Market Signals from Social Media

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

March 20, 2025

Abstract

This paper develops daily market-wide sentiment and attention indexes derived from millions of posts across major investor social media platforms. We find that sentiment extrapolates from past market-wide returns and exhibits a strong reversal. In contrast, attention predicts negative returns as a continuation of previous trends. The two indexes have distinct predictions for aggregate trading: abnormal trading rises when sentiment is low and attention is high. To identify the drivers of attention and sentiment, we use a shock to data sharing networks: We find sentiment spreads through real firm connections while attention does not. Moreover, attention rises after abnormally high trading, while sentiment rises after abnormally high returns. This extrapolative return pattern is asymmetric, primarily driven by negative market jumps. These findings provide new evidence on the daily market dynamics of sentiment and attention.

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 Azi Ben-Rephael, Joey Engelberg, Charles Martineau, and Enrichetta Ravina for their comments. We thank Zhi Da, Shaun Davies, Lin Peng, Ran Shi, Huan Tang and Nancy Xu for providing both data and comments. This draft has benefited from seminar and conference feedback at Rutgers, Chinese University of Hong Kong, CU Boulder, and the 2025 Front Range Finance Seminar. Remaining errors are our own.

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 they are economically distinct concepts. Consistent with this distinction, we find returns rise prior to high *sentiment* days, which is followed by a reversal over the next 20 days. By contrast, returns decline prior to high *attention* days, and this is 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: lagged *returns* predict sentiment, while lagged abnormal *trading volume* predicts attention. These findings highlight the importance of understanding these higher frequency patterns, as well as distinguishing 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 1.2. 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). The trading strategy is especially profitable on days when the market performs poorly: portfolio excess returns are two-thirds larger after days when the market declines by 1%. Imposing long-only and 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 — which is 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 then investigate the drivers of these sentiment and attention indexes. In both OLS regressions and in vector autoregressions — which accounting 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 both downward jumps in market returns 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).

We conclude by examining the role of network effects in driving market-wide sentiment and attention. Using the data sharing network developed by Bian et al. (2025) to capture real economic linkages across firms, we test whether sentiment and attention from central firms spill over to other firms. Exploiting the Apple Tracking Transparency (ATT) policy — which disrupted firm-to-firm data sharing by requiring user consent — as a shock to network connections, we find that the relationship between central firm sentiment and overall market sentiment significantly weakens after ATT policy implementation. This suggests that sentiment spills over along company networks, contributing to market-wide sentiment. In contrast, we find no significant change in the relationship between central firm attention and overall market attention, indicating that market-wide attention emerges through more decentralized mechanisms. These findings further emphasize the differing dynamics driving sentiment and attention, reinforcing the value of analyzing them separately.

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,

[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.

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 and the construction of the aggregate sentiment and attention indexes.

2.1 Firm-day social media sentiment and attention data

Our data contain firm-day observations on social media sentiment (bullishness vs. bearishness) 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 ("significantly relevant"). We use Ravenpack's Event Sentiment Score (ESS) to measure Seeking Alpha sentiment. To avoid posts by bots, we drop users who post over 1,000 times in any single day.

For StockTwits and Seeking Alpha, we measure sentiment about firm i on day t, by averaging sentiment over all posts or articles about the firm from market close (4:00 pm) on day $t-1$ to market close on day t. The resulting firm-day sentiment are measured over the same time period as 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 time period.

2.2 Other data

Our data on firm-related news events covered in traditional media is from the *Dow Jones Newswire*. Ravenpack 1.0 provides article-level sentiment and the number of articles by firm-day. We include only 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. We also 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 each firm-day observation to have at least 10 posts on StockTwits. Table 1 reports summary statistics across the three social media platforms. While all the platforms offer comparable firm coverage, Seeking Alpha has substantially fewer firm-day observations and posts per day, reflecting its likely due to its long-form content format. StockTwits has over five times the number of daily posts relative to Twitter.

StockTwits and Twitter provide daily coverage of firms comprising around 90% of the market capitalization in our sample. After restricting the sample to firm-day observations with at least 10 posts on StockTwits to ensure a high quality social signal, coverage remains strong at approximately 50% of market capitalization. While Seeking Alpha offers less breadth, it still covers nearly 40% of market capitalization.

2.4 Constructing aggregate indexes from firm-day information

To construct daily indexes of *market-level* sentiment and *market-level* 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.[2]

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

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

where $Signal_{i,t}^{P}$ is either attention or sentiment on 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. $\overline{Signal}_{i,-y(t)}^{P}$ 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 the 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.

[2]We obtain similar findings if we do not purge the news and firm-specific components from these signals. We also obtain similar, but noisier, results if we equal weight instead of value weight. See Appendix Tables A7 and A8.

For the final step, we combine the resulting platform-day signals into daily indexes of aggregate sentiment and attention, by performing two separate 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) from each analysis 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 attention PC1 explains 54% of the variation in attention. To put these in perspective, if the three signals were completely uncorrelated, the PC1 would only explain 33% of the variation. Both sentiment and attention signals place similar loadings on StockTwits and Twitter. For sentiment, the Seeking Alpha loading is approximately half the size of the other platforms' loading. By contrast, the attention index places close to no weight on Seeking Alpha, likely because its article volume is substantially lower than the number of posts on StockTwits and Twitter.

Our sentiment and attention indexes contain unique variation relative to existing daily measures of attention and sentiment in the literature, while having sensible relationships that help validate our construction. To examine these relationships, we regress our social media indexes on existing daily indicators: abnormal retail attention (*ARA*) and abnormal institutional attention (*AIA*) from Da et al. (2024), news attention indexes (*MAI-WSJ, MAI-NYT*) from Fisher et al. (2022), Twitter-derived economic policy uncertainty (*Twitter EU*) from Baker et al. (2021), and a market cap weighted news sentiment index, constructed from RavenPack (*RavenPack news*). Some specifications, also include our sentiment index when predicting the attention index and vice versa, as well as day-of-week, month-of-year, and year-quarter fixed effects to account for potential seasonality and within-week variation.

Table 3 presents the results. Columns 1-2, without fixed effects, show that existing indicators explain only a small portion of the variation in our sentiment and attention indexes—they collectively explain only 2.8% for sentiment and 18.2% for attention. This suggests that our measures capture information not reflected in existing indicators.

The correlations that we observe help to validate our measures. For example, Twitter EU is negatively and significantly related to our sentiment index across all specifications, indicating that periods of high economic policy uncertainty correspond to lower sentiment. Similarly, our attention index correlates strongly and positively with the attention indexes from Da et al. (2024), which are based on Google searches and Bloomberg queries. Notably, while both professional and retail attention measures correlate with our social media-derived attention index, the coefficient on ARA (a retail attention measure) is twice as large as that for professional attention. This pattern suggests our social media attention index primarily captures retail investor attention.

The even columns in Table 3 include our sentiment index as an explanatory variable when analyzing the attention index (and vice versa). We find a robust negative relation between sentiment and attention, but there is substantial variation in sentiment that is unexplained by attention, as well as attention that is unexplained by sentiment. Moreover, the significant correlation between sentiment and Twitter EU persists even after controlling for attention, and the significant correlations between attention and both ARA and AIA remains robust even when controlling for sentiment. These findings confirm that sentiment and attention contain independent variation, supporting their separate use in our empirical analyses.

