Market Signals from Social Media

J. Anthony Cookson — *CU Boulder*
Runjing Lu — *U. of Toronto*
William Mullins — *UC San Diego*
Marina Niessner* — *Indiana U.*

Preliminary Draft

January 27, 2025

Abstract

This paper develops daily market-wide sentiment and attention indexes using the content of millions of posts from major investor social media platforms. Our daily indexes predict market-wide returns and trading over the next 20 days, and the dynamics of sentiment differ from those of attention. A standard deviation increase in sentiment predicts a 37 basis point lower 20-day return for the S&P500 index, reflecting a reversal of the market run-up that precedes high sentiment. Greater attention also predicts negative returns, but as a continuation of previous trends. A dynamic trading strategy based on these signals delivers a Sharpe ratio of 1.2. The indexes also predict aggregate trading activity in distinct ways: abnormal trading is greater when sentiment is low and attention is high, consistent with bad news driving increases in aggregate trading. Moreover, we find that sentiment responds strongly to recent returns, but this impact is driven primarily by negative market jumps.

Keywords: Sentiment, Attention, Market-wide Signals, Social Media
JEL: G12, E71, G41

*Cookson (tony.cookson@colorado.edu); Lu (runjing.lu@utoronto.ca); Mullins (wmullins@ucsd.edu) and Niessner (mniessne@iu.edu). We thank Zhi Da, Shaun Davies, Charles Martineau, Lin Peng for providing attention data and valuable comments. Any remaining errors are our own.

> [A] large proportion of our positive activities depend on spontaneous optimism rather than on a mathematical expectation, whether moral or hedonistic or economic.
>
> — John Maynard Keynes in "General Theory Of Employment, Interest And Money" (1936, p. 144).

1. Introduction

How to characterize investor sentiment has been a major question in financial markets at least since Keynes (1936). In recent decades, research has explored how investor sentiment and aggregate beliefs are formed (Baker and Wurgler, 2006, Bordalo et al., 2024), spurring the development of models of extrapolation, diagnostic expectations, and memory among others (e.g., Bordalo et al., 2018, 2020). Although this work has had important implications for asset pricing (Barberis et al., 2018, Greenwood et al., 2019), most of this research is rooted in monthly aggregate return patterns. Shorter-term dynamics have received less attention — an important gap given that investor sentiment is increasingly expressed on social media and is subject to change at high frequency (e.g., Cookson et al., 2024b).

This paper develops daily attention and sentiment indexes drawn from millions of posts on three major investor social media platforms: StockTwits, Twitter, and Seeking Alpha. The social media setting clearly separates sentiment from attention, which is important because these concepts are economically distinct. Consistent with this distinction, we find returns rise prior to high *sentiment* days, followed by a reversal over the next 20 days. By contrast, returns decline prior to high *attention* days, followed by a continuation of negative returns. Beyond these patterns, we find that these signals contain unique return-relevant information, yielding a Sharpe ratio of 1.2 in a dynamic trading strategy. Moreover, there are important differences between the drivers of the sentiment and attention indexes: spikes in lagged *returns* predict sentiment, while increases in lagged abnormal *trading volume* predict attention. These findings highlight the importance of understanding these higher frequency patterns, as well as the distinction between sentiment and attention.

Our analysis begins with detailed data spanning a decade of *stock-specific* social media posts. We first residualize firm-day social media sentiment and attention signals by projecting them onto firm-level lagged average sentiment and attention, plus a rich set of controls for traditional news. Using these firm-day residuals, which have been stripped of lagged firm-level and news-driven components, we construct daily market sentiment and attention signals by (i) calculating the market-capitalization weighted average within each platform, and then (ii) combining them across platforms via principal component analysis. This procedure yields two daily indexes from 2013 through 2021: one for sentiment and one for attention. Although these indexes display novel daily patterns, they also reflect major and persistent market episodes, such as the onset of the Covid-19 pandemic, which saw sustained increases in attention and declines in sentiment.

With the market-level sentiment and attention indexes in hand, we then examine their relation to subsequent market returns and aggregate turnover. Both sentiment and attention predict negative returns over a 20-day window, but for different reasons. Negative returns after high sentiment reflect a reversal of a run-up in returns in the prior five days. By contrast, negative returns after high attention are a continuation of previously low returns. Critically, this pattern holds even when we include year-month fixed effects, which absorb the vast majority of existing sentiment indexes that vary at the year-month level.

To investigate the economic significance of this return predictability, we implement a dynamic trading strategy that determines the portfolio weight on risky assets on day $t + 1$ using the values of the attention and sentiment indexes on day t, following the approach in Campbell and Thompson (2008). This strategy generates portfolio excess returns averaging 4.6% over our 2013-2021 sample (50% cumulatively), with a Sharpe ratio of 1.22. Further, two-thirds of these portfolio excess returns are abnormal with respect to Fama and French (1993) and Carhart (1997) risk factors. This performance compares favorably to other daily market-level signals (e.g., Da et al., 2024) as well as the historical Sharpe ratio of the market (Bodie et al., 2011, Mehra and Prescott, 1985). Interestingly, portfolio excess returns are two-thirds larger after days when the market declines by 1%, indicating that the social media strategy is especially profitable on days when the market performs poorly. Imposing long-only, no leverage constraints (i.e., an allocation to the market index between 0% and 100%) leads to only minimal deterioration in portfolio performance, indicating that the returns are not driven by leverage or by going short. We also show that the portfolio returns cannot be replicated by a factor rotation strategy using the Fama-French or momentum factors. Collectively, these findings highlight the informativeness of the social media indexes.

Turning to market turnover, we estimate that aggregate abnormal turnover increases when social media attention is *high* and sentiment is *low*. These results are consistent with the idea that high attention and low sentiment typically occur when there is aggregate market stress. Further, our sentiment results hold controlling for the attention index, and vice versa, indicating that each index contains distinct information — consistent with differing return and trading dynamics for sentiment and attention. These market-wide trading patterns also hold after controlling for year-month fixed effects, as well as daily controls for abnormal Google search volume, Bloomberg attention, and coverage in the *New York Times* and the *Wall Street Journal*.

We conclude the paper by investigating the drivers of these sentiment and attention indexes. In both OLS regressions and in vector autoregressions — which account for joint dynamics of sentiment, attention, returns, and trading— we find that sentiment is predicted by lagged returns, while attention is predicted instead by lagged trading. Additionally, we examine how the sentiment and attention indexes evolve around sharp changes in the S&P500 index and the VIX. This analysis reveals a striking asymmetry: neither sentiment nor attention respond to positive market jumps, but following downward market jumps and spikes in the VIX, sentiment decreases while attention increases. These results are consistent with daily extrapolation (Barberis et al., 2018), and also reveal an asymmetry in how market-wide sentiment updates in response to market signals, resembling sentiment updates by journalists to recent market movements (e.g., Garcia, 2013).

Contributions to Related Literature. Our research makes several contributions to the economics and finance literature on sentiment, belief formation, and social media.

The monthly sentiment index in Baker and Wurgler (2006) began an empirical literature focused on understanding aggregate sentiment and the creation of a number of alternative sentiment indexes, primarily at the monthly level (e.g., Huang et al., 2015, DeVault et al., 2019, Jiang et al., 2019, Davies, 2022, Henderson et al., 2023). Our research departs from this literature by focusing on higher frequency patterns derived from social media. Specifically, our tests focus on the daily level, and by including year-month fixed effects, we show that our findings are entirely driven by higher frequency variation.

Within this broader literature on sentiment, a closely related idea is to capture household concerns by measuring daily search volume of negative search terms like "bankruptcy" or "recession" (Da et al., 2015) or by counting daily mentions of economic uncertainty terms (Baker et al., 2016). Our sentiment index is distinct from these ideas because it is built from variation in bullish sentiment about stocks, not attention to negative outcomes or the use of uncertainty terms. Moreover, we show that our sentiment index is distinct from mentions of economic uncertainty on Twitter (Baker et al., 2021). Moreover, we also introduce a social media-based daily aggregate attention index. This is an important contribution because some of the existing sentiment indexes, including daily uncertainty indexes, are a combination of sentiment and attention; the distinct dynamics of sentiment and attention in our results highlight the importance of separating them.

Our research contributes to the literature on investor social media (Chen et al., 2014, Avery et al., 2016). Recent work shows that social media signals can have firm-level predictive power (Cookson et al., 2024a, Dim, 2020), but also that social signals may be shared in a way that generates biases (Cookson, Engelberg, and Mullins, 2023a, Chen and Hwang, 2022, Cassella, Dim, and Karimli, 2023, Chen, Peng, and Zhou, 2024, Hirshleifer, Peng, and Wang, 2024). By extracting market-wide sentiment and attention signals from firm-specific posts on social media, this paper also contributes to the literature that uses social media as a lens to study broader economic phenomena (Bailey, Cao, Kuchler, Stroebel, and

Wong, 2018, Cookson and Niessner, 2020, Cookson, Mullins, and Niessner, 2024b). The focus tends to be at the firm level in this literature, in contrast to our market-level indexes of sentiment and attention.[1] Recent work has constructed related market-level attention indexes from Google searches, Bloomberg activity, and news articles (Fisher, Martineau, and Sheng, 2022, Da, Hua, Hung, and Peng, 2024). However, our indexes capture distinct and complementary information: our main findings hold when controlling for these alternative attention measures, likely because the information in our indexes is derived from social media.

Finally, our indexes and findings are also relevant to the recent literature on aggregate belief formation (e.g., Bordalo et al., 2018, Barberis et al., 2018, Bordalo et al., 2020), especially because sentiment in our context is an aggregate investor-contributed measure of investor beliefs. Consistent with recent models and evidence on extrapolation in other settings and at other frequencies (Da and Huang, 2020, Da et al., 2021), we find that sentiment is extrapolative in that it exhibits a strong connection with recent lagged returns. Additionally, our analysis of jumps highlights the asymmetry of this daily return extrapolation: sharp negative jumps drive sentiment and attention, but sharp positive jumps bear no relation to our market signals. Beyond showing that sentiment is extrapolative at a different frequency, the social media setting draws a more immediate connection to the beliefs of retail traders in shaping this relationship. The connection between retail investors and aggregate sentiment represented by the index may also shed light on the causes and consequences of trading frenzies connected to social media (Bradley et al., 2024, Cookson et al., 2023b).

2. Data

In this section, we describe our data sources and the construction of the aggregate sentiment and attention indexes.

[1] Cookson, Engelberg, and Mullins (2020) develops an aggregate sentiment index from posts on StockTwits around the onset of the Covid-19 pandemic. However, this research is narrowly focused on partisan differences in investor beliefs, not the content of the market signal.

2.1 Firm-day social media sentiment and attention data

Our data contain firm-day observations on social media sentiment (optimism versus pessimism) and attention across three major investor social media platforms: Twitter, StockTwits, and SeekingAlpha. The underlying data are at the message level for StockTwits, article level for Seeking Alpha, and firm-day level for Twitter. These data are the same as the sources in Cookson, Lu, Mullins, and Niessner (2024a); we obtain Seeking alpha data from Ravenpack 1.0 and Twitter data—including average sentiment and number of messages per firm-day—from Context Analytics.

To construct the firm-day datasets, we make the following choices. For each platform, we construct close-to-close measures of firm-day attention and sentiment. To ensure accurate measurement, we include only StockTwits posts that reference a single ticker (via a "cashtag," a dollar sign ($) followed by a ticker symbol) and Seeking Alpha articles with a relevance score to a specific ticker above 75 (i.e., "significantly relevant"). We use Ravenpack's Event Sentiment Score (ESS) to measure Seeking Alpha sentiment. To avoid posts by bots, we drop users making over 1,000 posts in any day.

To measure sentiment about firm i on day t for the post- and article-level data (StockTwits and Seeking Alpha), we average sentiment over all posts or articles about the firm from 4:00 pm (close) on day $t-1$ to 4:00 pm on day t. The resulting firm-day sentiment measure is comparable to the Twitter firm-day sentiment measure provided by Context Analytics. 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 over the same period.

2.2 Other data

Our data on firm-related news events covered in traditional news is from the *Dow Jones Newswire*. Ravenpack 1.0 provides article-level sentiment from these sources and the number of articles by firm-day. We retain articles with a ticker-specific relevance score above 75. To measure firm-day level sentiment, we average the article-level Ravenpack ESS of all relevant articles by firm-day. For firm-specific news we use 8-K filing dates from the SEC Analytics Suite by WRDS and earnings announcement dates from IBES.

2.3 Sample

As in Cookson et al. (2024a), we focus on the 1,500 most-discussed firms on StockTwits and require firm-day observations to have at least 10 posts on StockTwits. Table 1 presents the attributes of the data. While the three platforms provide similarly good coverage of firms, Seeking Alpha has a substantially lower number of firm-day observations, reflecting its less frequent coverage of the same firms. StockTwits contains more than five times the number of daily posts than Twitter. In turn, Seeking Alpha's number of posts per day is substantially lower than the other two platforms, likely due to its long-form nature.

Each day in our data, StockTwits and Twitter cover a large number of firms, comprising around 90% of the market capitalization in the sample. After restricting our sample to firm-day observations with at least ten posts on StockTwits, ensuring a high quality social signal, coverage remains strong at around 50% of the market capitalization. Seeking Alpha maintains less breadth but still covers around 30% of the market capitalization.

2.4 Constructing Aggregate Indexes from Firm-day information

To construct daily indexes of *aggregate* sentiment and *aggregate* attention from social media, we employ a three-step process. Starting with firm-day data on attention and sentiment across three platforms, we first remove firm-specific news events and slow-moving attention or sentiment averages from the firm-day signals . Next we use the resulting residuals to create value-weighted averages for each platform-day. Finally, we combine the respective platform-day signals into aggregate sentiment and attention indexes using principal component analysis.

For the first step in this procedure, we remove the firm-specific news and slow-moving attention or sentiment from the firm-day signals, as many posts reflect reactions to firm-specific information, which is not relevant to the aggregate signals. To do this we run the following firm-day regressions separately for each platform:

$$Signal_{i,t}^{P} = \Gamma^{P} \cdot X_{i,t} + \beta \cdot \overline{Signal}_{i,-y}^{P} + \epsilon_{i,t}, \tag{1}$$

where $Signal_{i,t}^{P}$ is either attention or sentiment on a platform P for firm i on day t; $X_{i,t}$ are indicators for traditional news coverage, 8-K filings, and earnings announcements from day $t-7$ through day t for firm i. $\overline{Signal}_{i,-y}^{P}$ denotes the average signal on platform P for firm i in the previous calendar year, which controls for firm-specific averages in attention or sentiment without introducing look-ahead bias of firm fixed effects. The estimates from these regressions are reported in Table 2 panel A. Although the lagged signal and firm news controls are statistically significant, a substantial share of variation is unexplained by this firm-specific information, and thus is left in the residuals.

In the second step, we aggregate for each platform the residuals from Equation 1 across firms within day by calculating a value-weighted average of residuals.

For the final step, we combine the resulting platform-day signals into daily indexes of aggregate sentiment and attention, by performing two principal component analyses (PCAs): one using the sentiment signals and another using the attention signals from StockTwits, Twitter and Seeking Alpha. The first principal components (PC1) of sentiment and attention constitute our daily sentiment and attention indexes, respectively. These are reported in Table 2 panel B.

In this analysis, the PC1 of sentiment explains 47% of the variation in the three component sentiment signals, while the PC1 of attention explains 54% of the variation in attention. To put these in perspective, the PC1 would explain 33% of the variation if the three signals were completely uncorrelated. The sentiment and attention signals both place similar loadings on StockTwits and Twitter. For Seeking Alpha, the sentiment signal loading is about half the size as for the other platforms. By contrast, the attention signal places close to no weight on Seeking Alpha, likely because there are far fewer articles on Seeking Alpha relative to the number of posts on StockTwits and Twitter.

Figure 1 plots the sentiment and attention indexes over time. The lighter-colored lines in the background of the figure represent our daily indexes, while the dark lines show 20-day rolling averages of each series. We also plot the level of the S&P 500 index for reference. Focusing on the lower-frequency movements, the sentiment and attention indexes appear to reflect different economic forces, with a correlation of -0.37. This divergence is particularly evident during the 2013-2015 stock market bull run, in the 2018-2019 trade war with China, and around the onset of the COVID-19 pandemic. During the bull run, sentiment was high while attention was low; the converse happened during the other two (negative) events.

While the slower-moving signals are easier to visualize and highlight the two indexes' distinct information content, our paper focuses on the information in the *daily* movements in the indexes. As the graph highlights, there is substantial variation in these higher frequency series. In the following section, we analyze how daily fluctuations in sentiment and attention relate to daily market returns and trading activity.

3. Returns and Turnover Following Sentiment and Attention

In this section, we examine the return implication of our sentiment and attention indexes. As a preview, in Figure 2, we present coefficient estimates from regressions of cumulative returns on sentiment at day 0 for an event window from $t = -5$ to $t = +20$. The figure highlights distinct return dynamics around high sentiment (panel a) versus those around high attention (panel b). High sentiment on day 0 is preceded by a five-day return run-up and is followed by a gradual and nearly full reversal over the next 20 days. By contrast, for attention we find the *opposite* pattern: high attention on day 0 is preceded by negative returns, which continue on a downward trajectory in the following days.

