.558*	1.865**
	(0.308)	(0.781)
Post × Sentiment (z)	-0.045	-0.096
	(0.071)	(0.235)
Post × Attention (z)	0.291	-0.391
	(0.224)	(0.368)
Treated × Sentiment (z)	-0.021	-0.627***
	(0.053)	(0.226)
Treated × Attention (z)	0.265	-0.718
	(0.218)	(0.596)
Sentiment (z)	0.035	1.026***
	(0.049)	(0.174)
Attention (z)	-0.605***	1.291***
	(0.178)	(0.172)
Post × Treated	0.086	0.393
	(0.119)	(0.398)
Treated	-0.034	-1.570***
	(0.081)	(0.285)
Firm FE	Y	Y
Date FE	Y	Y
Outcome Mean	-0.097	-0.408
Outcome SD	8.338	25.356
Observations	107,201	106,389
R^2	0.034	0.039

Note: This table compares how social signals about firms with long versus short StockTwits messages changed their predictive power for CAR t+1 and retail trading imbalance before versus after StockTwits character limit increase on May 8, 2019. Sample consists of firm-day observations with at least 10 messages on StockTwits between May 8, 2018 and May 8, 2020. The outcome is CAR t+1 scaled by 100 in column (1) and retail trading imbalance scaled by 100 in column (2). *Treated* is one if a firm's daily average number of characters per message is in the top quartile; the omitted category is those is in the bottom quartile. *Post* is one if a day is on or after May 8, 2019. Controls are DJNW sentiment and attention, 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table A7: How Did Informativeness of Components of Retail Trading Imbalance Change around StockTwits Character Limit Increase?
Firms with Long versus Short Messages

	Dep. variable: CAR t+1	
	(1)	(2)
Post × Treated × Retail trading imbalance (z)	0.184*	
	(0.106)	
Post × Retail trading imbalance (z)	-0.086	
	(0.084)	
Treated × Retail trading imbalance (z)	-0.013	
	(0.076)	
Retail trading imbalance (z)	0.035	
	(0.060)	
Post × Treated × Imbalance from sentiment (z)		0.216**
		(0.085)
Post × Treated × Imbalance from mid attention (z)		-0.013
		(0.105)
Post × Treated × Imbalance from high attention (z)		0.099
		(0.129)
Post × Treated × Residual imbalance (z)		0.191*
		(0.109)
Post × Imbalance from sentiment (z)		-0.056
		(0.073)
Post × Imbalance from mid attention (z)		-0.030
		(0.087)
Post × Imbalance from high attention (z)		-0.104
		(0.124)
Post × Residual imbalance (z)		-0.088
		(0.085)
Treated × Imbalance from sentiment (z)		-0.050
		(0.055)
Treated × Imbalance from mid attention (z)		0.092
		(0.060)
Treated × Imbalance from high attention (z)		0.353***
		(0.084)
Treated × Residual imbalance (z)		-0.039
		(0.077)
Imbalance from sentiment (z)		0.065
		(0.052)
Imbalance from mid attention (z)		-0.068
		(0.054)
Imbalance from high attention (z)		-0.436***
		(0.083)
Residual imbalance (z)		0.062
		(0.060)
Post × Treated		0.162
		(0.116)
Treated		-0.010
		(0.072)
Firm FE	Y	Y
Date FE	Y	Y
Outcome Mean	-0.092	-0.092
Outcome SD	8.341	8.341
Observations	106,367	106,367
R^2	0.033	0.034

Note: This table compares how components of retail trading imbalance predicted by StockTwits signals about firms with long versus short messages changed their predictive power for CAR t+1 before versus after StockTwits character limit increase on May 8, 2019. Sample consists of firm-day observations with at least 10 messages on StockTwits between May 8, 2018 and May 8, 2020. The outcome is CAR t+1 scaled by 100. *Post* is one if a day is on or after May 8, 2019. *Treated* is one if a firm's daily average number of characters per message is in the top quartile; the omitted category is firms whose daily average number of characters per message is in the bottom quartile. Controls are DJNW sentiment and attention, 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), volatility from the previous five days, lagged cumulative abnormal returns (previous five trading days and the 25 days before that), and DJNW coverage and sentiment. All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table A8: How Did the Informativeness of Components of Retail Trade for Next-Day Returns change Around the GameStop Event?

	Dep. variable: CAR t+1			
	(1) Retail imbalance	(2) PC signal	(3) StockTwits old	(4) StockTwits new
Post × Retail trading imbalance (z)	0.015			
	(0.028)			
Retail trading imbalance (z)	0.033			
	(0.024)			
Post × Imbalance from sentiment (z)		-0.107**	0.002	-0.108***
		(0.047)	(0.034)	(0.042)
Post × Imbalance from mid attention (z)		-0.007	-0.013	-0.084**
		(0.035)	(0.033)	(0.037)
Post × Imbalance from high attention (z)		-0.007	-0.053	-0.109*
		(0.054)	(0.058)	(0.065)
Post × Residual imbalance (z)		0.017	0.017	0.018
		(0.028)	(0.028)	(0.028)
Imbalance from sentiment (z)		0.109***	0.039	0.103***
		(0.041)	(0.028)	(0.037)
Imbalance from mid attention (z)		-0.105***	-0.047	0.010
		(0.032)	(0.029)	(0.028)
Imbalance from high attention (z)		-0.282***	-0.210***	-0.180***
		(0.050)	(0.052)	(0.052)
Residual imbalance (z)		0.039	0.039	0.036
		(0.024)	(0.024)	(0.024)
DJNW sentiment (z)	0.086***	0.096***	0.090***	0.089***
	(0.017)	(0.017)	(0.017)	(0.017)
DJNW attention (z)	-0.075**	-0.038	-0.048	-0.051*
	(0.031)	(0.031)	(0.030)	(0.030)
Controls	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Outcome Mean	-0.003	-0.003	-0.003	-0.003
Outcome SD	7.872	7.872	7.872	7.872
Observations	286,308	286,308	286,308	286,308
R^2	0.049	0.050	0.049	0.050