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. For reference, we also include the level of the S&P 500 index. Focusing on the lower-frequency movements, the sentiment and attention indexes appear to capture different economic forces, with a correlation of -0.37. This divergence is particularly evident during three key periods: the 2013-2015 stock market bull run, in the 2018-2019 trade war with China, and the onset of the COVID-19 pandemic. During the bull market period, sentiment was high while attention was low; conversely, during the two negative events (trade war and pandemic), low sentiment was coupled with high attention.

While the slower-moving signals are easier to visualize and highlight, our paper focuses on the information contained in the *daily* movements of these indexes. As the figure illustrates, 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 relation between our sentiment and attention indexes and subsequent returns. As a preview, in Figure 2, we present coefficient estimates from regressions of cumulative returns on day zero sentiment and attention for an event window from $t = -5$ to $t = +20$. The figure shows that the return dynamics around high sentiment (panel a) differ from those around high attention (panel b). High sentiment on day zero 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 high attention on day zero is preceded by *negative* returns, which continue on a downward trajectory on the following day and exhibit no reversal.

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 illustrate that the predictive content of these market-level social media signals does not reflect forward-looking information. Section 4 then examines the drivers of these indexes.

3.1 Do social media indexes predict returns or turnover?

Here we examine how daily sentiment and attention indexes predict returns using the following regression specification:

$$\text{Market return}_{t\to k} = \beta_1 Sentiment_t + \beta_2 Attention_t + \beta_3 (Sentiment \times Attention)_t + \Lambda_t + \epsilon_t \quad (2)$$

where Market return$_{t\to k}$ 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 $t-1$ as well as the preceding 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 on day $t-1$.[3]

Table 4 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 $t+15$. Table 4 also shows that attention is negatively related to returns on days t and $t+1$, but the relationship dissipates for days $t+2$ through $t+20$. In addition, the specifications in columns 2, 4, and 6 also include an interaction between sentiment and attention, which is positively and significantly related to day t returns with no reversal.

Table 5 investigates our second main outcome: turnover. Specifically, it shows 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 4. These results show that sentiment and attention indexes have the opposite relationship with

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

turnover compared to their relationship with returns. Specifically, high attention is contemporaneously related to *high* turnover, whereas high sentiment accompanies *low* turnover. Moreover, the dynamics differ: following high attention, abnormal trading continues to increase, and following high sentiment abnormal trading volume continues to decrease. Figure 3, which presents 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 tests 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 (Appendix Figure A2 and Table A5). Second, we examine the sensitivity of our results to the data quality requirement that we retain only firm-day observations with at least 10 StockTwits messages. In Appendix Figure A4, we require only 5 or more messages per firm-day, and obtain very similar findings.

3.2 Accounting for alternative drivers of return and turnover

In this subsection, we conduct two additional sets of robustness tests on our main analysis. First, in Appendix Table A6, we replace month-of-year and year-quarter fixed effects with year-month fixed effects; we obtain similar results. Year-month fixed effects flexibly control for slow moving factors that could jointly drive returns, sentiment, and attention – particularly for existing sentiment indexes measured at the monthly frequency (e.g., the Baker and Wurgler (2006) index). The similarity of the results when using within year-month variation indicates that our sentiment and attention indexes contain distinct information from existing alternative measures in the literature.

Second, in Appendix Table A9, we control for daily attention indexes from recent literature. Specifically, Da et al. (2024) develops 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) construct 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. Finally, to further capture aggregate firm news, we build a RavenPack news sentiment index by value weighting firm-level news for firms in our sample. As shown in Appendix Table A9, our results remain robust to the inclusion of these alternative proxies for attention or sentiment. Furthermore, in Appendix Figure A3 we replicate our results from Figures 2 and 3 with these additional controls, and again 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 subsection, we implement a dynamic trading strategy based on the daily aggregate sentiment and attention indexes to ensure that the return patterns we show are not driven by future information. First, we use information up to the 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 a 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} (the loadings for Ω_t^{m-1}). We then predict next-day returns for each trading day in month m:

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

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 (yielding 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 (representing a 100% short position) and 2 (representing 100% leverage). The portfolio return remains similar if we prohibit short selling and leverage by restricting w_t^{social} to be 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. Appendix Figure A7 displays the time series of portfolio weights of the risky asset, showing that the restriction that the portfolio weights must be between -1 and $+2$ rarely binds. 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 (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 6 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 statistically 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 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 6 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. This suggests our strategy is not driven by short-selling or leverage.[4]

Next, we examine whether the significant excess returns from the dynamic strategy yield *abnormal* returns beyond the Fama-French and momentum factors in Table 7. In columns 2 through 4 (where the factor returns are measured on day $t+1$) 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 positive and significant loadings — the market factor alone explains 15.2% of portfolio returns while reducing the magnitude of alpha by just one-third.

[4]Appendix Table A11 presents two robustness tests of these portfolio results: In Panel A, we winsorize the forecast returns from Eq. (4) at the 90% and 10% percentiles before constructing the portfolio weights as in Da et al. (2024), and in Panel B, we smooth portfolio weights using a trailing 5-day moving average rather than using weights directly from Eq. (5). These modifications reduce the annualized alpha to 3.45% and 3.8% respectively, but the Sharpe ratios remain above one (1.137 and 1.078).

In summary, these findings indicate that the market and momentum factors explain approximately one-third of the average portfolio excess returns.

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 subsection 4.1.

In subsequent subsections, we examine dynamic interdependence and feedback effects between variables over time. Specifically, we evaluate the drivers of sentiment and attention through three complementary analyses: (1) a vector autoregression (VAR) examining how returns and turnover drive the indexes, (2) an event study of index responses to abrupt changes in prices and volatility — *jumps* — and (3) an analysis of how spillovers of sentiment and attention via firm networks contribute to market-level social media signals.

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 the day t sentiment index. This pattern is consistent with extrapolative beliefs documented in previous work (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 A3, which flexibly accounts for any monthly-level variability, including effects from 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), compared to day $t-1$ turnover in the SPY ETF (column 4). This suggests that our attention index captures at least some of the dispersed information impounded in market trading (Hayek, 1945). Overall, 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 among sentiment, attention, market returns, and abnormal market turnover. Following Sims (1980), we summarize the properties of the fitted VAR model by examining the impulse responses to one standard deviation shocks in Market return$_t$ and Log(Ab. market turnover)$_t$.

Figure 5 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 the component stocks of the S&P500 (solid line) and abnormal turnover in the SPY (dashed line). Consistent with the 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 a response in sentiment, but increases attention for the next several days. Notably, a one-standard-deviation abnormal turnover shock in SPY does not lead to higher attention.

We next conduct two robustness tests on these results. First, we account flexibly for alternative attention measures in the VAR. Adding a daily retail attention index and an institutional attention index (from 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 respond 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 abnormal retail turnover.