The first subsection scrutinizes this evidence by studying how sentiment and attention indexes predict returns and market-wide trading. In the second subsection, we consider alternative drivers of returns and turnover. The third subsection implements a dynamic trading strategy to quantify the information content of these market-level social media signals. Section 4 then examines the drivers of these indexes.

3.1 Do social media indexes predict returns or turnover?

In this section, we examine how daily sentiment and attention indexes predict returns. To do so, we estimate the following regression specification:

$$\begin{aligned} \text{Market return}_{t\rightarrow k} &= \beta_1 Sentiment_t + \beta_2 Attention_t \\ &\quad + \beta_3 (Sentiment \times Attention)_t + \Lambda_t + \epsilon_t \end{aligned} \tag{2}$$

where Market return$_{t\rightarrow k}$ is the cumulative return between days t and $t+k$. $Sentiment_t$ and $Attention_t$ are the market-level indexes on day t. The Λ_t vector includes day-of-week, month-of-year, and year-quarter fixed effects to control for seasonality and slow-moving annual trends, as well as lagged market volatility (day $t-5$ through $t-1$) and lagged returns (day $t-5$ through $t-1$ and the previous 25 days). We also estimate the specification in Eq. 2 using turnover (S&P500 or SPY) as the dependent variable, additionally controlling for abnormal turnover in day $t-1$.[2]

Table 3 reports the estimates. Columns 1 and 2 present contemporaneous (day t) regressions, while columns 3-6 display regressions of future returns on day t sentiment and attention indexes. These results show that sentiment and attention exhibit distinct return dynamics. Sentiment is strongly and positively related to contemporaneous returns on day t, followed by a significant return reversal from day $t+1$ through day $t+20$ (columns 5 and 6). Figure 2, which presents the daily cumulative returns from day $t-5$, shows that this reversal flattens out around day $t+15$. The specifications in columns 2, 4, and 6 also include an interaction between sentiment and attention, which is positively and significantly related

[2]The attention and sentiment indexes are constructed using the 1,500 most-discussed stocks on StockTwits, rather than the components of the S&P500. We use the S&P500 index to capture the overall market's behavior and relate it to the market-wide signals we construct.

to day t returns with no reversal.

Table 4 investigates how day t sentiment and attention indexes predict abnormal turnover for S&P500 stocks in aggregate (Panel A) and for the SPY ETF (Panel B) using analogous specifications to Table 3. These results show that sentiment and attention indexes have opposite relationship with turnover compared to their relationship with returns. Specifically, high attention is contemporaneously related to *high* turnover, whereas high sentiment relates to *low* turnover. Moreover, the dynamics are distinct: following high attention, abnormal trading continues to increase, and following high sentiment abnormal trading volume continues to decrease. Figure 3, which presents the cumulative abnormal turnover starting from day $t-5$, shows that these turnover patterns flatten out by day $t+10$. Finally, the interaction between sentiment and attention indexes does not significantly predict abnormal turnover (see columns 2, 4, 6).

We also perform two robustness checks on these main results. First, we examine the relation of sentiment and attention indexes to *retail* turnover (Boehmer et al., 2021), finding similar patterns and dynamics to our main findings (Figure A2 and Table A2). Second, we examine robustness of our results to the data quality requirement that we retain only firm-day observations with at least 10 StockTwits messages. In A4 we loosen this restriction to 5 or more messages per firm-day and obtain very similar findings.

3.2 Accounting for Alternative Drivers of Return and Turnover

In this section, we conduct two sets of robustness checks on our main analysis.

First, in Appendix Table A3, we substitute month-of-year and year-quarter fixed effects with year-month fixed effects and continue to find similar results. Year-month fixed effects flexibly controls for slow moving factors that could jointly drive returns, sentiment, and attention – particularly existing sentiment indexes measured at the monthly frequency (e.g., the Baker and Wurgler, 2006 index and similar macroeconomic indexes). Finding similar results using the within year-month variation indicates that our sentiment and attention indexes contain distinct from existing alternatives in the literature.

Second, in Appendix Table A4, we control for daily attention indexes from recent literature. For example, Da et al. (2024) develop two daily value-weighted macro attention indexes: a retail index based on Google searches for tickers and an institutional index based on Bloomberg searches for tickers. Additionally, Fisher et al. (2022) build a daily macroeconomic news index using articles in the *New York Times* and the *Wall Street Journal*. More closely related to sentiment, Baker et al. (2021) develop a Twitter-based measure of economic policy uncertainty. As shown in the Appendix Table A4, our results are not sensitive to inclusion of these alternative proxies for attention. Furthermore, in the Appendix Figure A3 we replicate our results from Figures 2 and 3 with these additional controls, and find robust results. Taken together this evidence suggests that our indexes contain unique information not captured in Google searches, Bloomberg searches, or traditional news.

3.3 Trading Strategy

In this section, we implement a dynamic trading strategy based on the daily aggregate sentiment and attention indexes to gauge their economic relevance. First, we use information up to prior month to construct daily social media indexes for the current month. We then use these social media indexes to predict next-day returns in the current month. Finally, we construct the dynamic trading strategy using these return forecasts.

For each month m, we estimate a daily-level regression using data from the beginning of our sample through month $m-1$ following Welch and Goyal (2008):

$$\begin{aligned} r_{t+1} \quad = \quad & \beta_{1,m-1} Sentiment_t + \beta_{2,m-1} Attention_t \\ & + \beta_{3,m-1}(Sentiment \times Attention)_t + \gamma_{m-1}\Omega_t + \epsilon_t \end{aligned} \tag{3}$$

This specification follows Eq. 2, but focuses on next-day returns as the outcome variable and contains no fixed effects. These differences ensure that the predictions from this regression yield a tradeable signal. r_{t+1} is the excess market return measured as the S&P 500 return minus the risk-free rate, while Ω_t^m includes lagged market volatility (day t-5 through t-1) and lagged returns (day t-5 through t-1 and the previous 25 days). For each month m, we use data up to $m-1$ to obtain OLS estimates of $\beta_{1,m-1}$, $\beta_{2,m-1}$, $\beta_{3,m-1}$ and γ_{m-1} (loadings for Ω_t^{m-1}). We then predict next-day returns for each trading day in month m:

$$\begin{aligned} \hat{r}_{t+1} &= \hat{\beta}_{1,m-1} Sentiment_t + \hat{\beta}_{2,m-1} Attention_t \\ &\quad + \hat{\beta}_{3,m-1} (Sentiment \times Attention)_t + \hat{\gamma}_{m-1} \Omega_t \end{aligned} \tag{4}$$

This day $t+1$ forecast uses only information available through day t, preventing look-ahead bias and ensuring tradeability. We repeat this procedure monthly from February 2013 through December 2021 (for a total of 108 rolling regressions, forecasting 2,246 trading day returns).

Next, we construct portfolio weights on the risky asset as in Campbell and Thompson (2008):

$$w_t^{social} \equiv \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2} \tag{5}$$

where $\hat{r}_{t+1}$ is the out-of-sample forecast excess return using information through day t and $\hat{\sigma}_{t+1}^2$ is the variance of the daily return forecasts over the 20 days leading up to t. This strategy dynamically adjusts the risky asset allocation. We constrain w_t^{social} to be between -1 (100% short) and 2 (100% leverage). The portfolio return remains similar if we prohibit short selling and leverage by restricting w_t^{social} to lie between 0 and 1.

Figure 4 presents a graphical summary of the portfolio strategy. Panel (a) presents the buy-and-hold cumulative returns from the dynamic trading strategy separately for each of the 9 years of our sample. These cumulative annual returns range from a loss of one percentage point (2020) to a gain of nearly 10 percentage points (2013). Panel (b) presents the cumulative return plot from 2013 through 2021, showing a 50% gain over the full sample. In Panels (c) and (d), we construct return plots for one year following 100 randomly drawn start dates. Panel (c) presents all 100 paths, whereas Panel (d) shows the average return of those paths with a 90% confidence band. The strategy generates an average annualized excess return of approximately 4%, which is highly statistically different from zero.

Figure 5 displays the time series of portfolio weights of the risky asset. Panel (a) presents the baseline portfolio weights from Eq. (5), while Panel (b) presents their five-day moving average. The restriction that the portfolio weights must be between -1 and $+2$ rarely binds, as the strategy tends to produce an interior solution on all but a few extreme attention or sentiment days. Short selling occurs on 24% of days, while leverage is needed on only 3% of days.

Next, we evaluate whether *other* market signals explain the next-day portfolio returns with the following regression:

$$r_{t+1}^{P} = \alpha + \beta_m R_t^m + \beta_{smb} R_t^{smb} + \beta_{hml} R_t^{hml} + \beta_{mom} R_t^{mom} + \epsilon_{t+1} \tag{6}$$

where the outcome variable r_{t+1}^{P} is the date $t+1$ portfolio excess returns from allocating a weight of w_{signal} to the risky asset (i.e., the S&P500) and $1-w_{signal}$ to the risk-free asset. The regression controls for market excess returns (R_t^m), small minus big returns (R_t^{smb}), value minus growth returns (R_t^{hml}) and momentum returns (R_t^{mom}). These factor returns are observed on date t using daily data from Kenneth French's data library (Fama and French, 1993). This regression tests whether the return can be explained by a factor rotation using the three Fama-French factors, plus the momentum factor.

Table 5 presents estimates from Eq. (6) across different specifications with different combinations of factors for the factor rotation strategy. Column 1 is unconditional, column 2 includes market excess returns, column 3 includes the three original Fama-French factors, and column 4 additionally includes the momentum factor (Carhart, 1997). Panel A presents these estimates under our baseline restriction that the portfolio weight be in the range $[-1, +2]$, while Panel B further restricts the portfolio weight to $[0, 1]$.

Column 1 shows that unconditional excess returns are highly significant, with an annualized excess return (alpha) of 4.564%, and an annualized information ratio (equivalent to the Sharpe ratio in this regression) of 1.224. This Sharpe ratio is large relative to other daily market-level signals. For example, Da et al. (2024) report an out-of-sample Sharpe ratio of 0.46 and 0.17 when using abnormal retail attention and abnormal institutional attention as a signal, respectively. The social media Sharpe Ratio also exceeds the market Sharpe Ratio, which ranges from 0.3 to 0.5 in historical samples (Bodie et al., 2011, Mehra and Prescott, 1985).

When we control for date t market excess returns in Column 2, the social media alpha remains robust: annualized alpha is 4.75% with an information ratio of 1.246. However, the significant negative loading on date t market excess returns implies that the social media portfolio performs better following market *declines*. The magnitude of this estimate is economically large: a market excess return of -1% predicts social media portfolio returns will be 1.2 basis points higher the next day (roughly two-thirds of the daily alpha of 1.9 bps). In columns 3 and 4, we investigate whether portfolio returns are explained by a factor rotation using the SMB, HML or momentum factors. The intercept remains unchanged, and in contrast to market excess returns, none of these factors exhibits a significant relation to portfolio excess returns.

Panel B of Table 5 repeats this analysis with a long-only, no leverage constraint (i.e., with portfolio weights in [0,1]). The main difference is that information ratios in Panel B are larger than those in Panel A, while coefficient estimates remain similar throughout. This suggests our strategy is not driven by short-selling or leverage.

Table 6 presents two robustness tests of these portfolio results. First, in Panel A, we winsorize the forecast returns from Eq. (4) at the 90% and 10% percentiles before constructing the portfolio weights following Campbell and Thompson (2008), Da et al. (2024). This modification reduces our annualized alpha to 3.45%, but the Sharpe ratio remains above one (1.137). Second, in Panel B, we smooth portfolio weights using a trailing 5-day moving average rather than using weights directly from Eq. (5). Similar to the winsorization, using the 5-day moving average reduces the annualized alpha slightly to 3.8% while remaining significant, yet the Sharpe ratio of this strategy remains high at 1.078. Although the strategy requires daily turnover, its persistent performance with the 5-day moving average weights suggests that performance remains strong with lower portfolio turnover, and correspondingly reduced transaction costs.

Finally, we examine whether the significant excess returns from the dynamic strategy yield *abnormal* returns beyond the Fama-French 3 and momentum factors in Table 7. In columns 2 through 4, we find that the dynamic strategy yields an annualized alpha of 3%, which is statistically different from zero at the 5% level. The results are robust to the inclusion of the market, size, value and momentum factors. Size and value do not explain variation in portfolio excess returns while the market and momentum factors have a positive and significant loading.

Taken together, these findings indicate that the market and momentum factors explain approximately one-third of the average portfolio excess returns. The market factor alone explains 15.2% of portfolio returns while reducing the magnitude of alpha by just one-third. Appealing to an Oster (2019) argument, an omitted factor that captures an additional 5% of R^2 would need to be 6 times more important than the market factor to drive alpha to zero. Given that size, value and momentum factors collectively explain only *0.5%* of the variation in the portfolio excess returns (i.e., their contribution to increasing the R^2), it seems unlikely that an omitted factor explains the abnormal returns of our social media strategy.

4. Drivers of Sentiment and Attention

In this section, we explore the drivers of our sentiment and attention indexes. The return patterns in Figure 2 leading up to day 0 suggest that lagged returns might predict both sentiment and attention, which we explore using an OLS regression in the first subsection.

In subsequent subsections, we examine dynamic interdependence and feedback effects between variables over time. Specifically, we evaluate the drivers of sentiment and attention through two complementary analyses: (1) a vector autoregression (VAR) examining how returns and turnover drive the indexes, and (2) an event study of index responses to abrupt changes in prices and volatility, or "jumps."

4.1 Drivers of Sentiment and Attention Indexes

We begin by examining the drivers of our daily sentiment and attention indexes using an OLS regression specification:

$$\mathrm{Y}_t = \sum_{k=1}^{5} \beta_k \text{Market return}_{t-k} + \sum_{k=1}^{5} \gamma_k \text{Ab. log(market turnover)}_{t-k} + \Lambda_t + \epsilon_t \tag{7}$$

where Y_t is either the sentiment or the attention index observed on day t, and Market return$_{t-k}$ is measured by the S&P500 index's return on day $t-k$. Ab. log(market turnover)$_{t-k}$ is abnormal log market turnover on day $t-k$, measured as either a market-capitalization weighted average of abnormal turnover across S&P500 stocks or as abnormal turnover of the SPY exchange traded fund (ETF), the most popular S&P500 ETF . The Λ_t vector includes day-of-week, month-of-year, and year-quarter fixed effects to account for seasonality and slow-moving annual trends.

Table 8 presents the results. Day $t-1$ and $t-2$ returns positively and significantly predict day t sentiment index. This pattern is consistent with extrapolative beliefs documented in other asset pricing studies (e.g., Lakonishok et al., 1994, Case et al., 2012, Greenwood and Shleifer, 2014, Barberis et al., 2018). By contrast, day $t-1$ turnover negatively predicts day t sentiment. These relationships remain similar when we include year-month fixed effects in Appendix Table A1, which accounts flexibly for any monthly-level variability, including commonly used sentiment measures (e.g., Baker and Wurgler, 2006).

For the attention index, the most prominent driver is the previous day's turnover. Interestingly, column 3 shows that attention is more closely related to day $t-1$ turnover of S&P500 stocks (i.e., the aggregate trading in the S&P500 components), than to day $t-1$ turnover in the SPY ETF (column 4). This suggests that our attention index captures the dispersed information impounded in market trading (Hayek, 1945). In short, high market-wide abnormal turnover tends to predict higher attention the following day.

4.2 VAR models

We next estimate a VAR model, which allows for interdependence between sentiment, attention, market return, and market abnormal turnover:

$$
\begin{aligned}
\text{Sentiment}_t &= c_1 + \sum_{\tau=1}^{T} \Theta_\tau^{(1)} \cdot \mathbf{L}_{t-\tau} + \varepsilon_{1t} \\
\text{Attention}_t &= c_2 + \sum_{\tau=1}^{T} \Theta_\tau^{(2)} \cdot \mathbf{L}_{t-\tau} + \varepsilon_{2t} \\
\text{Market return}_t &= c_3 + \sum_{\tau=1}^{T} \Theta_\tau^{(3)} \cdot \mathbf{L}_{t-\tau} + \varepsilon_{3t} \\
\text{Log(Ab. market turnover)}_t &= c_4 + \sum_{\tau=1}^{T} \Theta_\tau^{(4)} \cdot \mathbf{L}_{t-\tau} + \varepsilon_{4t}
\end{aligned}
\tag{8}
$$

where we regress each of the dependent variables — Sentiment_t, Attention_t, Market return_t and $\text{Log(Ab. market turnover)}_t$ — on T daily lags of all four variables. Our main specifications use $T = 10$ daily lags, selected by optimizing over the AIC. The vector $\mathbf{L}_{t-\tau}$ contains the four dependent variables lagged τ days, with the coefficient vector $\Theta_\tau^{(i)}$ containing the corresponding loadings. The fitted VAR model thus captures the joint dynamics and feedback of sentiment, attention, market returns and abnormal market turnover. Following (Sims, 1980), we summarize the properties of the fitted VAR model by examining the impulse response to one standard deviation shocks in Market return_t and $\text{Log(Ab. market turnover)}_t$.