Note: This table reports how the informativeness of components of retail trading imbalance for next-day returns changed before versus after the GameStop event on January 28, 2021. Sample consists of firm-day observations with at least 10 messages on StockTwits between February 1, 2020 and December 31, 2021, excluding January 2021. The outcome is CAR t+1 scaled by 100. *Post* is one if a day is on or after February 1, 2021. The components come from an auxiliary regression of retail trading imbalance on all social media sentiment signals and terciles of the first PC of attention in column 2 (PC signal), from a regression on sentiment and terciles of attention based on messages from users who joined StockTwits before 2020 in column 3 (StockTwits old), and from a regression on sentiment and terciles of attention based on messages from users who joined StockTwits in or after 2020 in column 4 (StockTwits new). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table A9: How Did the Informativeness of the Social Signal for Next-Day Returns and Retail Trade Change around GameStop Event?

Signal from StockTwits Users Who Joined before versus after January 2020

	(1) CAR t+1	(2) Retail imbalance t
Post × New user × Sentiment (z)	-0.095*	-0.021
	(0.049)	(0.133)
Post × New user × Attention (z)	0.038	-0.246**
	(0.032)	(0.114)
New user × Sentiment (z)	0.056	-0.214**
	(0.041)	(0.103)
New user × Attention (z)	-0.017	0.130
	(0.031)	(0.160)
Post × Sentiment (z)	-0.000	-0.030
	(0.032)	(0.102)
Post × Attention (z)	-0.020	-0.113
	(0.093)	(0.240)
Sentiment (z)	0.034	0.465***
	(0.026)	(0.068)
Attention (z)	-0.057	0.437*
	(0.048)	(0.236)
Post × New user	-0.024**	-0.172***
	(0.010)	(0.028)
New user	0.024***	0.098***
	(0.009)	(0.023)
DJNW sentiment (z)	0.087***	0.221***
	(0.017)	(0.034)
DJNW attention (z)	-0.069**	0.199***
	(0.031)	(0.052)
Controls	Y	Y
Firm FE	Y	Y
Date FE	Y	Y
Outcome Mean	-0.004	-1.390
Outcome SD	7.867	22.388
Observations	575,666	572,698
R^2	0.049	0.028

Note: This table compares how social signals from new versus old StockTwits users changed their predictive power for CAR t+1 and retail trading imbalance before versus after the GameStop event on January 28, 2021. Sample consists of firm-day observations with at least 10 messages on StockTwits between February 1,2020 and December,31,2021, excluding January 2021. The outcome is CAR t+1 scaled by 100 in column (1) and retail trading imbalance scaled by 100 in column (2). *Post* is one if a day is on or after February 1, 2021. *New user* is one if the social signals are from users who joined StockTwits in 2020 or 2021; the comparison group is the social signals from users who joined before 2020. Controls are DJNW sentiment and attention, 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level

Table A10: How Did the Informativeness of Components of Retail Trade for Next-Day Returns Change Around the GameStop Event?
Signal from StockTwits Users Who Joined before versus after January 2020

	(1) CAR t+1
Post × New user × Imbalance from sentiment (z)	-0.108**
	(0.048)
Post × New user × Imbalance from mid attention (z)	-0.064
	(0.039)
Post × New user × Imbalance from high attention (z)	-0.046
	(0.030)
Post × New user × Residual imbalance (z)	-0.001
	(0.002)
New user × Imbalance from sentiment (z)	0.064
	(0.040)
New user × Imbalance from mid attention (z)	0.053
	(0.033)
New user × Imbalance from high attention (z)	0.027
	(0.022)
New user × Residual imbalance (z)	-0.002
	(0.002)
Post × Imbalance from sentiment (z)	0.004
	(0.032)
Post × Imbalance from mid attention (z)	-0.014
	(0.032)
Post × Imbalance from high attention (z)	-0.055
	(0.057)
Post × Residual imbalance (z)	0.017
	(0.028)
Imbalance from sentiment (z)	0.039
	(0.026)
Imbalance from mid attention (z)	-0.041
	(0.028)
Imbalance from high attention (z)	-0.199***
	(0.050)
Residual imbalance (z)	0.039
	(0.024)
Post × New user	-0.020*
	(0.010)
New user	0.022**
	(0.009)
Controls	Y
Firm FE	Y
Date FE	Y
Outcome Mean	-0.003
Outcome SD	7.872
Observations	572,616
R^2	0.049

Note: This table reports the change in the informativeness for next-day returns of components of retail trading imbalance predicted by social signals from new versus old StockTwits users changed before versus after the GameStop event on January 28, 2021. Sample consists of firm-day observations with at least 10 messages on StockTwits between February 1, 2020 and December 31, 2021, excluding January 2021. The outcome is CAR t+1 scaled by 100. *Post* is one if a day is on or after February 1, 2021. *New user* is one if the social signals are from users who joined StockTwits in 2020 or 2021; the comparison group is the social signals from users who joined before 2020. Controls are 8-K report date indicators, earnings announcement indicators, lagged retail order imbalance (previous five trading days and the 25 days before that), and lagged cumulative abnormal returns (previous five trading days and the 25 days before that). All regressions control for firm fixed effects and date fixed effects. Standard errors are clustered by firm and by date.
*** 1%, ** 5%, * 10% significance level