4.3 Sentiment and attention around jumps

We next examine how our indexes behave around *jumps* in price and volatility. We classify market jumps as days that had at least a 2 percentage point change in returns (either positive or negative), and volatility jumps as 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. To ensure non-overlapping windows, if there are multiple jumps in a sequence, we only classify the first as a jump event.

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

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

where Social media 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 separately estimate 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 event in the sample. By only estimating leads and lags from $t-4$ to $t+10$, we fix the reference period as 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 preceding 25 days).

Figure 6 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 a marked asymmetry in the response of sentiment and attention to positive versus negative market jumps. Specifically, positive jumps (good news) associated with significant changes in sentiment or attention. In contrast, negative jumps lead to a sharp and persistent drop in sentiment and an increase 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 A8).

Figure 7 presents an analogous set of results for volatility jump days. Large spikes in the VIX can be interpreted as negative news, similar to negative market jumps, but there is no clear VIX analogue to 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 A9).

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:

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

where Social media index is either the sentiment or attention index. This specification includes event day indicators (α_l) for days -1, 0, and +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 days -15 to -2. 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 preceding 25 days). In more demanding specifications, we also control for changes in the 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 to 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 6, we estimate that attention rises significantly on negative jump days but 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).

4.4 Sentiment and attention of central firms

In this subsection, we examine a hypothesis for the origin of market social media sentiment and attention — the idea that these signals originate from shocks to firms that spill over along firms' economic networks. While firm networks have many dimensions, here we focus on a data sharing network among firms. Bian et al. (2025) shows that data sharing constitutes a real economic network that induces greater similarity among connected firms. In addition to presenting evidence that connectedness and real outcomes are correlated, Bian et al. (2025) leverages the introduction of the Apple Tracking Transparency (ATT) policy, which — with the rollout of iOS 14.2 on April 26, 2021 — required user consent for firm-to-firm data sharing, effectively disrupting the connections within this data sharing network.

Here, we explore whether sentiment and attention spill over from central firms to other firms in the economy, and in the process contribute to market-level social media signals. To test this idea, we define central firms as those ranking in the top 20 in 2019 based on one of three measures: eigenvector centrality, betweenness centrality, and degree centrality. We do this using data generously provided by the authors of Bian et al. (2025). All other firms are classified as non-central. We use the ATT policy as a negative shock to network connections, which disrupts the network's spillovers in the data economy.

We estimate the following specification:

$$\begin{aligned}\text{Social media index}_{t+k}^{all} \quad &= \quad \beta_0 Post\ ATT_t + \beta_1 \text{Social media index}_t^{central} + \\ &\quad \beta_2 Post\ ATT_t \times \text{Social media index}_t^{central} + \Lambda_t + \epsilon_t \qquad (9)\end{aligned}$$

where Social media index$_{t+k}^{all}$ is the overall sentiment or attention index on day $t+k$ (where $k = 0, 1, 2$), constructed using all firms in our sample. Social media index$_t^{central}$ is the central firm sentiment or attention index, which is constructed using only central firms. We expect central firm social media signals to exhibit a strong positive relation to overall social media signals, thus $\beta_1 > 0$. Our coefficient of interest is β_2, which captures the change in the relation between central firm and overall social media signals after the ATT policy disrupted the network. If social media signals spill over across the data network from central to non-central firms, we would expect the relation to weaken after the ATT policy, and hence $\beta_2 < 0$. Λ_t consists of day-of-week and event quarter fixed effects. Additionally, we consider a version of this specification that uses the index constructed using only *non-central* firms as the dependent variable.

Table 10 presents the results from estimating Eq. (9) over the period from May 2020 through December 2021, which balances pre and post ATT months. Panel A focuses on sentiment; as expected, β_1 is strongly positive across all specifications — whether the dependent variable is sentiment or attention index, and whether the index is based on all or non-central firms only — reflecting the fact that central firms are an important component of the overall indexes and are related to the sentiment and attention of other firms. However, the significant negative interactions between the Post ATT indicator and central firm sentiment in Panel A columns 1-3 indicate that shutting off the transmission of data along the data network significantly dampens the connection between central firm and overall sentiment. When we use non-central firm sentiment as the dependent variable in Panel A columns 4-6, we find that the coefficient on the interaction is very similar. These findings point to important spillovers from central firms to non-central firms along the data sharing network, which in turn contribute to market-wide sentiment. Table A13 shows that these results are robust when we examine simple correlations pre versus post the ATT policy.

Turning to attention in Panel B, we do not see a significant change in the relation between central firm and overall attention following the ATT policy. This finding suggests that market-wide attention is aggregated differently to sentiment. Rather than attention spilling over from central firms to other firms, it likely emerges in a more decentralized manner. This interpretation is consistent with our findings that attention is more strongly connected to disaggregated S&P500 turnover than turnover in SPY, the most widely traded S&P500 ETF. Consistent with the other results in this study, this finding underscores that sentiment and attention reflect different underlying dynamics, reinforcing the importance of disaggregating these forces into separate measures.

Lastly, we note that the relation between central firm and overall (or non-central firm) indexes in Table 10 weakens by day $t+2$ for attention, but remains strong for sentiment. In Figure 8, we show how the coefficient estimates change when we extend k further into the future. Over time, both sentiment and attention coefficients decrease to zero.

5. Conclusion

Social media has become a conduit of our collective attention. Recent market events like GameStop and the Silicon Valley Bank run highlight social media's role as a key venue for expressing sentiment about market events. This paper leverages the rise of investor social media to develop daily social media-based sentiment and attention indexes.

As well as predicting future returns, the indexes provide novel perspectives on both the *economic content* and the *timing* of changes to market-wide sentiment. Regarding economic content, we show that attention and sentiment have sharply differing return dynamics. A key advantage of the social media setting is that it allows for a natural separation between these two concepts. This is important both conceptually and empirically, because existing sentiment indexes often conflate the two. Regarding timing, our daily indexes capture daily market dynamics missed by most existing research using monthly sentiment indexes. This higher frequency variation is particularly relevant given the sustained increase in daily trading activity (SIFMA, 2024), 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 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.

Bian, Bo, Qiushi Huang, Ye Li, and Huan Tang, 2025, Data as a networked asset, *Working Paper* .

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.

Hodrick, Robert J, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357–386.

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.