Figure 6 presents impulse response functions for sentiment and attention from day $t+1$ through $t+20$ following two separate shocks: a one standard deviation increase in market return on day t (panels a and c), and a one standard deviation increase in market turnover on day t (panels b and d). We proxy for market turnover using abnormal aggregate turnover in composite stocks of S&P500 (solid line) and abnormal turnover in SPY (dashed line). Consistent with the simple lagged OLS results in Table 8, sentiment increases for one to two days after a positive return shock, while attention decreases for two days. By contrast, a one-standard-deviation abnormal turnover shock in S&P500 stocks does not trigger responses in sentiment but increases attention for the next several days. Notably, a one standard deviation abnormal turnover shock SPY does not lead to higher attention.

We next conduct two robustness checks on these results. First, we account flexibly for alternative attention measures in the VAR estimation. Adding daily retail attention index based on Google searches for tickers and institutional attention index based on Bloomberg searches for tickers (Da et al., 2024) to the VAR model results in qualitatively similar patterns (see Appendix Figure A5). Second, given that retail investors dominate social media platforms, we examine whether the social media indexes responds differently to retail trading. In Appendix Figure A6, we replace total turnover with turnover based on retail trades as measured in Boehmer et al. (2021). We find similar responses of sentiment and attention to a shock in returns, but stronger responses to a shock in retail abnormal turnover.

4.3 Sentiment and attention around jumps

We next examine how our indexes behave around sudden price and volatility jumps. We classify market *jumps* as days that had at least a 2 percentage point change in returns (up or down). We also classify volatility jumps — daily spikes in the VIX exceeding 15pp, 20pp, and 25pp thresholds. For these jump days, we examine how the sentiment and attention indexes evolve from 4 days before to 10 days after the jump day. We analyze positive and negative return jumps separately. To ensure non-overlapping windows, if there are multiple large market movements in a row, we only consider the first one as a jump event.

We first examine how sentiment and attention indexes behave around market jumps, separately considering negative and positive jumps to allow for asymmetric responses. We estimate the following event-study regression:

$$\text{Social index}_t = \sum_{\tau=-4}^{10} \beta_\tau \text{Pos jump}_0 + \sum_{\tau=-4}^{10} \gamma_\tau \text{Neg jump}_0 + \theta \text{Neg jump}_0 + \Lambda_t + \epsilon_t$$

where Social index$_t$ is the sentiment (or attention) index on day t. Pos (Neg) jump$_0$ equals one for all days in the [-15,+10] event window around a positive (negative) jump at day $\tau = 0$ in event time. We estimate separately lead and lag coefficients for positive jumps (β_τ) and for negative jumps (γ_τ) from $\tau = -4$ through $\tau = +10$. In this analysis, we include days $t - 15$ through $t + 10$ around each market jump events in the sample. By only estimating leads and lags from $t - 4$ to $t + 10$, we set the reference period to be days $t - 15$ though $t - 5$. As in Eq. 2, the Λ_t vector includes day-of-week, month-of-year, and year-quarter fixed effects to account for seasonality and slow-moving annual trends, as well as lagged market volatility (day $t - 5$ through $t - 1$), and lagged market returns (day $t - 5$ through $t - 1$ and the previous 25 days).

Figure 7 plots the estimates for β_τ (panels a and c) and γ_τ (panels b and d). There are no trends in sentiment or attention prior to market jumps, indicating that these large market price movements are more likely to drive sentiment and attention rather than the other way around. There is an interesting asymmetry in the response of sentiment and attention to positive versus negative market jumps. Specifically, positive jumps (good news) do not correspond to significant changes in sentiment or attention. In contrast, negative jumps lead to a sharp and persistent drop in sentiment and increases in attention. Moreover, results are similar if we use alternative market jump definitions (+/- 1.5 percentage points movements in the S&P500 index) or exclude jumps that coincide with FOMC announcement days (Appendix Figure A7).

Figure 8 presents an analogous set of results for volatility jump days. Large positive spikes in the VIX can be interpreted as negative news, similar to negative market jumps, with no clear analogue to the positive market jumps. Consistent with earlier results on sentiment and attention responses around "bad news," proxied for by negative market jumps, we find that sentiment and attention indexes behave very similarly around spikes in volatility. Sentiment drops sharply on the event day, and remains below normal levels for several days. In contrast, attention increases somewhat and persists at elevated levels for several days. These patterns remain similar when we exclude volatility jumps that coincide with FOMC announcement days (see Appendix Figure A8).

Next, we examine these patterns in regression form in Table 9. This allows us to study the movements of sentiment and attention indexes around market jumps while controlling for other determinants of the indexes. We estimate the following specification over [-15,10] day event windows around jumps:

$$\begin{aligned}\text{Social index}_t \quad = \quad & \sum_l \alpha_l + \beta_0 \text{Neg jump}_0 + \beta_1 \text{Neg jump}_0 Day_{-1} + \beta_2 \text{Neg jump}_0 Day_0 \\ & + \beta_3 \text{Neg jump}_0 Day_{+1} + \beta_4 \text{Neg jump}_0 Day_{+2\to+10} + \Lambda_t + \epsilon_t\end{aligned}$$

where Social index is either the sentiment or attention index. This specification includes event day indicators (α_l) for day -1, day 0, day +1, and days +2 though +10 in event time relative to the jump day ((day 0), and their interactions with Neg jump_0, an indicator for negative jump events on day 0; positive jumps serve as the reference group. Given the event window [-15,+10], the baseline period spans 15 to 2 days before each jump. The Λ_t vector includes day-of-week, month-of-year, and year-quarter fixed effects controlling for seasonality in calendar time t as well as lagged market volatility (day $t-5$ through $t-1$), lagged market returns (day $t-5$ through $t-1$ and the previous 25 days). In stricter specifications, we also control for changes in VIX and MOVE indexes on jump days.

The results in Table 9 highlight the sharp drop in sentiment and spike in attention around negative market jump days. Specifically, on negative jump days sentiment declines on average by 0.768 standard deviations, relative positive jump days (column 1). This decline in sentiment persists (0.572 standard deviations below the baseline, on average) on the day after a negative market jump, gradually returning to baseline within the event window. These patterns are robust to controlling for recent market returns, volatility, and seasonal effects, and remain similar when we additionally control for changes in the VIX and the MOVE indexes in column 2. These findings highlight that sentiment responds significantly to stock market movements, and does not merely reflect underlying changes to volatility or bond markets.

Columns 3 and 4 present analogous specifications for the attention index. Mirroring our graphical evidence in Figure 7, we estimate that attention raises significantly on negative jump days and not on positive jump days. The effect is economically large: attention is 0.679 standard deviations higher on negative jump days than on positive jump days (column 3). However, unlike sentiment, the attention index reverts to normal levels more quickly, showing no statistically significant difference from baseline the day after the negative jump. As with the sentiment results, these findings are robust to controlling for recent returns and volatility and calendar fixed effects, as well as to controlling for changes in the VIX and MOVE indexes (column 4).

5. Conclusion

As social media has become pervasive, it has become a conduit of our collective attention. Recent market events like GameStop and Silicon Valley Bank highlight social media's role as a key venue for expressing sentiment about market events. This paper leverages these trends to develop daily sentiment and attention indexes using social media data.

Apart from significantly predicting future returns, our indexes provide a novel perspectives on both the *economic content* and *timing* of changes to market-wide sentiment. Relating to the economic content, attention and sentiment have sharply differing return dynamics. Our social media setting allows for a natural separation between these two concepts. This is an important insight given that existing sentiment indexes often conflate sentiment and attention factors. Relating to timing, our daily indexes capture within-month market dynamics missed by most existing research using monthly sentiment indexes. This higher frequency variation is particularly relevant given significant daily trading activity especially from retail investors who favor short-term strategies (Odean, 1999, Barber and Odean, 2000).

Moreover, our results on the drivers of daily sentiment and attention offer useful new facts for behavioral updating models. For example, while our results are broadly consistent with extrapolative belief models, daily sentiment exhibits an important asymmetry: showing *no* response to positive market jumps but sharp, persistent declines after negative ones. This result — and its contrast with monthly extrapolative patterns — presents a challenge to understanding how the sentiment drawn from the daily news cycle relates to sentiment drawn from slower moving cycles in the broader economy.

References

Avery, Christopher N, Judith A Chevalier, and Richard J Zeckhauser, 2016, The “caps” prediction system and stock market returns, *Review of Finance* 20, 1363–1381.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2018, Social connectedness: Measurement, determinants, and effects, *Journal of Economic Perspectives* 32, 259–80.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The journal of Finance* 61, 1645–1680.

Baker, Scott R, Nicholas Bloom, S Davis, and Thomas Renault, 2021, Twitter-derived measures of economic uncertainty .

Baker, Scott R, Nicholas Bloom, and Steven J Davis, 2016, Measuring economic policy uncertainty, *The Quarterly Journal of Economics* 131, 1593–1636.

Barber, Brad M, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *The journal of Finance* 55, 773–806.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2018, Extrapolation and bubbles, *Journal of Financial Economics* 129, 203–227.

Bodie, Zvi, Alex Kane, and Alan J Marcus, 2011, *Investments* (McGraw-hill).

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *Journal of Finance* 76, 2249–2305.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, Matthew OBrien, and Andrei Shleifer, 2024, Long-term expectations and aggregate fluctuations, *NBER Macroeconomics Annual* 38, 311–347.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2018, Diagnostic expectations and credit cycles, *The Journal of Finance* 73, 199–227.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2020, Memory, attention, and choice, *The Quarterly journal of economics* 135, 1399–1442.

Bradley, Daniel, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao, 2024, Place your bets? the value of investment research on reddit’s wallstreetbets, *The Review of Financial Studies* 37, 1409–1459.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

Case, Karl E., Robert J. Shiller, and Anne K. Thompson, 2012, What have they been thinking? homebuyer behavior in hot and cold markets, *Brookings Papers on Economic activity* 265–315.

Cassella, Stefano, Chukwuma Dim, and Tural Karimli, 2023, Optimism shifting, *Available at SSRN 4557313* .

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of crowds: The value of stock opinions transmitted through social media, *Review of Financial Studies* 27, 1367–1403.

Chen, Hailiang, and Byoung-Hyoun Hwang, 2022, Listening in on investors' thoughts and conversations, *Journal of Financial Economics* 145, 426–444.

Chen, Shuaiyu, Lin Peng, and Dexin Zhou, 2024, Wisdom or whims? decoding investor trading strategies with large language models, *SSRN Working Paper (June 19, 2024)* .

Cookson, J Anthony, Joseph Engelberg, and William Mullins, 2023a, Echo Chambers, *Review of Financial Studies* 36, 450–500.

Cookson, J Anthony, Joseph E Engelberg, and William Mullins, 2020, Does partisanship shape investor beliefs? Evidence from the COVID-19 pandemic, *Review of Asset Pricing Studies* 10, 863–893.

Cookson, J. Anthony, Corbin Fox, Javier Gil-Bazo, Juan Felipe Imbet, and Christoph Schiller, 2023b, Social media as a bank run catalyst, *Working paper* .

Cookson, J Anthony, Runjing Lu, William Mullins, and Marina Niessner, 2024a, The social signal, *Journal of Financial Economics* 158, 103870.

Cookson, J Anthony, William Mullins, and Marina Niessner, 2024b, Social media and finance, *Oxford Research Encyclopedia of Economics and Finance* .

Cookson, J Anthony, and Marina Niessner, 2020, Why don't we agree? Evidence from a social network of investors, *Journal of Finance* 75, 173–228.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all fears investor sentiment and asset prices, *The Review of Financial Studies* 28, 1–32.

Da, Zhi, Jian Hua, Tim Chih-Ching Hung, and Lin Peng, 2024, Market returns and a tale of two types of attention, *Available at SSRN 3551662* .

Da, Zhi, and Xing Huang, 2020, Harnessing the wisdom of crowds, *Management Science* 66, 1847–1867.

Da, Zhi, Xing Huang, and Lawrence J Jin, 2021, Extrapolative beliefs in the cross-section: What can we learn from the crowds?, *Journal of Financial Economics* 140, 175–196.

Davies, Shaun William, 2022, Speculation sentiment, *Journal of Financial and Quantitative Analysis* 57, 2485–2515.

DeVault, Luke, Richard Sias, and Laura Starks, 2019, Sentiment metrics and investor demand, *The Journal of Finance* 74, 985–1024.

Dim, Chukwuma, 2020, Social media analysts' skills: Insights from text-implied beliefs, *Available at SSRN 3813252* .

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Fisher, Adlai, Charles Martineau, and Jinfei Sheng, 2022, Macroeconomic attention and announcement risk premia, *The Review of Financial Studies* 35, 5057–5093.

Garcia, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267–1300.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *The Review of Financial Studies* 27, 714–746.

Greenwood, Robin, Andrei Shleifer, and Yang You, 2019, Bubbles for fama, *Journal of Financial Economics* 131, 20–43.

Hayek, FA, 1945, The use of knowledge in society, *American Economic Review* 35.

Henderson, Brian J, Neil D Pearson, and Li Wang, 2023, Retail derivatives and sentiment: A sentiment measure constructed from issuances of retail structured equity products, *The Journal of Finance* 78, 2365–2407.

Hirshleifer, David, Lin Peng, and Qiguang Wang, 2024, News diffusion in social networks and stock market reactions, *The Review of Financial Studies* hhae025.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *The Review of Financial Studies* 28, 791–837.

Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou, 2019, Manager sentiment and stock returns, *Journal of Financial Economics* 132, 126–149.

Keynes, John Maynard, 1936, The general theory of employment, interest and money.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Mehra, Rajnish, and Edward C Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

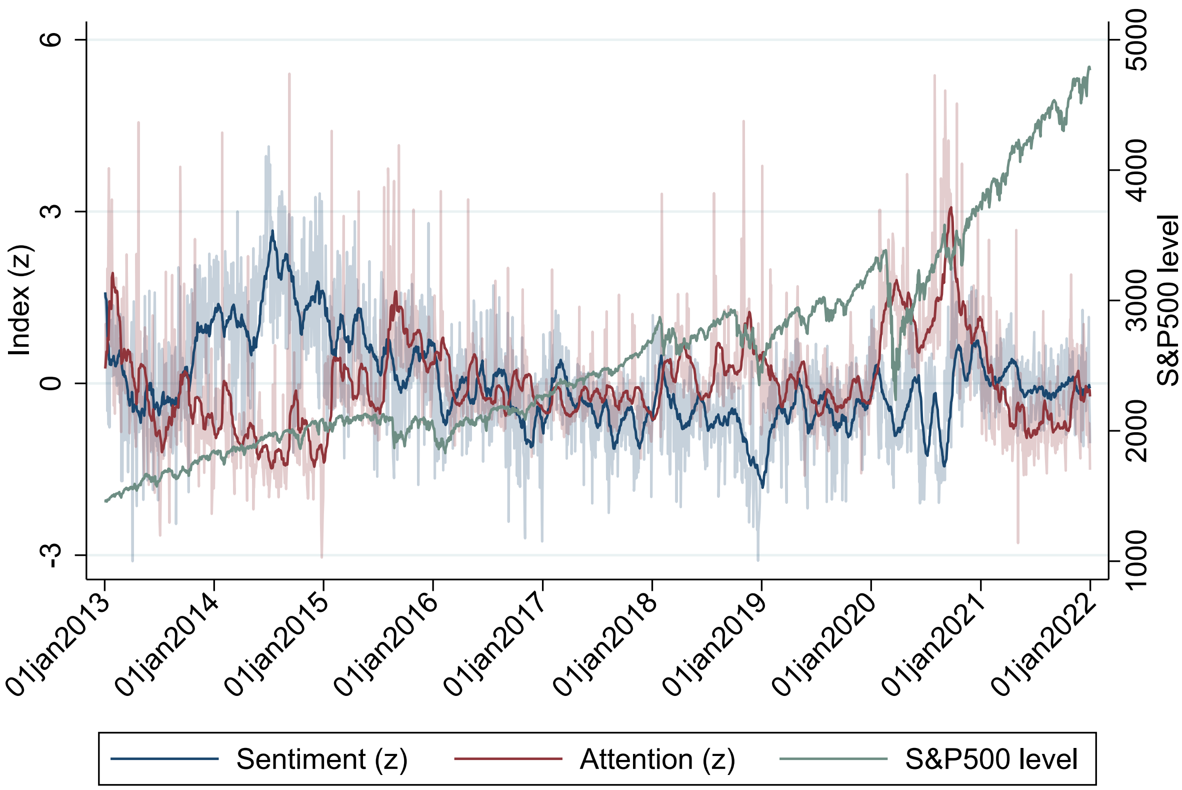
Odean, Terrance, 1999, Do investors trade too much?, *American economic review* 89, 1279–1298.

Oster, Emily, 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37, 187–204.

Sims, Christopher A, 1980, Macroeconomics and reality, *Econometrica: journal of the Econometric Society* 1–48.

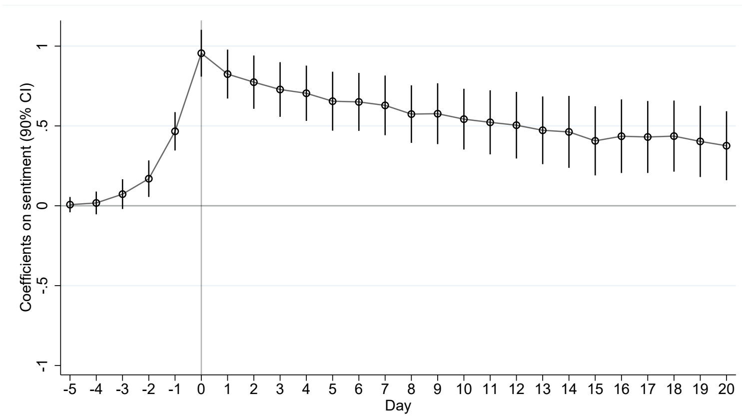
Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Figure 1: Time Series of Sentiment and Attention Indexes

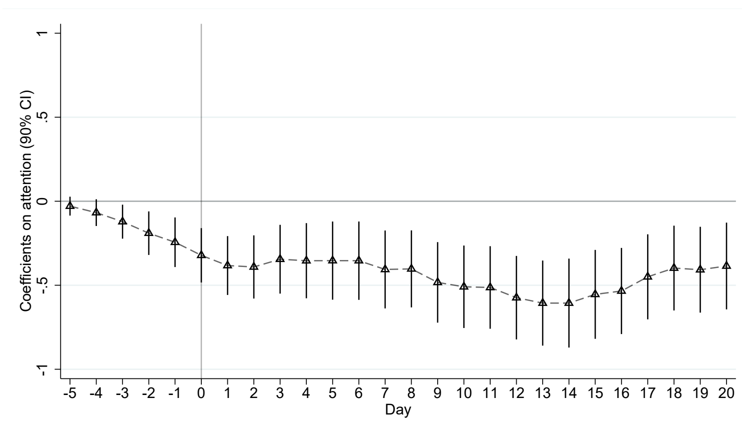


Note: This figure plots the time series for sentiment index (blue) and attention index (red) benchmarked against the S&P 500 index level (green). The lighter-colored lines plot the daily series of sentiment and attention indexes while the darker-colored lines plot the corresponding 20-day rolling average of each series. Sentiment (attention) index is the first principal component from a principal component analysis of platform-day level market-weighted average residualized sentiment (attention) signal across firms, normalized to have a mean zero and standard deviation of one. Platform specific firm-day level residualized signal is obtained by regressing firm-day level signal on the firm-specific annual average in the prior year and indicators for presence of firm news (8K, Earnings announcement, or DJNW news coverage) on day t-7 through t, seperately for each platform. Standard errors are calculated via Newey-West with 6 lags. See Section 2.4 for index construction.

Figure 2: How Do Cumulative Returns Relate to Sentiment and Attention Indexes?



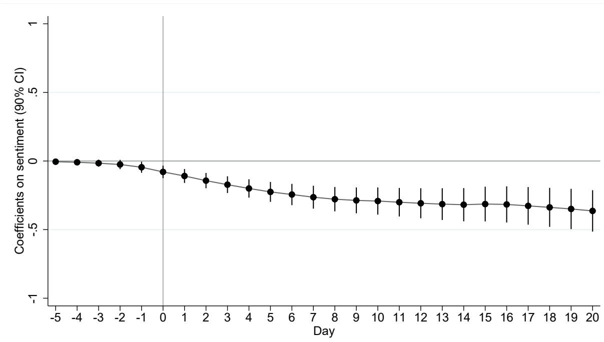
(a) Sentiment



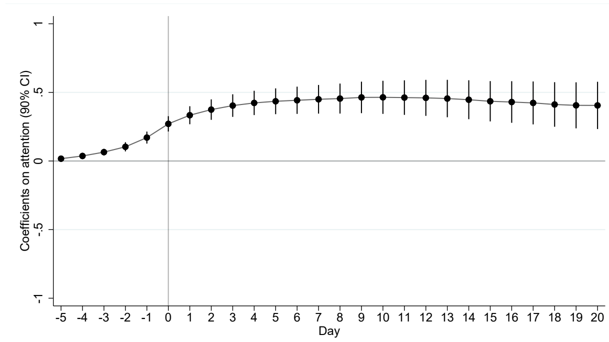
(b) Attention

Note: This figure plots the estimated coefficients (and 90% confidence intervals) on sentiment and attention indexes by regressing cumulative S&P 500 returns starting from day $t-5$ on sentiment index, attention index, and their interactions at day 0 for an event window between days $t = -5$ and $t = +20$. The regressions control for return volatilities in the prior 6-10 trading days, past returns in the prior 6-10 trading days and the previous 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Standard errors are calculated via Newey-West with 6 lags.

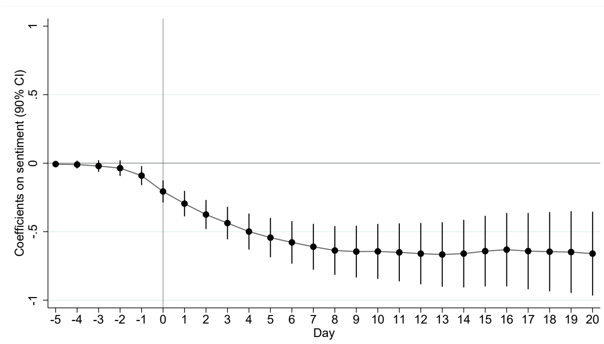
Figure 3: How Does Cumulative Abnormal Turnover Relate to Sentiment and Attention Indexes?



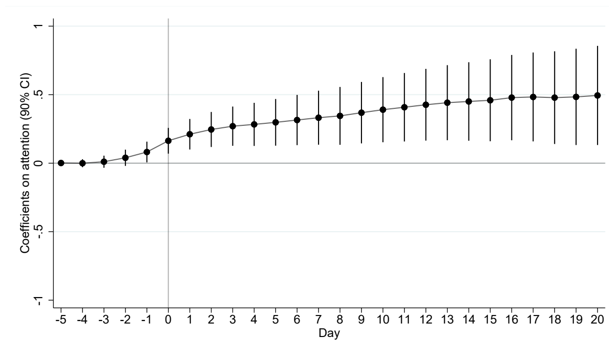
(a) Sentiment and S&P 500 turnover



(b) Attention and S&P 500 turnover



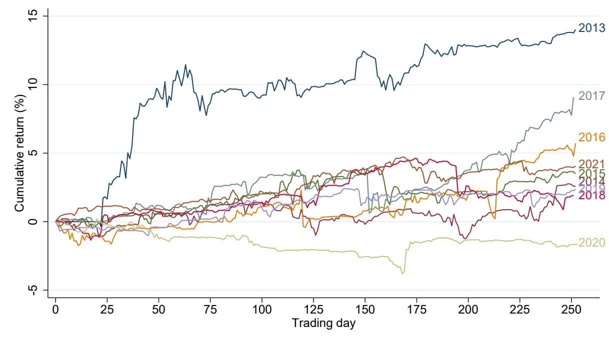
(c) Sentiment and SPY turnover



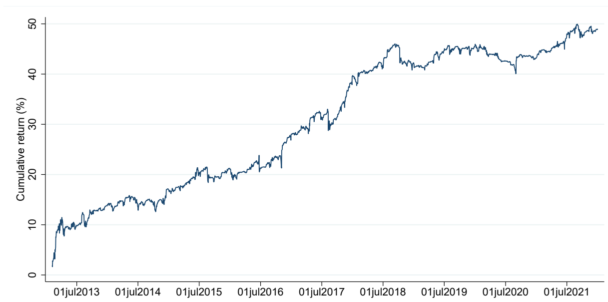
(d) Attention and SPY turnover

Note: This figure plots the estimated coefficients (and 90% confidence intervals) on sentiment and attention indexes by regressing cumulative abnormal turnover starting from day $t-5$ on sentiment index, attention index, and their interactions at day 0 for an event window between days $t=-5$ and $t=+20$. Cumulative abnormal turnover is the log turnover less the mean log turnover in the prior 140 through 20 days. S&P 500 turnover is the market-weighted turnover across all S&P 500 firms based on total trading. SPY turnover is the turnover for the SPY index based on total trading. The regressions control for abnormal turnover on day $t-1$, return volatilities in the prior 6-10 trading days, past returns in the prior 6-10 trading days and the previous 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Standard errors are calculated via Newey-West with 6 lags.

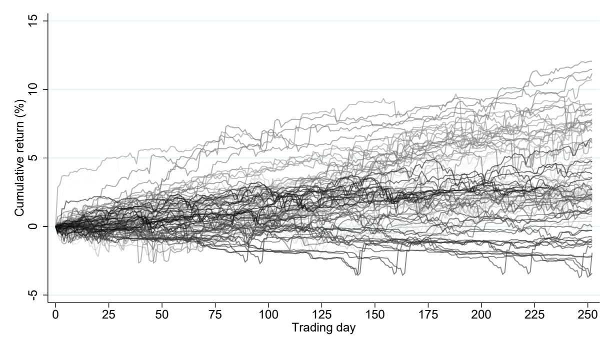
Figure 4: Dynamic Trading Strategy Based on Social Media Indexes
Cumulative Excess Return



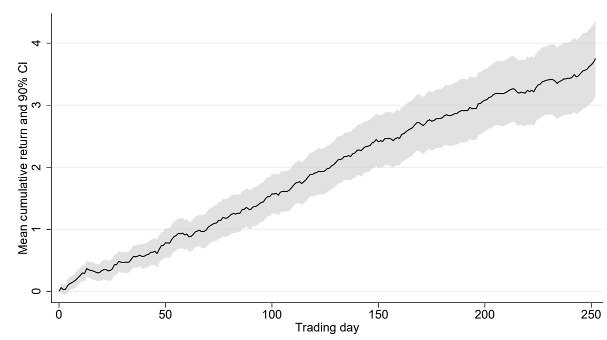
(a) Cumulated from 1st trading day each year



(b) Cumulated from 2013 through 2021



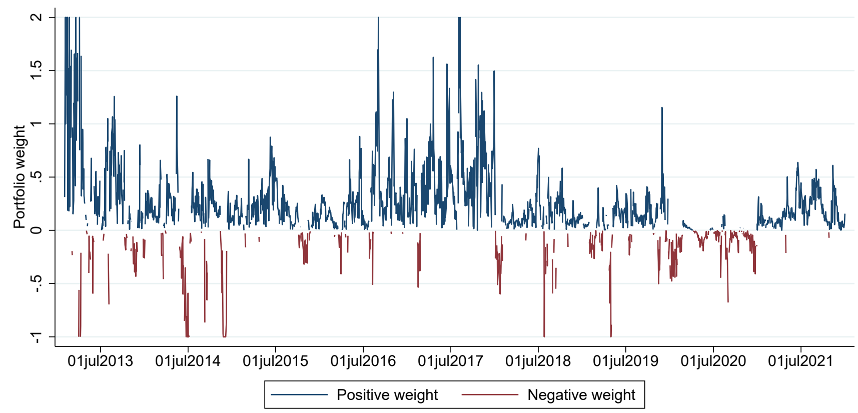
(c) Cumulated from 100 random start dates



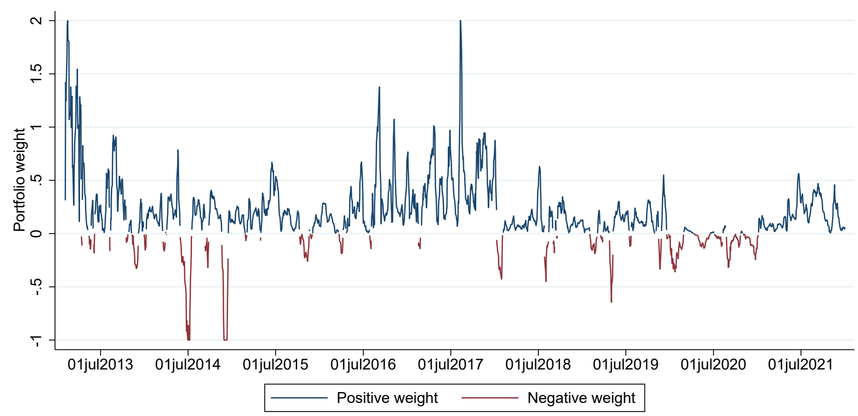
(d) Average and 90% confidence for Panel C

Note: This figure plots the cumulative buy-and-hold excess return from a dynamic trading strategy based on social media indexes constructed using information up to last month (see Section 3.3). Panel A presents the cumulative return from the first trading day to the last trading day of each year, separately for the 9 years of our sample. Panel B presents the return plot from 2013 through 2021. In Panel C and D, we construct return plots for one year following 100 randomly drawn start dates. Panel C presents all 100 paths, whereas Panel D presents the average return of those paths with a 90% confidence band.

Figure 5: Dynamic Trading Strategy Based on Social Media Indexes
Time Series of Portfolio Weights



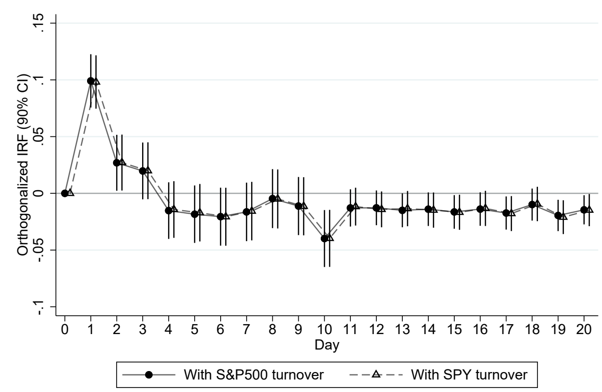
(a) Baseline portfolio weight



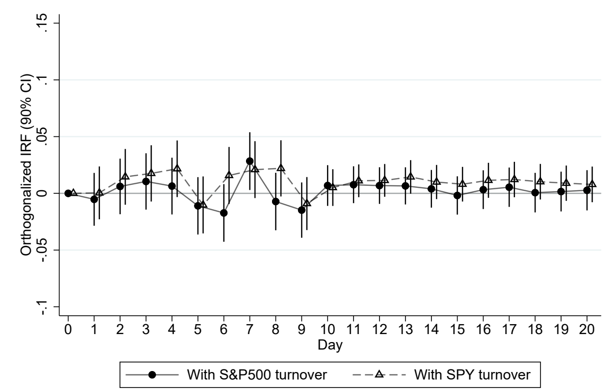
(b) Rolling 5-day avg. weight

Note: This figure plots the portfolio weights for a dynamic trading strategy based on social media indexes constructed using information up to last month (see Section 3.3). Panel A plots the daily weight while Panel B plots the 5-day rolling average. On average, 76% (78%) of days put positive weights on the market return and 3% (2%) of days have leverage, i.e., a portfolio weight exceeding 1, in Panel A (Panel B).