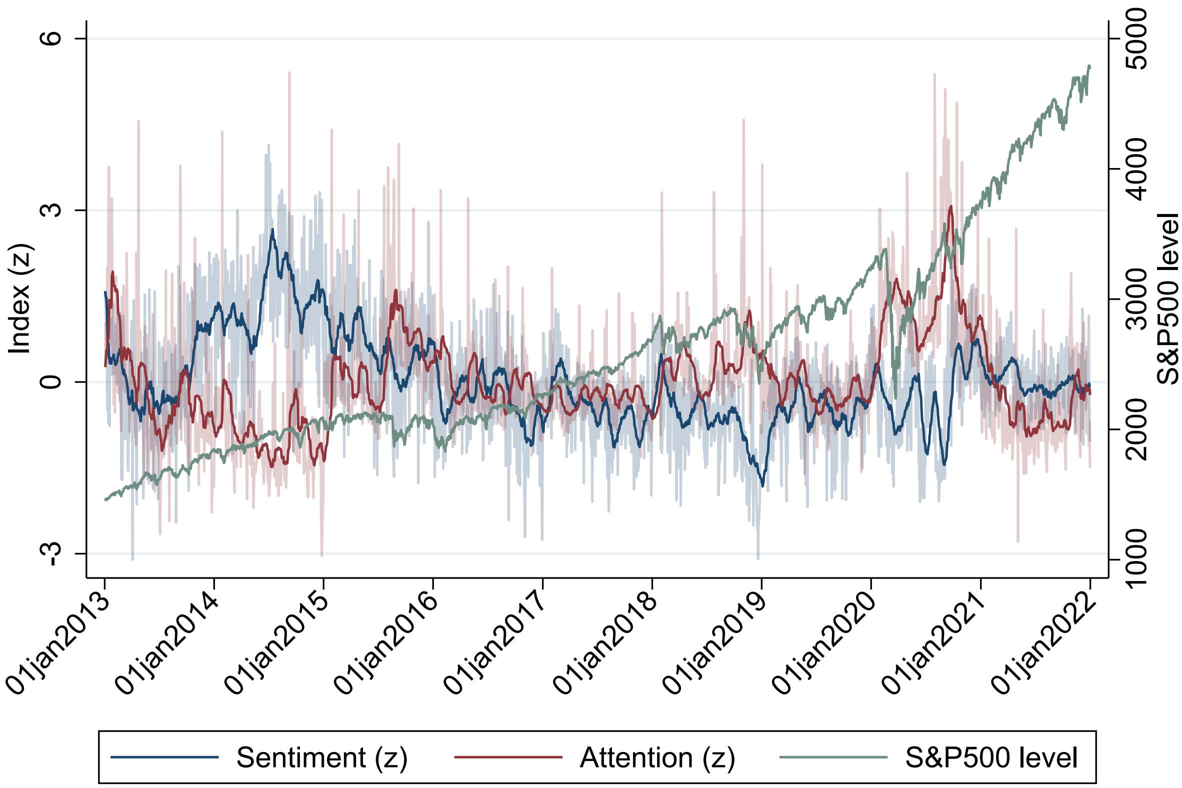
Odean, Terrance, 1999, Do investors trade too much?, *American economic review* 89, 1279–1298.

SIFMA, 2024, 2024 Capital Markets Fact Book, https://www.sifma.org/wp-content/uploads/2023/07/2024-SIFMA-Capital-Markets-Factbook.pdf (accessed March 19, 2025).

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

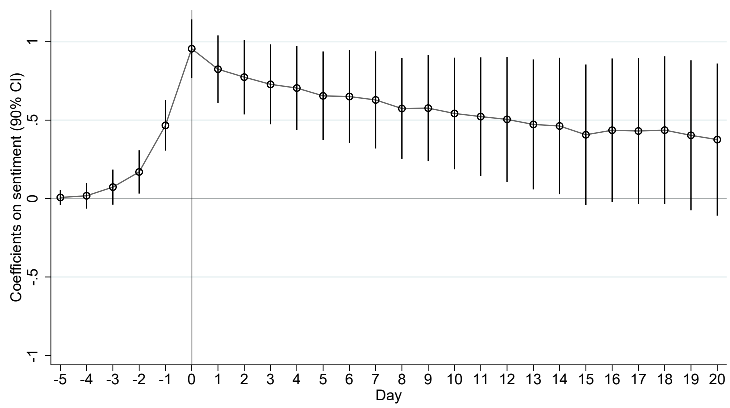
Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Figure 1: Time series of sentiment and attention indexes

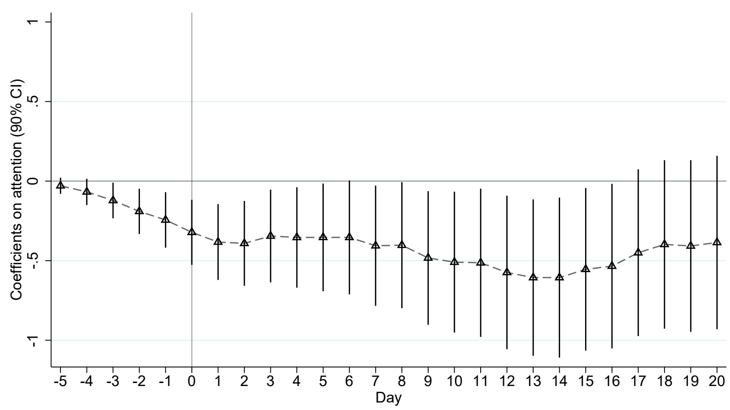


This figure plots the time series of the sentiment index (blue) and the attention index (red) benchmarked against the S&P 500 index level (green). The lighter-colored lines plot the daily series of sentiment and attention indexes, while the darker-colored lines plot the corresponding 20-day rolling average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

Figure 2: How cumulative returns relate to sentiment and attention indexes



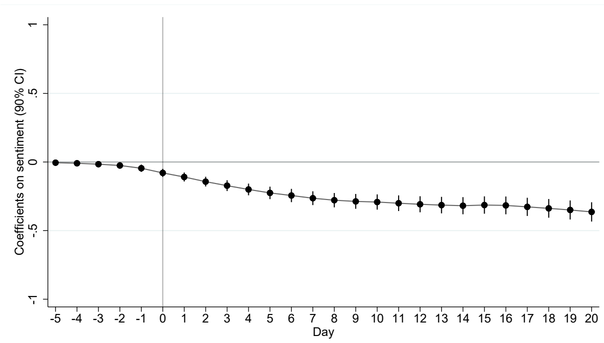
(a) Sentiment



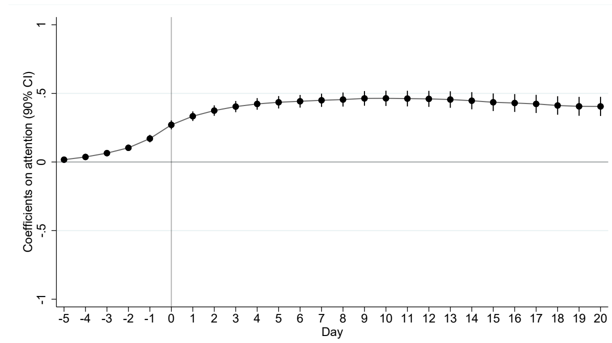
(b) Attention

This figure plots the estimated coefficients (and 90% confidence intervals) on the sentiment and attention indexes by regressing cumulative S&P 500 returns starting from day $t-5$ on the sentiment index, the attention index, and their interactions on day 0 for an event window between days $t=-5$ and $t=+20$. The regressions control for return volatility over the preceding 6-10 trading days, past returns over the preceding 6-10 trading days and the preceding 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Standard errors are adjusted following Hodrick (1992).

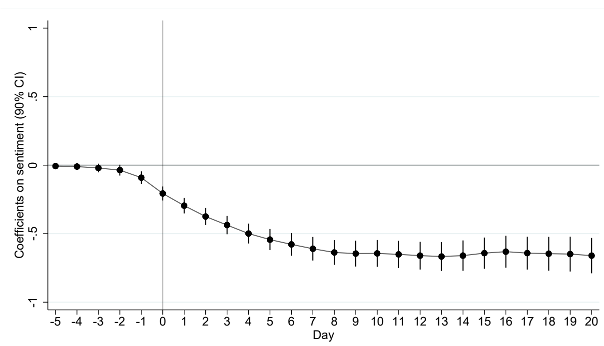
Figure 3: How cumulative abnormal turnover relates to sentiment and attention indexes



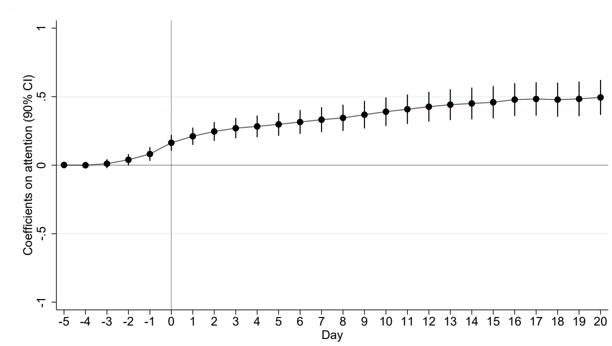
(a) Sentiment and S&P 500 turnover



(b) Attention and S&P 500 turnover



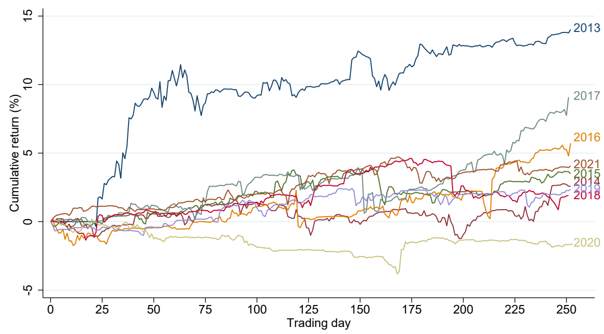
(c) Sentiment and SPY turnover



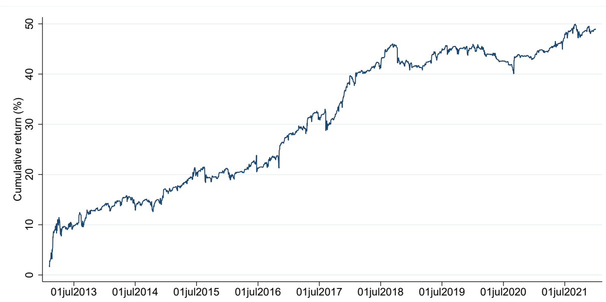
(d) Attention and SPY turnover

This figure plots the estimated coefficients (and 90% confidence intervals) on the sentiment and attention indexes by regressing cumulative abnormal turnover starting from day $t-5$ on the sentiment index, the attention index, and their interactions on day 0 for an event window between days $t=-5$ and $t=+20$. Cumulative abnormal turnover is the log turnover minus the mean log turnover in the preceding 140 through 20 day period. 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 volatility over the preceding 6-10 trading days, returns over the preceding 6-10 trading days and the preceding 11-35 trading days, and time fixed effects (day-of-week, month-of-year, and year-quarter). Standard errors are adjusted following Hodrick (1992).

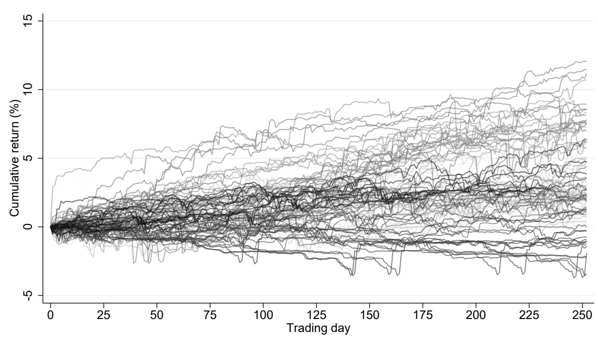
Figure 4: Dynamic trading strategy based on social media indexes
Cumulative excess return



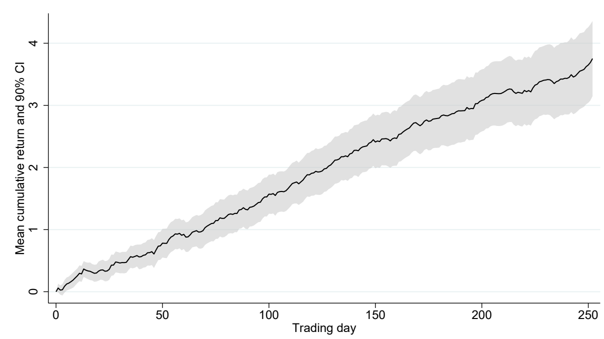
(a) Cumulated from 1st trading day each year



(b) Cumulated from 2013 through 2021



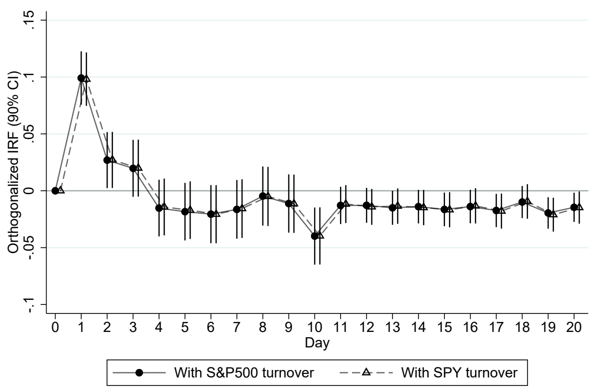
(c) Cumulated from 100 random start dates



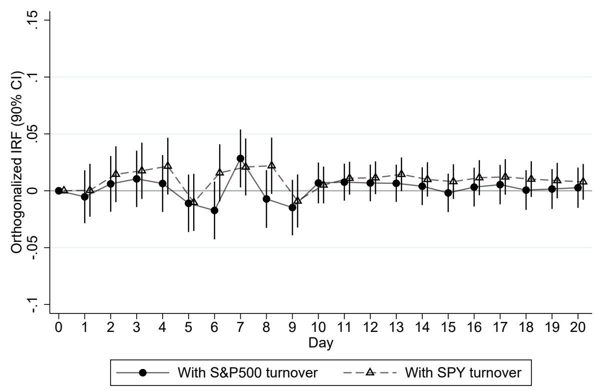
(d) Average and 90% confidence for Panel C

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 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.