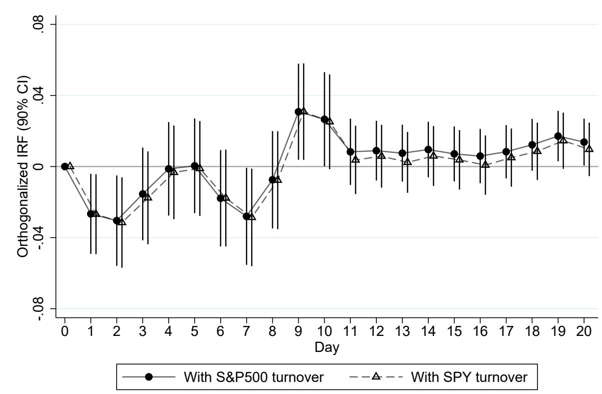
Figure 6: What Predicts Sentiment and Attention Indexes?
Impulse Response Function from a VAR Model



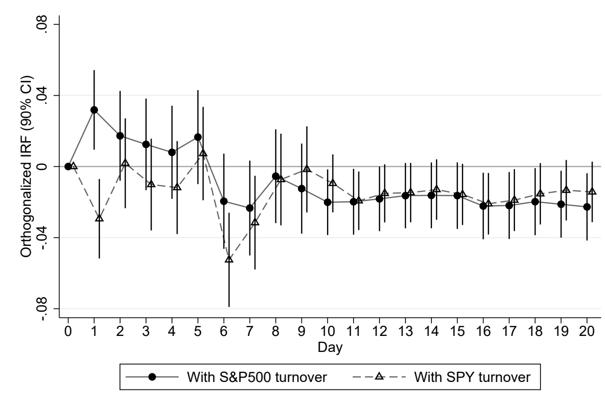
(a) Sentiment response to return



(b) Sentiment response to turnover



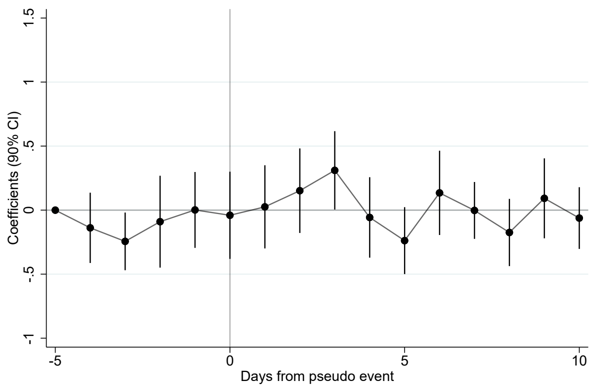
(c) Attention response to return



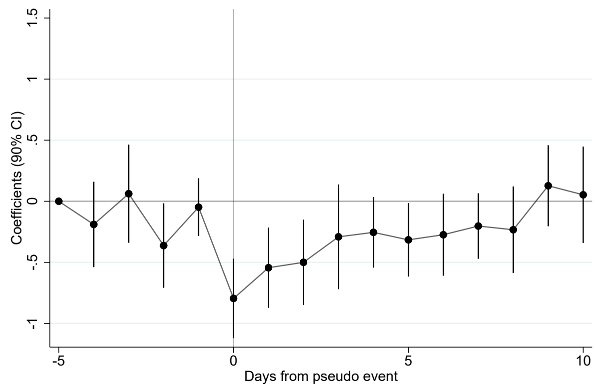
(d) Attention response to turnover

Note: This figure plots the orthogonalized impulse response function (and 90% confidence intervals) of sentiment and attention indexes on day t+1 through day t+20 to a standard-deviation change in returns or turnover on day t. Returns refer to S&P 500 daily return while turnover refers to abnormal log(turnover) based on market-weighed trading across all S&P 500 firms ("with S&P500 turnover") or trading of SPY index ("with SPY turnover"). Appendix Figure A5 presents robustness checks by adding daily attention indexes newly developed in the literature to the VAR system.

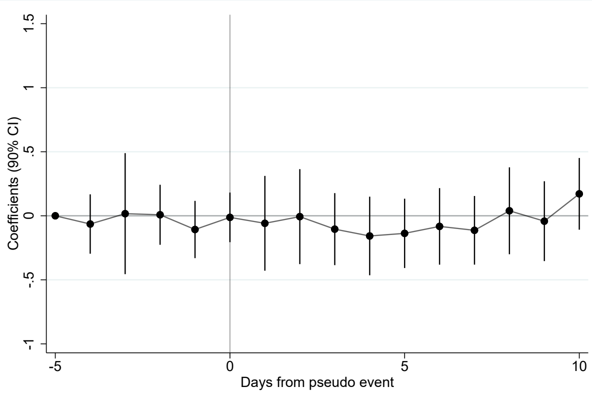
Figure 7: How Do Sentiment and Attention Indexes Change around Return Jumps



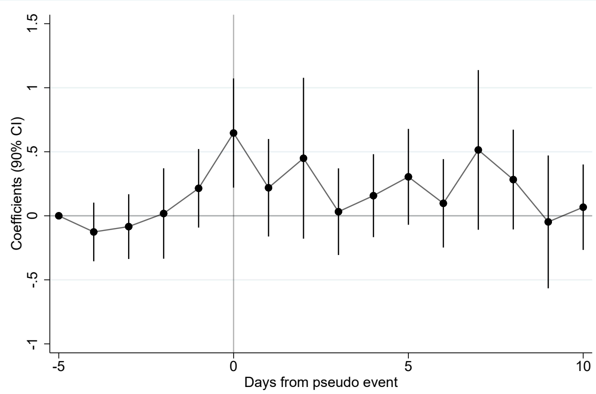
(a) Sentiment, positive jumps



(b) Sentiment, negative jumps



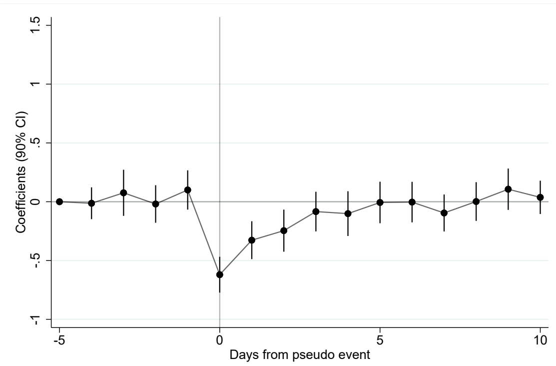
(c) Attention, positive jumps



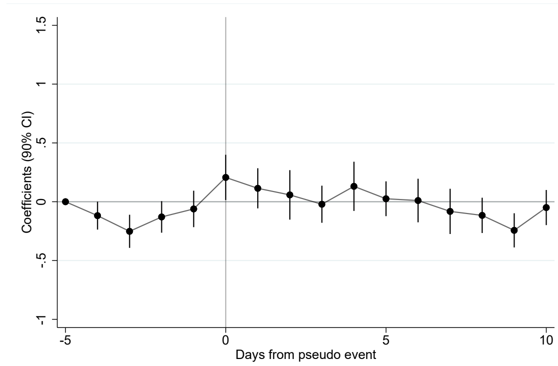
(d) Attention, negative jumps

Note: This figure plots how sentiment index (first row) and attention index (second row) change from day t-4 through t+10 around days with extreme return jumps. We categorize days with S&P 500 returns $\leq$ -2pp as negative jumps and days with S&P 500 returns $\geq$ +2pp as positive jumps. These events are further required to be at least 10 days apart from the last corresponding event of its type, leaving us with 13 negative return jumps and 22 positive return jumps. To obtain the coefficients (and 90% confidence intervals), we regress sentiment (or attention) index on indicators for positive return jumps (or negative return jumps) and the interactions between them and the indicators for day t-4, ..., t+10 around an event. Days t-15 through t-5 are used as the reference group and represented with a dot on day t-5. All regressions control for lagged volatilities (day t-5 through t-1), lagged returns (day t-5 through t-1 and the previous 25 days), DOW, MOY, and YQ fixed effects. Standard errors are calculated via Newey-West with 6 lags. Appendix Figure A7 presents robustness checks by excluding jumps coinciding with FOMC announcements and redefining jump events using +/-1.5pp as thresholds.

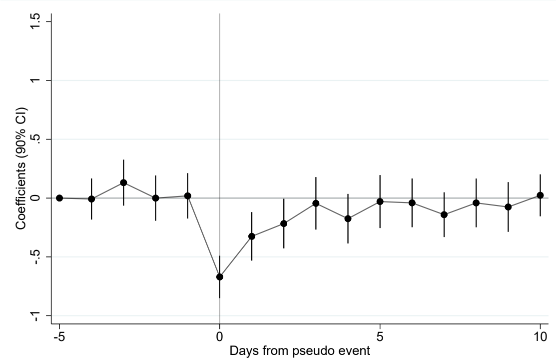
Figure 8: How Do Sentiment and Attention Indexes Change around High VIX Change?



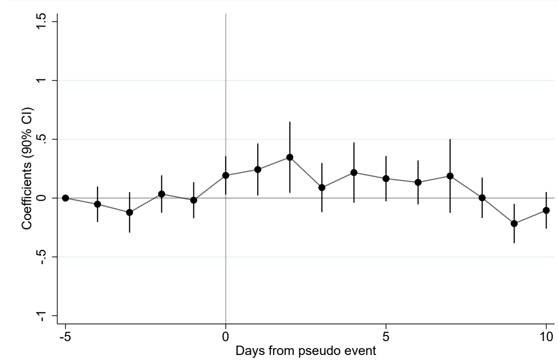
(a) Sentiment, high-Δ Vix, 15pp



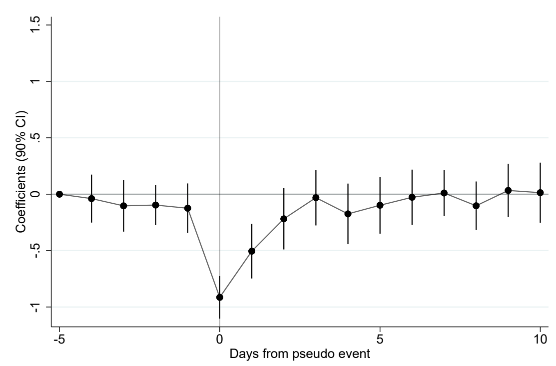
(b) Attention, high-Δ Vix, 15pp



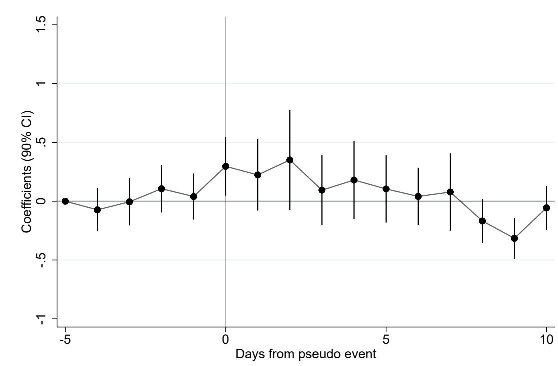
(c) Sentiment, high-Δ Vix, 20pp



(d) Attention, high-Δ Vix, 20pp



(e) Sentiment, high-Δ Vix, 25pp



(f) Attention, high-Δ Vix, 25pp

Note: This figure plots how sentiment index (first row) and attention index (second row) change from day t-4 through t+10 around days with a sharp increase in VIX. We categorize days with $\geq +15$, 20, or 25pp change in Vix from the prior day as high-Δ Vix events. These events are further required to be at least 10 days apart from the last event, leaving us with 46, 40, or 28 high-Δ Vix events. To obtain the coefficients (and 90% confidence intervals), we regress sentiment (or attention) index on indicators for high change in VIX and the interactions between them and the indicators for day t-4, ..., t+10 around an event. Days t-15 through t-5 are used as the reference group and represented with a dot on day t-5. All regressions control for lagged volatilities (day t-5 through t-1), lagged returns (day t-5 through t-1 and the previous 25 days), DOW, MOY, and YQ fixed effects. Standard errors are calculated via Newey-West with 6 lags.

Table 1: Summary Statistics of Investor Social Media

	StockTwits	Twitter	Seeking Alpha
# firms	1,500	1,287	1,294
# firm-day obs	1,870,488	1,312,173	218,604
# posts per day	49,007	8,598	223
# firms covered per day	825	502	90
Market cap. covered per day	93%	88%	38%
# firms with $\geq$10 posts per day	354	226	58
Market cap. with $\geq$10 posts per day	52%	50%	28%

Note: This table reports summary statistics for the social media platforms we use to calculate sentiment and attention indexes. We start from the 1,500 firms with the most StockTwits posts. # firms counts the number of unique firms with any posts from 2013 through 2021. # firm-days are firm-day observations with at least one post. All rows starting with # posts per day are daily averages. Market cap. refers to percentage of the market capitalization of firms with at least one post over the total of the 1500 firms (recomputed daily). The final two rows restrict to our main sample, which focuses on firm-days with at least 10 posts on StockTwits.

Table 2: Sentiment and Attention Index Construction

Panel A: Residualizing regressions for platform-day signal

	Dep. var.: Sentiment$_{i,t}$ (z)			Dep. var.: Attention$_{i,t}$ (z)		
	ST	TW	SA	ST	TW	SA
Firm annual avg$_{i,y(t)-1}$	0.373***	0.569***	0.295***	0.834***	0.789***	0.531***
	(0.018)	(0.015)	(0.026)	(0.084)	(0.043)	(0.046)
Firm news controls	Y	Y	Y	Y	Y	Y
Observations	738,438	738,438	738,438	738,438	738,438	738,438
R^2	0.0349	0.1093	0.0665	0.0811	0.4612	0.4031

Panel B: PCA of platform-day signal

	Sentiment PC1	Attention PC1
StockTwits	0.649	0.707
	(0.020)	(0.014)
Twitter	0.675	0.706
	(0.013)	(0.016)
Seeking Alpha	0.352	0.040
	(0.091)	(0.099)
Fraction(%)	46.876	53.696
	(1.207)	(2.525)

Note: **Panel A** reports the residualizing regressions that absorb firm-level information from each platform's social media sentiment and attention signals. Each regression uses data from a single platform at the *firm-day* level, and is separately estimated for attention and sentiment. The regressions include indicators for firm news (i.e., DJNW sentiment and attention, earnings announcements, 8-K filings) occurring on days $t-7$ through t. We also control for the firm-average value of the signal (attention or sentiment) for the previous calendar year. For each platform-level regression we take the resulting firm-day residuals and aggregate them to the daily level using market value weights. **Panel B** reports a principal component analysis of these platform-level daily time series separately for attention and sentiment. Standard errors in paratheses are computed using a bootstrap method, involving 1,000 random sample iterations of firms with replacement.

Table 3: Do Sentiment and Attention Indexes Predict Returns?

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+2∼20	(6) Day t+2∼20
Sentiment$_t(z)$	0.524***	0.544***	-0.105***	-0.108***	-0.271**	-0.264**
	(0.041)	(0.042)	(0.035)	(0.037)	(0.117)	(0.125)
Attention$_t(z)$	-0.095***	-0.097***	-0.067**	-0.067**	-0.145	-0.146
	(0.029)	(0.030)	(0.033)	(0.033)	(0.142)	(0.142)
Sentiment $\times$ Attention$_t(z)$		0.158***		-0.022		0.061
		(0.038)		(0.031)		(0.118)
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.173	0.192	0.035	0.036	0.392	0.392

Note: This table reports how sentiment and attention indexes relate to day t, day t+1, and day t+2∼t+20 S&P 500 cumulative return. All regressions control for return volatilities in the prior 6-10 trading days, past returns in the prior 6-10 trading days and the previous 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Sentiment (attention) index is the first principal component from a principal component analysis of platform-day level market-weighted average residualized sentiment (attention) signal across firms, normalized to have a mean zero and standard deviation of one. Platform specific firm-day level residualized signal is obtained by regressing firm-day level signal on the firm-specific annual average in the prior year and indicators for presence of firm news (8K, Earnings announcement, or DJNW news coverage) on day t-7 through t, seperately for each platform. Standard errors are calculated via Newey-West with 6 lags. See Section 2.4 for index construction.
*** 1%, ** 5%, * 10% significance level

Table 4: Do Sentiment and Attention Indexes Predict Turnover?

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+2∼20	(6) Day t+2∼20
Panel A: S&P turnover						
Sentiment$_t(z)$	-0.020***	-0.021***	-0.017***	-0.018***	-0.192**	-0.199***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.074)	(0.076)
Attention$_t(z)$	0.071***	0.071***	0.042***	0.042***	0.057	0.058
	(0.007)	(0.007)	(0.006)	(0.005)	(0.077)	(0.078)
Sentiment × Attention$_t(z)$		-0.007		-0.008		-0.060
		(0.005)		(0.005)		(0.064)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.596	0.597	0.483	0.484	0.749	0.749
Panel B: SPY turnover						
Sentiment$_t(z)$	-0.078***	-0.080***	-0.057***	-0.058***	-0.207	-0.206
	(0.008)	(0.009)	(0.010)	(0.010)	(0.140)	(0.143)
Attention$_t(z)$	0.054***	0.054***	0.025**	0.025**	0.162	0.162
	(0.009)	(0.009)	(0.010)	(0.010)	(0.165)	(0.166)
Sentiment × Attention$_t(z)$		-0.010		-0.008		0.005
		(0.007)		(0.009)		(0.119)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.627	0.627	0.533	0.533	0.695	0.695
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y

Note: This table reports how sentiment and attention indexes relate to day t, day t+1, and day t+2∼t+20 cumulative abnormal turnover. Each panel represents a different outcome: panel A S&P 500 cumulative abnormal turnover and panel B SPY cumulative abnormal turnover. All regressions control for day t cumulative abnormal turnover, return volatilities in the prior 6-10 trading days, past returns in the prior 6-10 trading days and the previous 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Sentiment (attention) index is the first principal component from a principal component analysis of platform-day level market-weighted average residualized sentiment (attention) signal across firms, normalized to have a mean zero and standard deviation of one. Platform specific firm-day level residualized signal is obtained by regressing firm-day level signal on the firm-specific annual average in the prior year and indicators for presence of firm news (8K, Earnings announcement, or DJNW news coverage) on day t-7 through t, seperately for each platform. Standard errors are calculated via Newey-West with 6 lags. See Section 2.4 for index construction.
*** 1%, ** 5%, * 10% significance level

Table 5: Dynamic Trading Strategy Based on Social Media Indexes

	Dependent var.: Portfolio excess return$_{t+1}$(%)			
	(1)	(2)	(3)	(4)
Panel A: Weight* $\in$ *[-1,+2]				
Alpha	0.018***	0.019***	0.019***	0.019***
	(0.005)	(0.005)	(0.005)	(0.005)
Market excess return$_t$		-0.012***	-0.013***	-0.012**
		(0.005)	(0.005)	(0.005)
SMB$_t$			0.009	0.009
			(0.009)	(0.009)
HML$_t$			-0.006	-0.010
			(0.005)	(0.007)
MOM$_t$				-0.004
				(0.005)
Observations	2,246	2,246	2,246	2,246
R^2	–	0.002	0.003	0.003
Alpha (annualized)	4.564***	4.754***	4.739***	4.731***
	(1.249)	(1.278)	(1.278)	(1.279)
Information ratio (annualized)	1.224	1.246	1.242	1.239
Panel B: Weight* $\in$ *[0,1]				
Alpha	0.016***	0.017***	0.017***	0.017***
	(0.004)	(0.004)	(0.004)	(0.004)
Market excess return$_t$		-0.011***	-0.012***	-0.012***
		(0.004)	(0.004)	(0.004)
SMB$_t$			0.010	0.010
			(0.008)	(0.008)
HML$_t$			-0.006	-0.009
			(0.004)	(0.006)
MOM$_t$				-0.003
				(0.005)
Observations	2,246	2,246	2,246	2,246
R^2	–	0.003	0.004	0.004
Alpha (annualized)	4.079***	4.259***	4.244***	4.239***
	(0.981)	(1.018)	(1.018)	(1.019)
Information ratio (annualized)	1.393	1.402	1.396	1.394

Note: This table reports the excess return and factor loadings for a dynamic trading strategy based on social media indexes constructed using information up to last month (see Section 3.3). Column 1 shows the unconditional excess returns. Column 2 controls for date t market excess return. Column 3 additionally includes small minus big returns and value minus growth returns. Column 4 further adds momentum returns. Panel A permits short-selling and leverage by allowing portfolio weights to range from -1 to +2. Panel B restricts portfolio weights to a range of 0 to +1, thereby prohibiting short-selling and leverage. Standard errors are calculated via Newey-West with 6 lags.