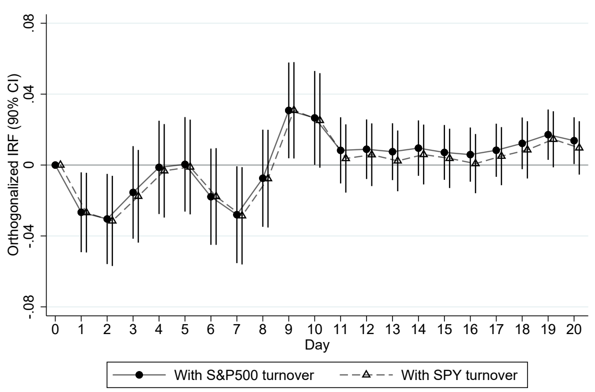
Figure 5: What predicts sentiment and attention indexes?
Impulse-response function from a VAR model



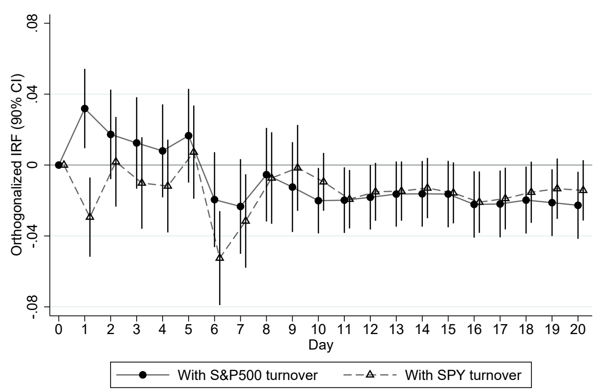
(a) Sentiment response to return



(b) Sentiment response to turnover



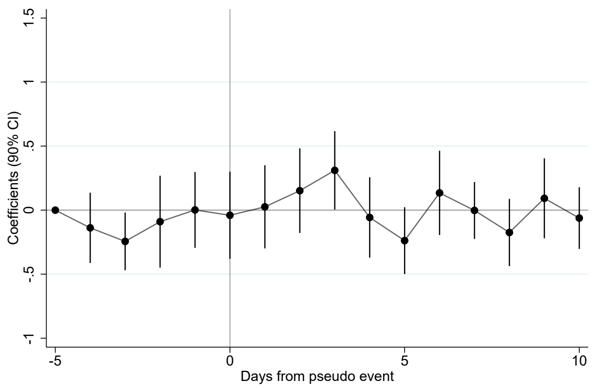
(c) Attention response to return



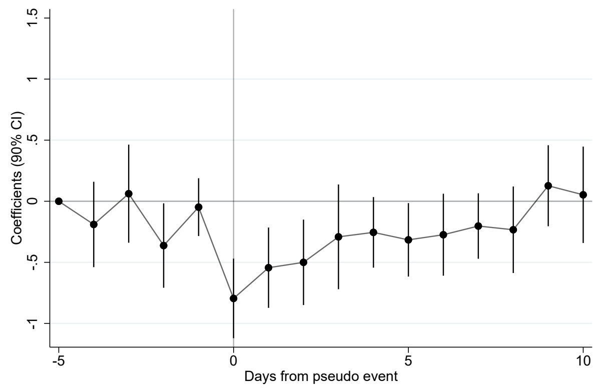
(d) Attention response to turnover

This figure plots the orthogonalized impulse response function (and 90% confidence intervals) of sentiment and attention indexes to a standard-deviation change in returns or turnover on day t. Returns refer to the 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 the SPY index ("with SPY turnover"). Appendix Figure A5 adds daily attention indexes to the VAR system.

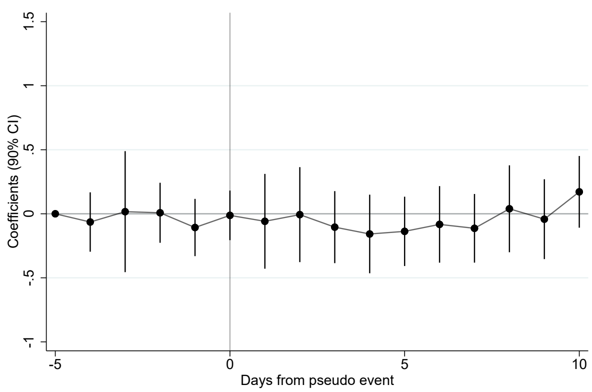
Figure 6: How sentiment and attention indexes change around jumps in returns



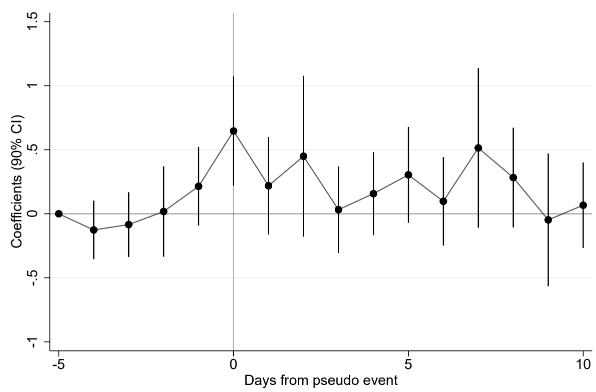
(a) Sentiment, positive jumps



(b) Sentiment, negative jumps



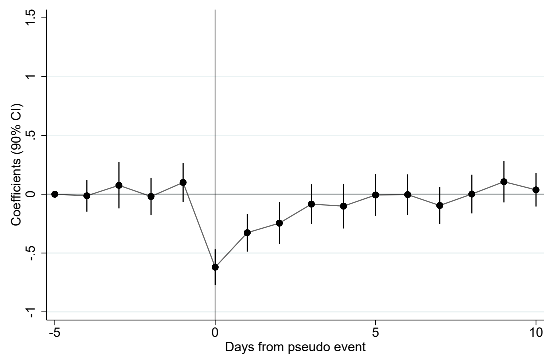
(c) Attention, positive jumps



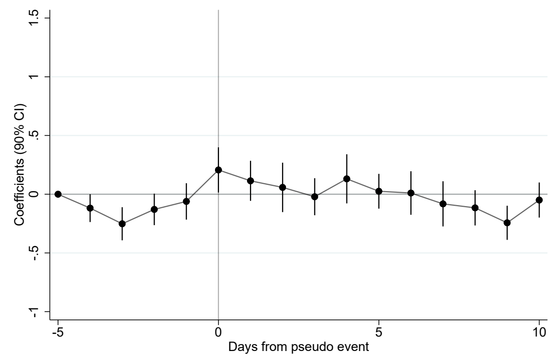
(d) Attention, negative jumps

This figure plots how the sentiment index (first row) and the 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 days $t-4$ through $t+10$ around an event. Days $t-15$ through $t-5$ are used as the reference group and a represented with a dot on day $t-5$. All regressions control for lagged volatility (days $t-5$ through $t-1$), lagged returns (days $t-5$ through $t-1$ and the previous 25 days), DOW, MOY, and YQ fixed effects. Standard errors are Newey-West with 6 lags. Appendix Figure A8 presents robustness checks by excluding jumps coinciding with FOMC announcements and redefining jump events using +/-1.5pp as thresholds.

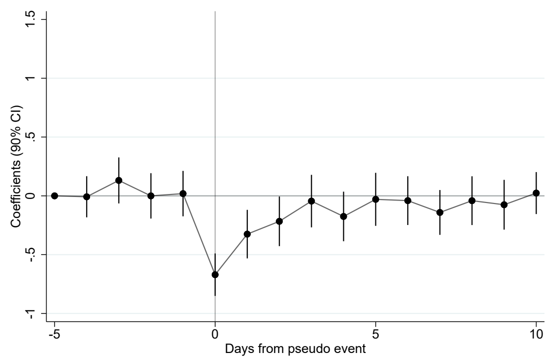
Figure 7: How sentiment and attention indexes change around jumps in the VIX



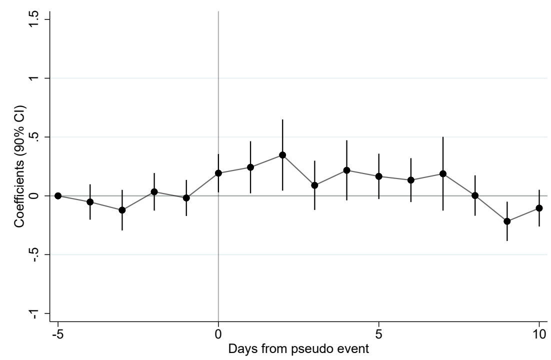
(a) Sentiment, high-Δ Vix, 15pp



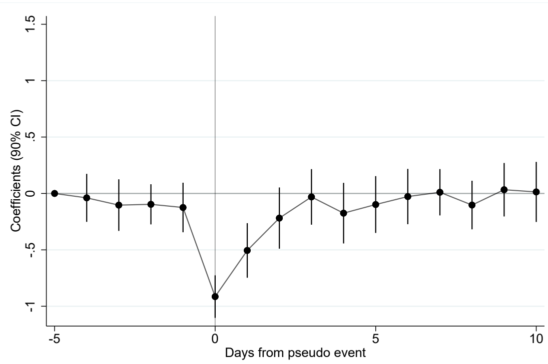
(b) Attention, high-Δ Vix, 15pp



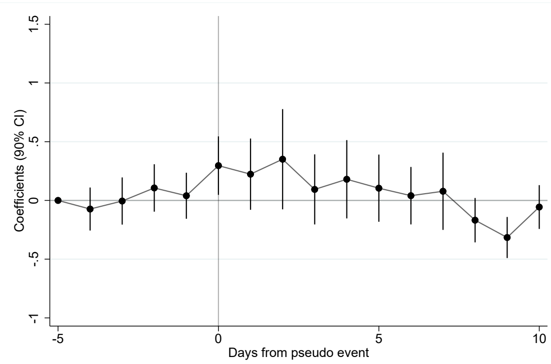
(c) Sentiment, high-Δ Vix, 20pp



(d) Attention, high-Δ Vix, 20pp



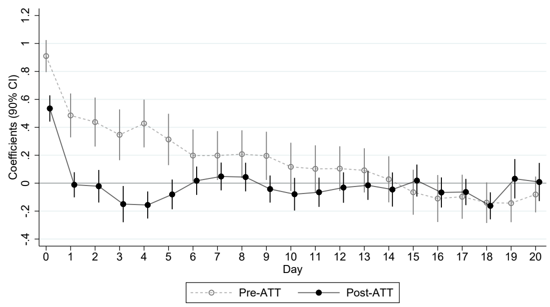
(e) Sentiment, high-Δ Vix, 25pp



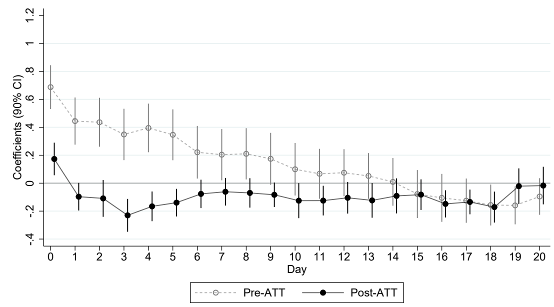
(f) Attention, high-Δ Vix, 25pp

This figure plots how the sentiment index (first row) and the attention index (second row) change from days $t-4$ through $t+10$ around days with a jump in the VIX. We categorize days with $\geq +15$, 20, or 25pp change in VIX from the prior day as high-Δ Vix events, or jumps in the VIX. 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 days $t-4$ through $t+10$ around an event. Days $t-15$ through $t-5$ are the reference group and are represented with a dot on day $t-5$. All regressions control for lagged volatility (day $t-5$ through $t-1$), lagged returns (day $t-5$ through $t-1$ and the preceding 25 days), DOW, MOY, and YQ fixed effects. Newey-West standard errors with 6 lags.

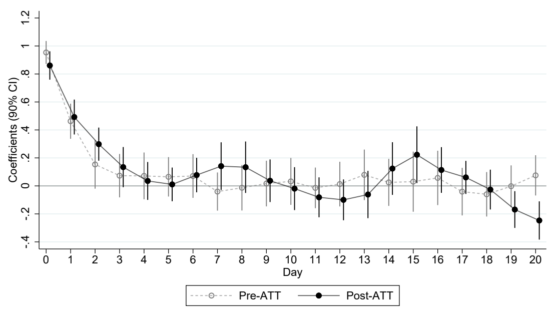
Figure 8: How central-firm social media indexes relate to overall and non-central-firm indexes before vs. after ATT policy?



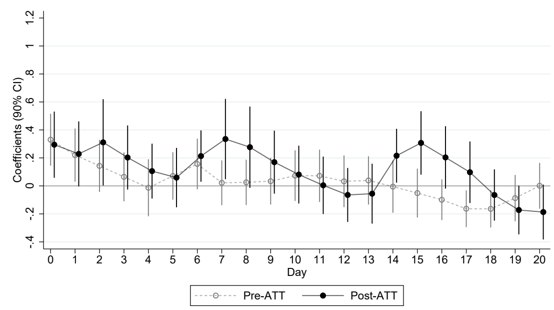
(a) How central-firm sentiment relates to overall sentiment



(b) How central-firm sentiment relates to non-central sentiment



(c) How central-firm attention relates to overall attention



(d) How central-firm attention relates to non-central attention

This figure plots how central-firm social media indexes relate to overall and non-central-firm indexes before vs. after the ATT policy change on April 26, 2021. The sample period spans 2020 May through 2021 December; pre-ATT is defined as the period from May 2020 through April 2021 and post-ATT from May 2021 through December 2021 (end of our data). Event quarter is defined in three-month intervals around ATT policy change: event quarter 0 refers to months -2, -1, and 0 relative to the policy (February, March, April 2021), event quarter 1 refers to months 1, 2, and 3 (May, June, July 2021), etc. All regressions control for DOW fixed effects and event-quarter fixed effects. Standard errors are Newey-West with 6 lags.

Table 1: Summary statistics

	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 ≥10 posts per day	354	226	58
Market cap. with ≥10 posts per day	52%	50%	28%

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 that day divided by the market cap. of all 1,500 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)

Panel A reports coefficients from 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 (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) over 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 parentheses are computed using a bootstrap method, with 1,000 random samples of firms with replacement. * $p < .1$; ** $p < .05$; *** $p < .01$.