Table 6: Dynamic Trading Strategy Based on Social Media Indexes, *Robustness*

	Dependent var.: Portfolio excess return$_{t+1}$(%)			
	(1)	(2)	(3)	(4)
Panel A: Return winsorized 10%				
Alpha	0.014***	0.014***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Market excess return$_t$		-0.007**	-0.008**	-0.007**
		(0.004)	(0.004)	(0.004)
SMB$_t$			0.009	0.009
			(0.008)	(0.008)
HML$_t$			-0.006	-0.007
			(0.004)	(0.006)
MOM$_t$				-0.002
				(0.005)
Observations	2,246	2,246	2,246	2,246
R^2	–	0.001	0.002	0.002
Alpha (annualized)	3.451***	3.563***	3.551***	3.546***
	(1.016)	(1.038)	(1.038)	(1.039)
Information ratio (annualized)	1.137	1.150	1.145	1.144
Panel B: 5-day avg weight				
Alpha	0.015***	0.016***	0.016***	0.016***
	(0.005)	(0.005)	(0.005)	(0.005)
Market excess return$_t$		-0.009*	-0.009*	-0.009*
		(0.005)	(0.005)	(0.005)
SMB$_t$			0.001	0.001
			(0.009)	(0.009)
HML$_t$			-0.004	-0.004
			(0.005)	(0.007)
MOM$_t$				0.000
				(0.005)
Observations	2,245	2,245	2,245	2,245
R^2	–	0.002	0.002	0.002
Alpha (annualized)	3.800***	3.943***	3.925***	3.926***
	(1.181)	(1.191)	(1.195)	(1.195)
Information ratio (annualized)	1.078	1.109	1.101	1.101

Note: This table presents robustness checks for Table 5 Panel A by using an alternative outcome and portfolio weight. Panel A winsorizes forecast returns at 10% level before calculating portfolio weight. Panel B replaces portfolio weight with rolling 5-day average. Everything else follows Table 5 Panel A.

Table 7: Dynamic Strategy: Abnormal Returns and Factor Decomposition

	Dependent var.: Portfolio excess return$_{t+1}$(%)			
	(1)	(2)	(3)	(4)
Alpha	0.018***	0.012**	0.012**	0.012**
	(0.005)	(0.005)	(0.005)	(0.005)
Market excess return$_{t+1}$		0.096***	0.098***	0.097***
		(0.016)	(0.017)	(0.017)
SMB$_{t+1}$			-0.011	-0.010
			(0.009)	(0.009)
HML$_{t+1}$			-0.011	0.002
			(0.009)	(0.010)
MOM$_{t+1}$				0.018**
				(0.008)
Observations	2,246	2,246	2,246	2,246
R^2	–	0.152	0.154	0.157
Alpha (annualized)	4.564***	3.051**	2.984**	3.020**
	(1.249)	(1.226)	(1.234)	(1.232)
Information ratio (annualized)	1.224	0.833	0.810	0.821

Note: This table presents tests for whether the dyanmic strategy produces abnormal returns beyond the Fama and French (1993) risk factors plus the Carhart (1997) momentum factor. Column 1 repeats the unconditional portfolio returns. Column 2 asks whether these returns are abnormal with respect to the market factor. Column 3 controls for the three Fama-French factors. Column 4 additionally includes the momentum factor.

Table 8: What Predicts Social Media Sentiment and Attention Indexes?

	Dependent var.: $\text{Sentiment}_t(z)$		Dependent var.: $\text{Attention}_t(z)$	
	(1) S&P turnover	(2) SPY turnover	(3) S&P turnover	(4) SPY turnover
Return_{t-1}	0.144***	0.139***	-0.024*	-0.035**
	(0.027)	(0.028)	(0.013)	(0.014)
Return_{t-2}	0.074***	0.073***	-0.024	-0.031*
	(0.017)	(0.019)	(0.015)	(0.018)
Return_{t-3}	0.020	0.019	-0.011	-0.020
	(0.015)	(0.015)	(0.014)	(0.015)
Return_{t-4}	0.003	0.005	-0.007	-0.022
	(0.016)	(0.016)	(0.013)	(0.015)
Return_{t-5}	0.008	0.007	-0.000	-0.019
	(0.014)	(0.014)	(0.017)	(0.018)
Ab. $\log(\text{turnover})_{t-1}$	-0.208**	-0.136**	0.897***	0.167***
	(0.096)	(0.057)	(0.091)	(0.055)
Ab. $\log(\text{turnover})_{t-2}$	0.024	0.039	0.016	0.058
	(0.091)	(0.055)	(0.078)	(0.050)
Ab. $\log(\text{turnover})_{t-3}$	-0.035	-0.021	0.028	-0.025
	(0.085)	(0.055)	(0.073)	(0.053)
Ab. $\log(\text{turnover})_{t-4}$	-0.035	0.015	-0.025	-0.089
	(0.105)	(0.058)	(0.077)	(0.058)
Ab. $\log(\text{turnover})_{t-5}$	-0.196**	-0.090*	0.115	-0.005
	(0.090)	(0.052)	(0.078)	(0.055)
DOW FE	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y
Observations	2267	2267	2267	2267
R^2	0.535	0.533	0.533	0.505

Note: This table predicts day t sentiment and attention indexes using day t-5 through day t-1 S&P 500 daily return and abnormal daily turnover. Columns 1 and 3 use abnormal daily turnover based on trading of firms included in S&P 500 while columns 2 and 4 use abnormal daily turnover based on trading of SPY. All regressions control for DOW, MOY, and YQ fixed effects. Standard errors are calculated via Newey-West with 6 lags. Table A1 repeats this specification using year-month fixed effects.
*** 1%, ** 5%, * 10% significance level

Table 9: How Do Sentiment and Attention Indexes Change around jumps?

	Dependent var.: Sentiment$_t(z)$		Dependent var.: Attention$_t(z)$	
	(1)	(2)	(3)	(4)
Neg jump$_0\times$ Day$_{-1}$	-0.056	-0.057	0.342	0.294
	(0.226)	(0.240)	(0.213)	(0.218)
Neg jump$_0\times$ Day$_0$	-0.768***	-0.715**	0.679**	0.643**
	(0.290)	(0.300)	(0.272)	(0.276)
Neg jump$_0\times$ Day$_{+1}$	-0.572**	-0.562**	0.296	0.225
	(0.264)	(0.281)	(0.317)	(0.328)
Neg jump$_0\times$ Day$_{+2\to+10}$	-0.259	-0.275	0.282	0.302
	(0.160)	(0.169)	(0.206)	(0.218)
Neg jump$_0$	0.174	0.296**	-0.511***	-0.584***
	(0.130)	(0.141)	(0.177)	(0.175)
Volatility$_{t-5\to t-1}$	0.148**	0.123*	0.075	0.069
	(0.066)	(0.069)	(0.073)	(0.074)
CR$_{t-1\to t-5}$	0.067***	0.061***	-0.036**	-0.029*
	(0.013)	(0.014)	(0.015)	(0.016)
CR$_{t-30\to t-6}$	0.019*	0.015	0.003	0.013
	(0.012)	(0.012)	(0.012)	(0.012)
Change in VIX$_0$		-0.005**		0.004
		(0.003)		(0.003)
Change in MOVE$_0$		0.018**		0.007
		(0.008)		(0.012)
DOW FE	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y
Relative day controls	Y	Y	Y	Y
Observations	895	843	895	843
R^2	0.472	0.443	0.602	0.619

Note: This table presents how sentiment and attention indexes change around return jump events. Positive (negative) jumps are defined as days with S&P 500 returns $\geq$ +2% ($\leq$ -2%). jumps are further required to be at least 10 days apart from the last return jump, leaving us with 14 negative jumps and 21 positive jumps during the sample period. We then regresses sentiment and attention indexes on an indicator for negative jumps ("Neg jump$_0$"), indicators for day -1, day 0, day +1, and day +2 through +10 from jumps, and the interaction between the two. Positive jumps and days -15 through -2 from jumps are used as the reference group. All regressions control for lagged volatilities (day t-5 through t-1), lagged returns (day t-5 through t-1 and the previous 25 days), as well as DOW, MOY, and YQ fixed effects. Columns 2 and 4 additionally control for the change in VIX and MOVE indexes on the jump day. Standard errors are calculated via Newey-West with 6 lags. Table **??** reports an alternative specification where we include indicators for both positive and negative jumps.
*** 1%, ** 5%, * 10% significance level

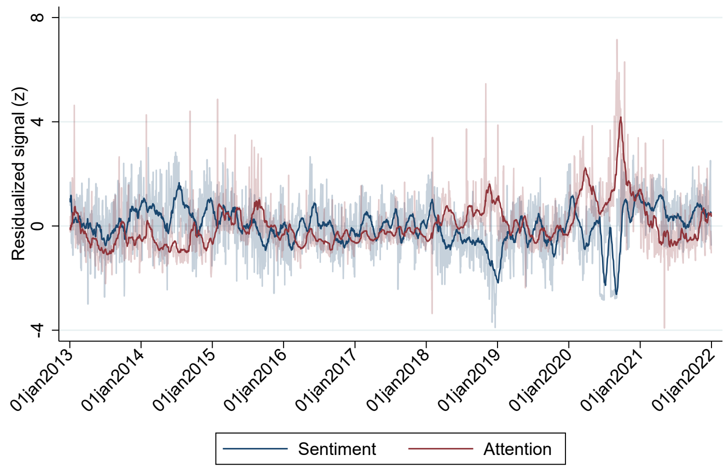
Internet Appendix:

Market Signals from Social Media

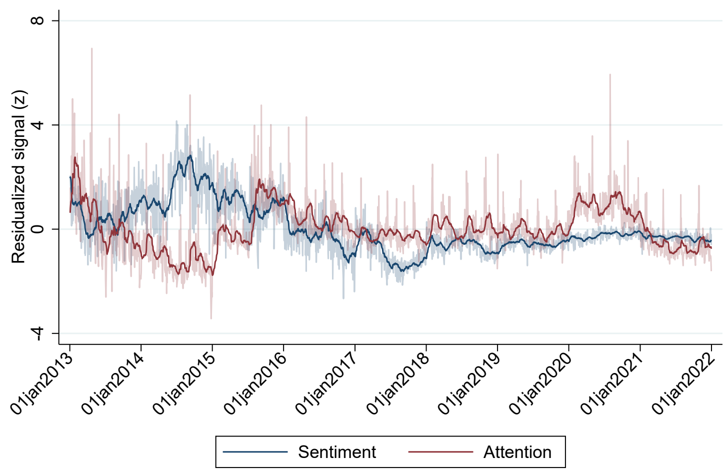
by J. Anthony Cookson, Runjing Lu, William Mullins and Marina Niessner[1]

[1] J. Anthony Cookson: CU Boulder (tony.cookson@colorado.edu); Runjing Lu: University of Toronto (runjing.lu@utoronto.ca); William Mullins: UC San Diego (wmullins@ucsd.edu) and Niessner: Indiana University (mniessne@iu.edu)

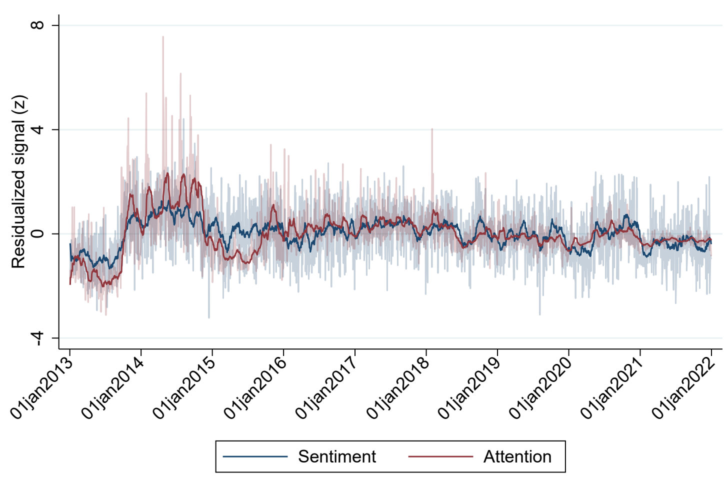
Figure A1: Time Series of Platform-Level Sentiment and Attention



(a) StockTwits



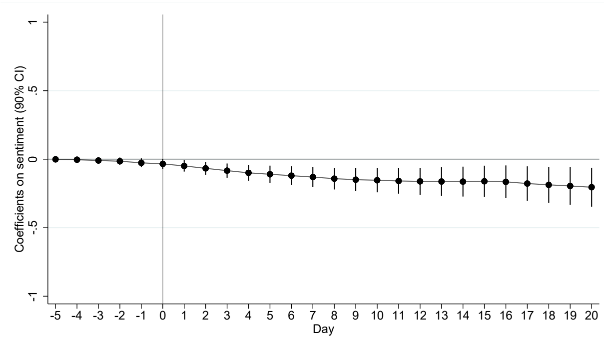
(b) Twitter



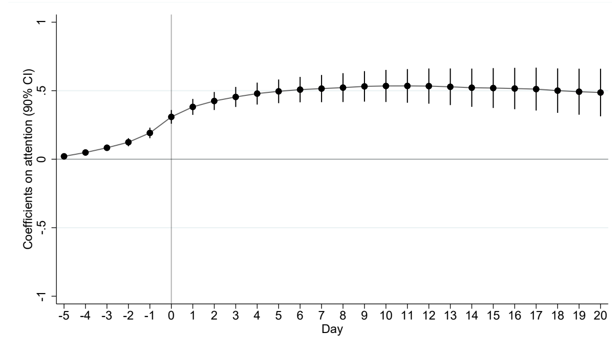
(c) Seeking Alpha

Note: This figure plots the time series for platform-level sentiment and attention signals. The lighter-colored lines plot the daily series of sentiment (blue) and attention (red) while the darker-colored lines plot the corresponding 20-day rolling average of each series. See Section 2.4 for index construction.

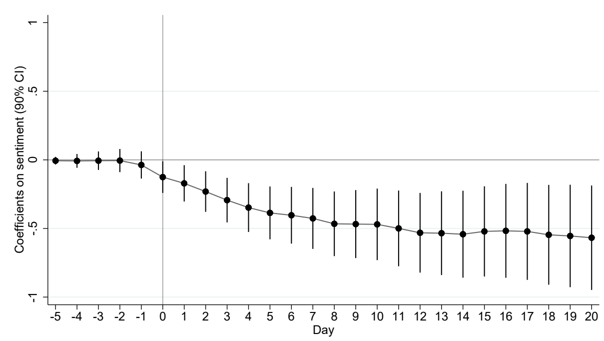
Figure A2: How Does Cumulative Abnormal *Retail* Turnover Relate to Sentiment and Attention Indexes?



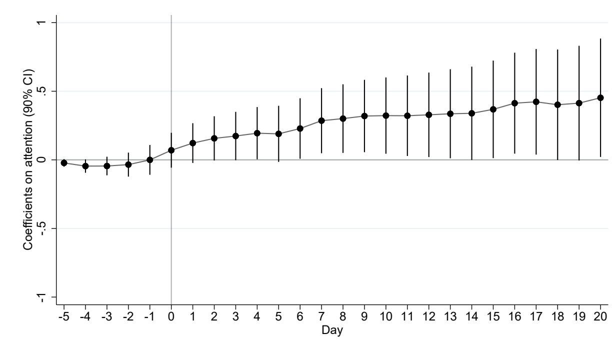
(a) Sentiment and S&P 500 turnover



(b) Attention and S&P 500 turnover



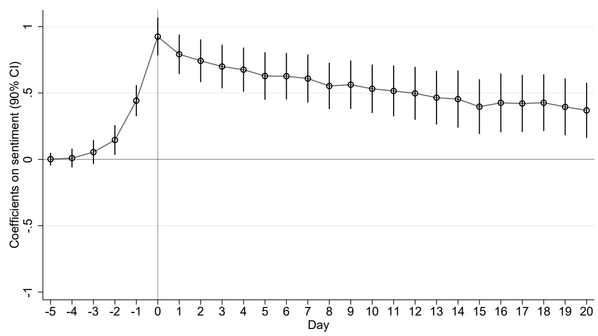
(c) Sentiment and SPY turnover



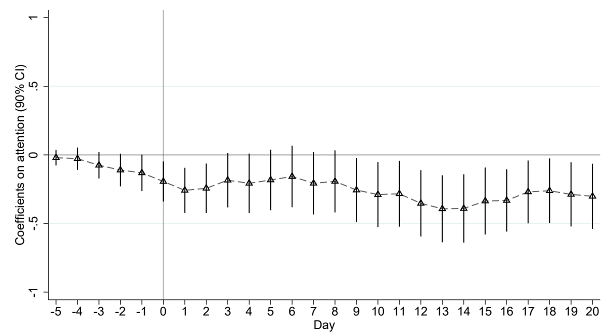
(d) Attention and SPY turnover

Note: This figure plots the estimated coefficients (and 90% confidence intervals) on sentiment and attention indexes by regressing cumulative abnormal *retail* turnover starting from day $t-5$ on sentiment index, attention index, and their interactions at day 0 for an event window between days $t=-5$ and $t=+20$. Cumulative abnormal turnover is the log turnover less the mean log turnover in the prior 140 through 20 days. S&P 500 retail turnover is the market-weighted turnover across all S&P 500 firms based on retail trading as measured in (Boehmer et al., 2021). SPY retail turnover is the turnover for the SPY index based on retail trading. Everything else follows those in Figure 3. Standard errors are calculated via Newey-West with 6 lags.

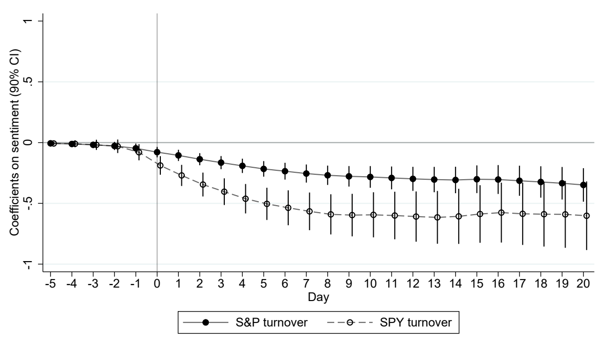
Figure A3: How Do Cumulative Returns and Cumulative Abnormal Turnover Relate to Sentiment and Attention Indexes?
With Additional Controls



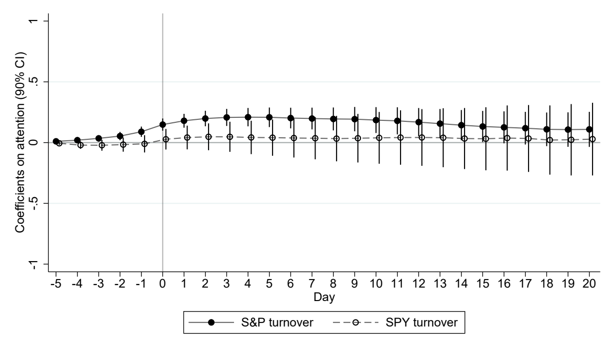
(a) Sentiment and S&P 500 return



(b) Attention and S&P 500 return



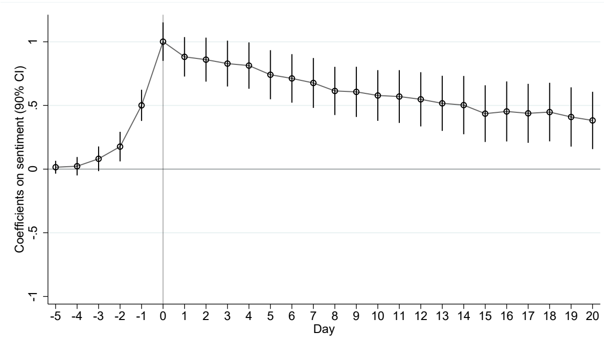
(c) Sentiment and turnover



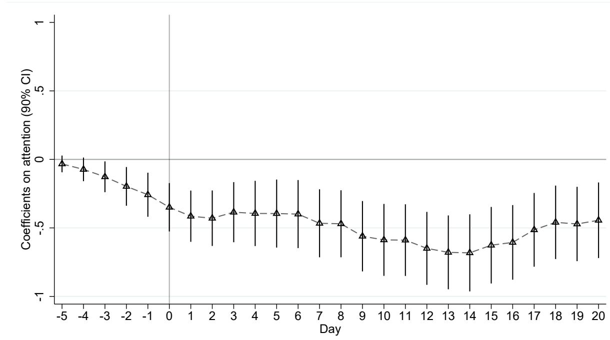
(d) Attention and turnover

Note: This figure repeats Figure 2 and Figure 3 by additionally controlling for other attention and sentiment signals (ARA, AIA, MAI - WSJ, MAI - NYT, and Twitter Economic Uncertainty). Everything else mirrors those in the original figures. Standard errors are calculated via Newey-West with 6 lags.

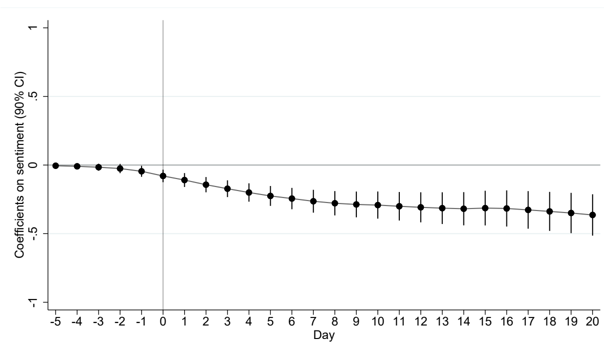
Figure A4: How Do Cumulative Returns and Cumulative Abnormal Turnover Relate to Sentiment and Attention Indexes?
Alternative Sample



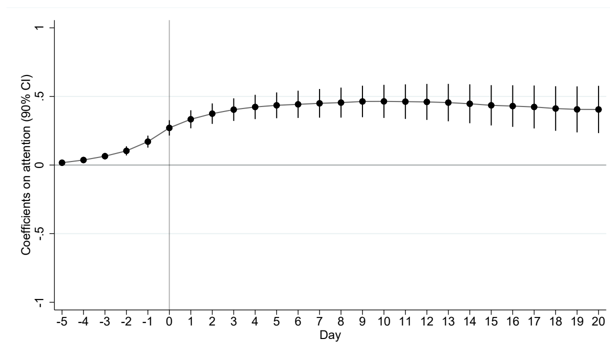
(a) Sentiment



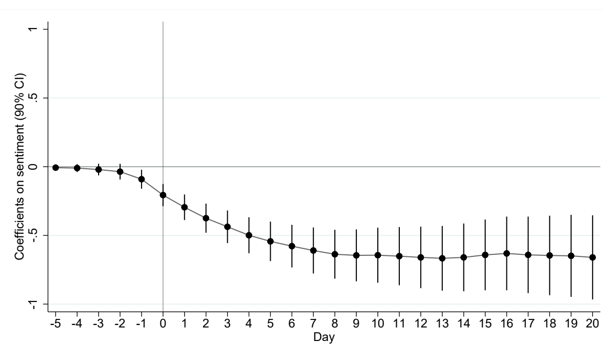
(b) Attention



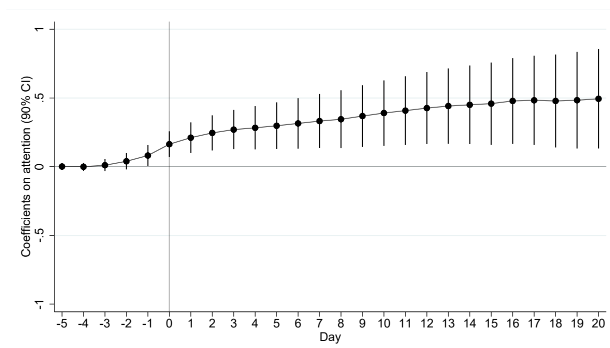
(c) Sentiment and S&P 500 turnover



(d) Attention and S&P 500 turnover



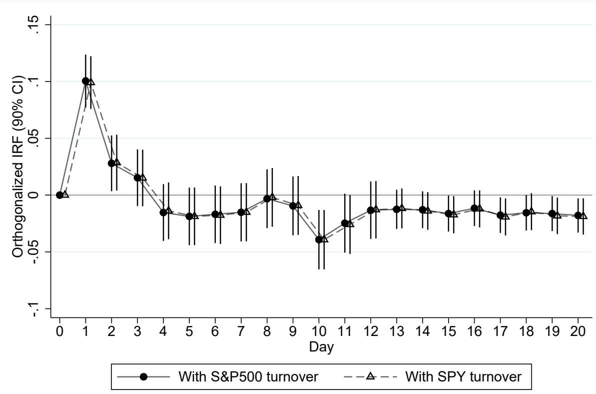
(e) Sentiment and SPY turnover



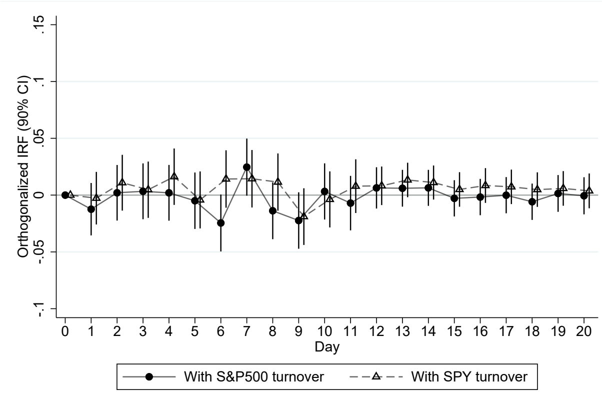
(f) Attention and SPY turnover

Note: This figure repeats Figures 2 and 3 using an alternative sample, where we focus on firm-day observations with at least 5 StockTwits posts. Everything else follows those in the corresponding tables.

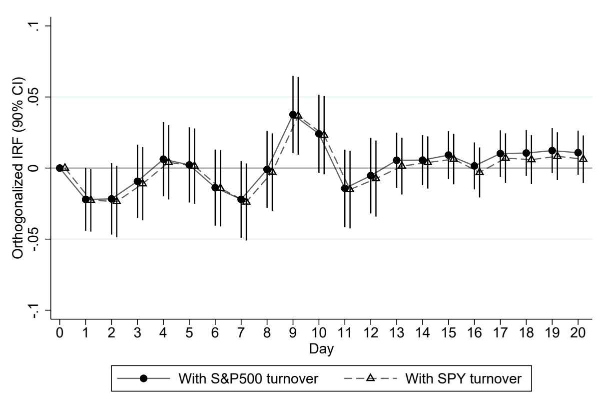
Figure A5: What Predicts Sentiment and Attention Indexes?
Impulse Response Function from a VAR Model
With Additional Attention Controls



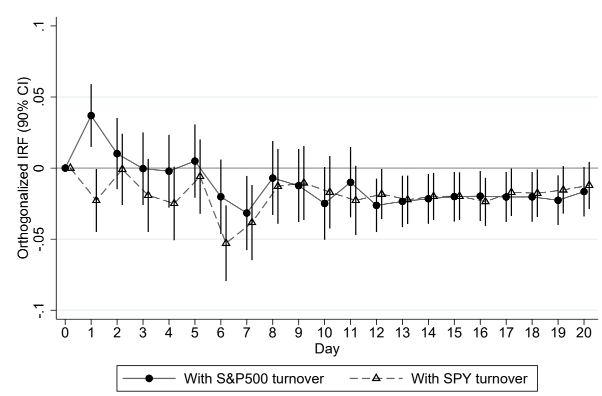
(a) Sentiment response to return



(b) Sentiment response to turnover



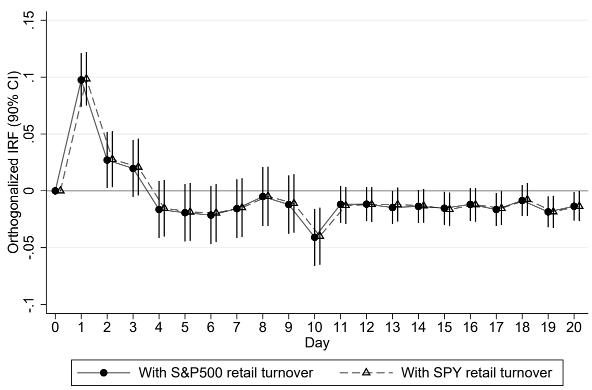
(c) Attention response to return



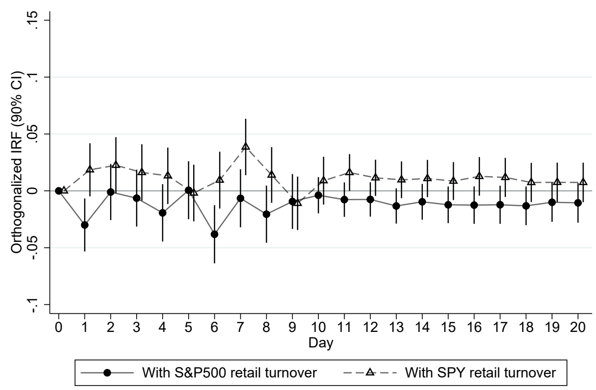
(d) Attention response to turnover

Note: This figure repeats the VAR models in Figure 6 while further including ARA and AIA. Everything else mirrors Figure 6.

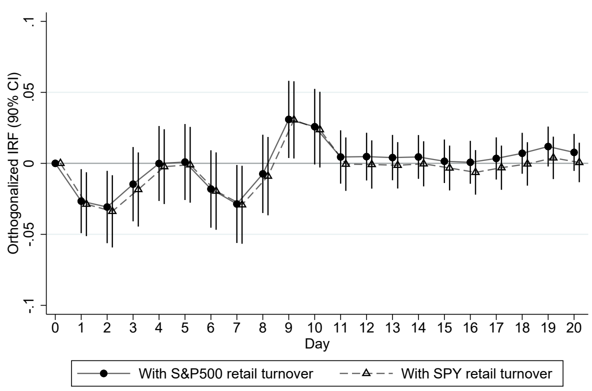
Figure A6: What Predicts Sentiment and Attention Indexes?
Impulse Response Function from a VAR Model
Turnover Based on Retail Trading



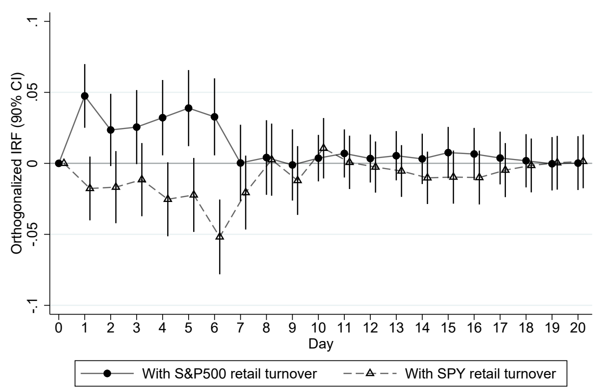
(a) Sentiment response to return



(b) Sentiment response to turnover



(c) Attention response to return

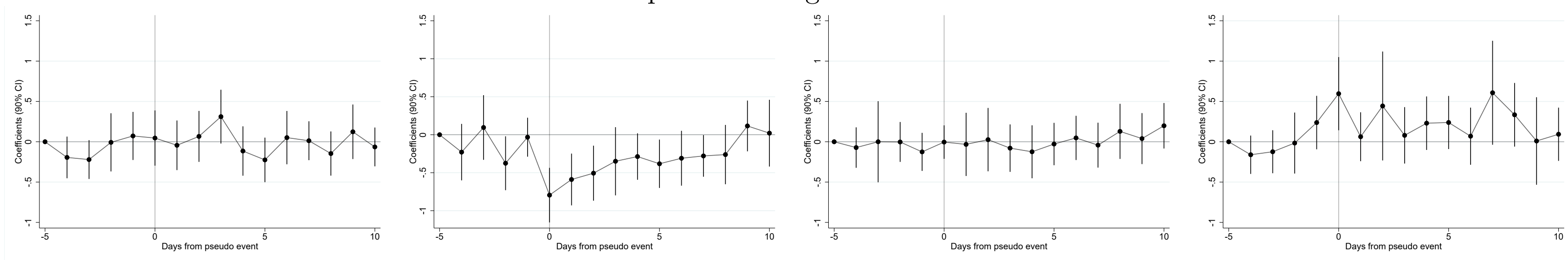


(d) Attention response to turnover

Note: This figure repeats the VAR models in Figure 6 while replacing total turnover with respective retail turnover. Everything else mirrors Figure 6.

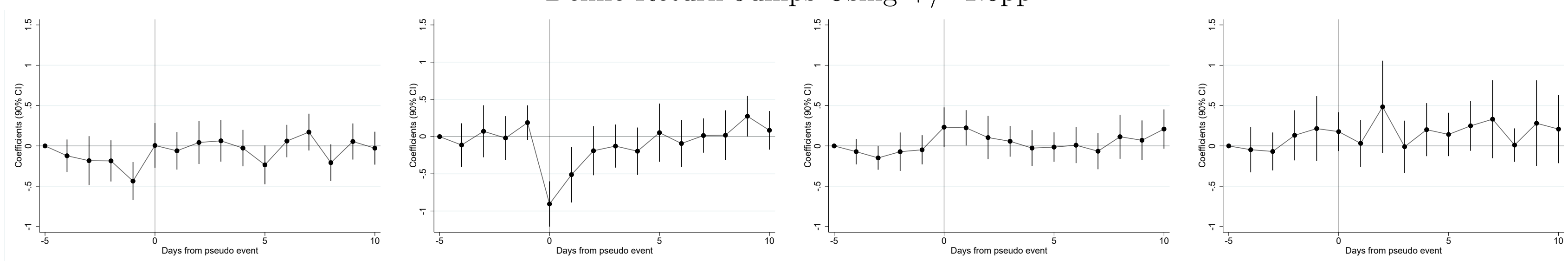
Figure A7: How Do Sentiment and Attention Indexes Change around Return Jumps
Alternative Definitions of Return Jumps

Exclude Return Jumps Coinciding with FOMC Announcements



(a) Sentiment, positive jumps **(b)** Sentiment, negative jumps **(c)** Attention, positive jumps **(d)** Attention, negative jumps

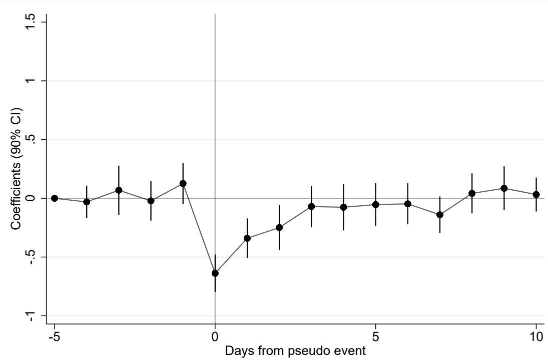
Define Return Jumps Using +/- 1.5pp



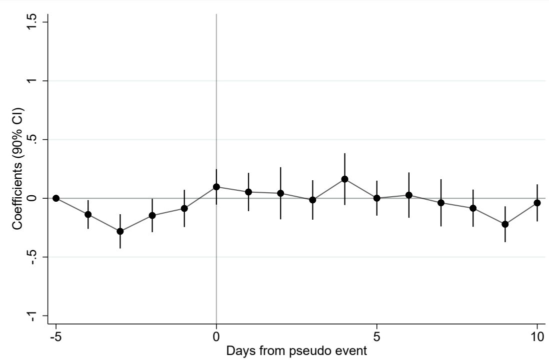
(e) Sentiment, positive jumps **(f)** Sentiment, negative jumps **(g)** Attention, positive jumps **(h)** Attention, negative jumps

Note: This figure repeats Figure 7 by using alternative definitions of return jumps. First row excludes return jumps on the same day of a FOMC meeting, and second row defines days with S&P 500 returns $\leq$ -1.5pp as negative jumps and days with S&P 500 returns $\geq$ +1.5pp as positive jumps. Everything else mirrors those in Figure 7.

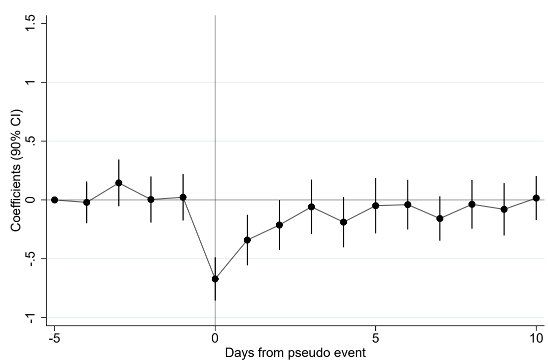
Figure A8: How Do Sentiment and Attention Indexes Change around High VIX Change?
Exclude High VIX Changes Coinciding with FOMC Announcements



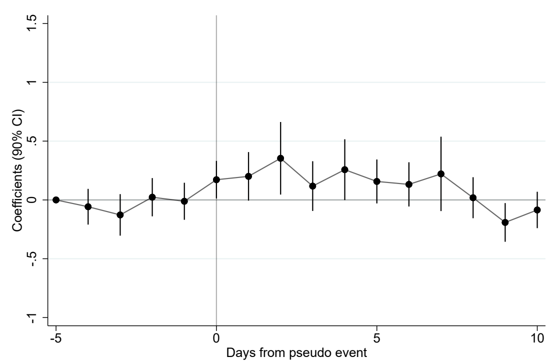
(a) Sentiment, high-Δ Vix, 15pp



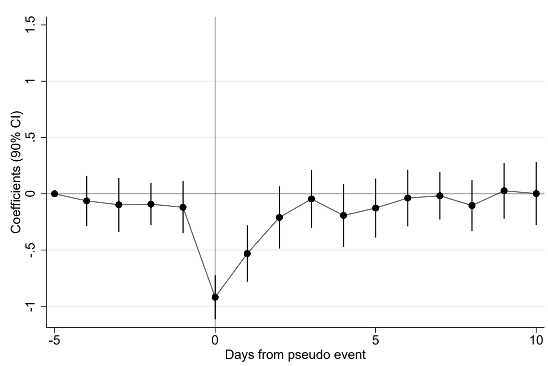
(b) Attention, high-Δ Vix, 15pp



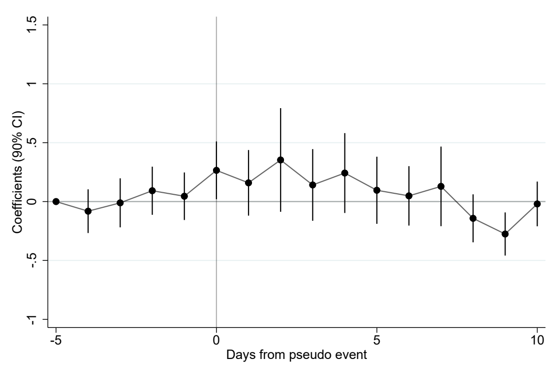
(c) Sentiment, high-Δ Vix, 20pp



(d) Attention, high-Δ Vix, 20pp



(e) Sentiment, high-Δ Vix, 25pp



(f) Attention, high-Δ Vix, 25pp

Note: This figure repeats Figure 8 while excluding high change in VIX that occurs on the same day of a FOMC meeting. Everything else mirrors those in Figure 8.

Table A1: What Predicts Social Media Sentiment and Attention Indexes?
Alternative Time Fixed Effects

	Dependent var.: Sentiment$_t(z)$		Dependent var.: Attention$_t(z)$	
	(1) S&P total	(2) SPY total	(3) S&P total	(4) SPY total
Return$_{t-1}$	0.138***	0.132***	-0.017	-0.027**
	(0.026)	(0.027)	(0.013)	(0.013)
Return$_{t-2}$	0.067***	0.064***	-0.016	-0.021
	(0.017)	(0.018)	(0.014)	(0.015)
Return$_{t-3}$	0.018	0.018	-0.004	-0.013
	(0.014)	(0.014)	(0.013)	(0.014)
Return$_{t-4}$	0.003	0.006	0.002	-0.012
	(0.014)	(0.015)	(0.014)	(0.014)
Return$_{t-5}$	0.010	0.011	0.006	-0.013
	(0.013)	(0.013)	(0.017)	(0.018)
Ab. log(turnover)$_{t-1}$	-0.137	-0.101*	0.858***	0.122**
	(0.091)	(0.055)	(0.092)	(0.057)
Ab. log(turnover)$_{t-2}$	0.061	0.056	0.005	0.033
	(0.092)	(0.055)	(0.081)	(0.052)
Ab. log(turnover)$_{t-3}$	-0.003	0.001	0.031	-0.039
	(0.086)	(0.055)	(0.073)	(0.054)
Ab. log(turnover)$_{t-4}$	-0.006	0.037	-0.034	-0.114**
	(0.107)	(0.060)	(0.076)	(0.056)
Ab. log(turnover)$_{t-5}$	-0.142	-0.053	0.101	-0.055
	(0.087)	(0.050)	(0.080)	(0.055)
DOW FE	Y	Y	Y	Y
YM FE	Y	Y	Y	Y
Observations	2267	2267	2267	2267
R^2	0.535	0.533	0.533	0.505

Note: This table repeats Table 8 while replacing MOY and YQ fixed effects with YM fixed effects. Everything else mirrors Table 8. Standard errors are calculated via Newey-West with 6 lags.
*** 1%, ** 5%, * 10% significance level

Table A2: Do Sentiment and Attention Indexes Predict *Retail* Turnover?

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+2∼20	(6) Day t+2∼20
Panel A: S&P retail turnover						
$\text{Sentiment}_t(z)$	0.000	-0.000	-0.007	-0.008	-0.093	-0.104
	(0.004)	(0.004)	(0.005)	(0.005)	(0.069)	(0.071)
$\text{Attention}_t(z)$	0.085***	0.085***	0.052***	0.052***	0.007	0.008
	(0.007)	(0.006)	(0.005)	(0.005)	(0.073)	(0.073)
Sentiment × $\text{Attention}_t(z)$		-0.007*		-0.010**		-0.097
		(0.004)		(0.004)		(0.061)
Observations	2,267	2,267	2,266	2,266	2,247	2,247
R^2	0.742	0.743	0.611	0.612	0.817	0.818
Panel B: SPY retail turnover						
$\text{Sentiment}_t(z)$	-0.078***	-0.080***	-0.057***	-0.058***	-0.207	-0.206
	(0.008)	(0.009)	(0.010)	(0.010)	(0.140)	(0.143)
$\text{Attention}_t(z)$	0.054***	0.054***	0.025**	0.025**	0.162	0.162
	(0.009)	(0.009)	(0.010)	(0.010)	(0.165)	(0.166)
Sentiment × $\text{Attention}_t(z)$		-0.010		-0.008		0.005
		(0.007)		(0.009)		(0.119)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.627	0.627	0.533	0.533	0.695	0.695
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y

Note: This table reports how sentiment and attention indexes predict day t, day t+1, and day t+2∼t+20 retail turnover. Each panel represents a different outcome: panel A S&P 500 cumulative abnormal retail turnover and panel B SPY cumulative abnormal retail turnover. Everything else mirror those in Table 4. Standard errors are calculated via Newey-West with 6 lags.
*** 1%, ** 5%, * 10% significance level

Table A3: Do Sentiment and Attention Indexes Predict Returns and Turnover?
Year-Month Fixed Effects

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+2∼20	(6) Day t+2∼20
Panel A: S&P return						
Sentiment$_t(z)$	0.583***	0.609***	-0.141***	-0.144***	-0.430***	-0.440***
	(0.043)	(0.045)	(0.037)	(0.040)	(0.074)	(0.080)
Attention$_t(z)$	-0.077***	-0.076***	-0.032	-0.032	0.145	0.145
	(0.026)	(0.029)	(0.032)	(0.032)	(0.109)	(0.108)
Sentiment × Attention$_t(z)$		0.190***		-0.022		-0.074
		(0.040)		(0.036)		(0.093)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.222	0.245	0.085	0.086	0.668	0.668
Panel B: S&P turnover						
Sentiment$_t(z)$	-0.021***	-0.022***	-0.016***	-0.017***	-0.101***	-0.112***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.038)	(0.039)
Attention$_t(z)$	0.075***	0.075***	0.040***	0.040***	-0.146***	-0.146***
	(0.008)	(0.007)	(0.006)	(0.005)	(0.043)	(0.043)
Sentiment × Attention$_t(z)$		-0.010*		-0.012**		-0.079**
		(0.005)		(0.005)		(0.035)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.616	0.617	0.517	0.519	0.893	0.894
Panel C: SPY turnover						
Sentiment$_t(z)$	-0.087***	-0.090***	-0.060***	-0.062***	-0.163**	-0.175**
	(0.009)	(0.009)	(0.010)	(0.010)	(0.070)	(0.073)
Attention$_t(z)$	0.059***	0.059***	0.020**	0.020**	-0.225***	-0.225***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.086)	(0.086)
Sentiment × Attention$_t(z)$		-0.021***		-0.022**		-0.090
		(0.007)		(0.009)		(0.065)
Observations	2,267	2,267	2,266	2,266	2,267	2,267
R^2	0.653	0.654	0.576	0.578	0.881	0.881
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y
YM FE	Y	Y	Y	Y	Y	Y

Note: This table reports how sentiment and attention indexes predict day t, day t+1, and day t+2∼t+20 S&P 500 cumulative return and cumulative abnormal turnover. Panel A follows Table 3 and panels B and C follow Table 4 panels A and B, except that we replace MOY and YQ fixed effects with Ym fixed effects. Standard errors are calculated via Newey-West with 6 lags.
*** 1%, ** 5%, * 10% significance level

Table A4: Do Sentiment and Attention Indexes Predict Return and Turnover? *With Additional Controls*

	(1)	(2)	(3)	(4)	(5)
Panel A: Return					
$\text{Sentiment}_t(z)$	0.544***	0.542***	0.543***	0.543***	0.539***
	(0.042)	(0.043)	(0.042)	(0.042)	(0.042)
$\text{Attention}_t(z)$	-0.097***	-0.110***	-0.064**	-0.100***	-0.092***
	(0.030)	(0.032)	(0.030)	(0.030)	(0.030)
$\text{Sentiment} \times \text{Attention}_t(z)$	0.158***	0.161***	0.153***	0.156***	0.158***
	(0.038)	(0.038)	(0.037)	(0.038)	(0.038)
$\text{ARA}_t(z)$		0.031			
		(0.038)			
$\text{AIA}_t(z)$			-0.125**		
			(0.055)		
$\text{MAI (WSJ)}_t(z)$				-0.046*	
				(0.028)	
$\text{MAI (NYT)}_t(z)$				0.055	
				(0.036)	
$\text{Twitter EU}_t(z)$					-0.080***
					(0.028)
Observations	2,267	2,267	2,267	2,267	2,267
R^2	0.192	0.192	0.199	0.195	0.196
Panel B: S&P turnover					
$\text{Sentiment}_t(z)$	-0.021***	-0.023***	-0.021***	-0.021***	-0.020***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
$\text{Attention}_t(z)$	0.071***	0.057***	0.055***	0.071***	0.069***
	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
$\text{Sentiment} \times \text{Attention}_t(z)$	-0.007	-0.003	-0.004	-0.006	-0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
$\text{ARA}_t(z)$		0.034***			
		(0.008)			
$\text{AIA}_t(z)$			0.071***		
			(0.006)		
$\text{MAI (WSJ)}_t(z)$				0.016***	
				(0.004)	
$\text{MAI (NYT)}_t(z)$				0.007*	
				(0.004)	
$\text{Twitter EU}_t(z)$					0.027***
					(0.003)
Observations	2,267	2,267	2,267	2,267	2,267
R^2	0.597	0.607	0.642	0.601	0.606
Panel C: SPY turnover					
$\text{Sentiment}_t(z)$	-0.080***	-0.081***	-0.080***	-0.079***	-0.077***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
$\text{Attention}_t(z)$	0.054***	0.043***	0.033***	0.054***	0.051***
	(0.009)	(0.010)	(0.008)	(0.009)	(0.009)
$\text{Sentiment} \times \text{Attention}_t(z)$	-0.010	-0.007	-0.006	-0.009	-0.010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$\text{ARA}_t(z)$		0.025**			
		(0.010)			
$\text{AIA}_t(z)$			0.082***		
			(0.009)		
$\text{MAI (WSJ)}_t(z)$				0.028***	
				(0.007)	
$\text{MAI (NYT)}_t(z)$				0.001	
				(0.007)	
$\text{Twitter EU}_t(z)$					0.049***
					(0.007)
Observations	2,267	2,267	2,267	2,267	2,267
R^2	0.627	0.629	0.645	0.631	0.637
Controls	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y

Note: This table repeats Table 3 and Table 4 while including additional attention and sentiment measures: ARA, AIA, MAI (WSJ), MAI (NYT), and Twitter EU index. Everything else follows those in column 2 of the corresponding tables. Standard errors are calculated via Newey-West with 6 lags.
*** 1%, ** 5%, * 10% significance level