Table 3: How social media sentiment and attention indexes relate to other sentiment and attention indexes

	(1)	(2)	(3)	(4)
Panel A: Sentiment$_t$				
ARA$_t$ (z)	-0.079***	0.021	0.021	0.068**
	(0.030)	(0.032)	(0.026)	(0.027)
AIA$_t$ (z)	0.134***	0.155***	-0.032	-0.009
	(0.032)	(0.030)	(0.025)	(0.025)
MAI (WSJ)$_t$ (z)	-0.051**	-0.098***	-0.022	-0.025
	(0.026)	(0.028)	(0.019)	(0.018)
MAI (NYT)$_t$ (z)	0.047*	0.064***	-0.026	-0.023
	(0.025)	(0.024)	(0.017)	(0.017)
Twitter EU$_t$ (z)	-0.078***	-0.045**	-0.048**	-0.047**
	(0.030)	(0.022)	(0.023)	(0.022)
RavenPack news$_t$ (z)	-0.035	-0.029	0.021	0.019
	(0.030)	(0.028)	(0.020)	(0.019)
Attention$_t$ (z)		-0.294***		-0.147***
		(0.049)		(0.030)
Observations	2,267	2,267	2,267	2,267
R^2	0.028	0.099	0.509	0.518
DOW FE	N	N	Y	Y
MOY FE	N	N	Y	Y
YQ FE	N	N	Y	Y
Panel B: Attention$_t$				
ARA$_t$ (z)	0.342***	0.322***	0.315***	0.318***
	(0.059)	(0.056)	(0.043)	(0.042)
AIA$_t$ (z)	0.073**	0.106***	0.162***	0.158***
	(0.032)	(0.031)	(0.026)	(0.026)
MAI (WSJ)$_t$ (z)	-0.161***	-0.173***	-0.017	-0.020
	(0.036)	(0.036)	(0.015)	(0.015)
MAI (NYT)$_t$ (z)	0.059**	0.070***	0.023	0.020
	(0.023)	(0.022)	(0.016)	(0.016)
Twitter EU$_t$ (z)	0.110**	0.090*	0.004	-0.002
	(0.054)	(0.050)	(0.017)	(0.016)
RavenPack news$_t$ (z)	0.021	0.012	-0.016	-0.014
	(0.030)	(0.028)	(0.019)	(0.018)
Sentiment$_t$ (z)		-0.248***		-0.123***
		(0.030)		(0.023)
Observations	2,267	2,267	2,267	2,267
R^2	0.182	0.242	0.589	0.597
DOW FE	N	N	Y	Y
MOY FE	N	N	Y	Y
YQ FE	N	N	Y	Y

This table regresses social media sentiment and attention indexes on other daily attention and sentiment indexes. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table 4: 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***	-0.106***	-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***	-0.068**	-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	2,267	2,267	2,267	2,267
R^2	0.173	0.192	0.035	0.036	0.392	0.392

This table reports how sentiment and attention indexes predict day t, day $t+1$, and day $t+2+20$ S&P 500 cumulative returns. All regressions control for return volatility 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 and attention indexes are the market-weighted average firm-level residualized sentiment and attention signal across S&P 500 firms, normalized to have zero mean and a standard deviation of one. The firm-level residualized signal is obtained by regressing raw firm-level signal on the annual signal average in the prior year and indicators for firm news (8K, Earnings announcement, or DJNW coverage) from $t-7$ through t. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table 5: 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						
$\text{Sentiment}_t(z)$	-0.020***	-0.021***	-0.018***	-0.019***	-0.192**	-0.199***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.074)	(0.076)
$\text{Attention}_t(z)$	0.071***	0.071***	0.042***	0.042***	0.057	0.058
	(0.007)	(0.007)	(0.006)	(0.005)	(0.077)	(0.078)
$\text{Sentiment} \times \text{Attention}_t(z)$		-0.007		-0.008		-0.060
		(0.005)		(0.005)		(0.064)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.596	0.597	0.482	0.483	0.749	0.749
Panel B: SPY 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***	0.024**	0.025**	0.162	0.162
	(0.009)	(0.009)	(0.010)	(0.010)	(0.165)	(0.166)
$\text{Sentiment} \times \text{Attention}_t(z)$		-0.010		-0.008		0.005
		(0.007)		(0.009)		(0.119)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.627	0.627	0.533	0.533	0.695	0.695

This table reports how sentiment and attention indexes predict day t, day $t+1$, and day $t+2 \sim t+20$ cumulative abnormal turnover. The outcome in panel A is S&P 500 cumulative abnormal turnover; in panel B it is SPY cumulative abnormal turnover. All regressions control for day t cumulative abnormal turnover, return volatility in the prior 6-10 trading days, returns over trading days $t-10$ through $t-6$ and the $t-35$ through $t-11$, and time fixed effects (day-of-week, month-of-year, and year-quarter). Sentiment and attention indexes are the market-weighted average firm-level residualized signal across S&P 500 firms, normalized to have zero mean and a standard deviation of one. Firm-level residualized signal is obtained by regressing the raw firm-level signal on the annual signal average over the preceding year, and on indicators for firm news (8K, Earnings announcement, or DJNW coverage) from $t-7$ through t. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table 6: Dynamic trading strategy based on social media indexes

	Dependent var.: Portfolio excess return$_{t+1}$(%)				
	(1)	(2)	(3)	(4)	(5)
Panel A: Weight ∈ [-1,+2]					
Alpha	0.018***	0.019***	0.019***	0.019***	0.019***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Market excess return$_t$		-0.012***	-0.013***	-0.012**	-0.139
		(0.005)	(0.005)	(0.005)	(0.095)
SMB$_t$			0.009	0.009	0.004
			(0.009)	(0.009)	(0.010)
HML$_t$			-0.006	-0.010	-0.031**
			(0.005)	(0.007)	(0.015)
MOM$_t$				-0.004	-0.001
				(0.005)	(0.006)
Observations	2,246	2,246	2,246	2,246	2,246
R^2	0.000	0.002	0.003	0.003	0.007
Alpha (annualized)	4.564***	4.754***	4.739***	4.731***	4.697***
	(1.249)	(1.278)	(1.278)	(1.279)	(1.317)
Information ratio (annualized)	1.224	1.246	1.242	1.239	1.195
FF12 industry excess return$_t$	N	N	N	N	Y
Panel B: Weight ∈ [0,1]					
Alpha	0.016***	0.017***	0.017***	0.017***	0.017***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Market excess return$_t$		-0.011***	-0.012***	-0.012***	-0.065
		(0.004)	(0.004)	(0.004)	(0.079)
SMB$_t$			0.010	0.010	0.004
			(0.008)	(0.008)	(0.008)
HML$_t$			-0.006	-0.009	-0.020
			(0.004)	(0.006)	(0.013)
MOM$_t$				-0.003	0.000
				(0.005)	(0.005)
Observations	2,246	2,246	2,246	2,246	2,246
R^2	0.000	0.003	0.004	0.004	0.007
Alpha (annualized)	4.079***	4.259***	4.244***	4.239***	4.271***
	(0.981)	(1.018)	(1.018)	(1.019)	(1.045)
Information ratio (annualized)	1.393	1.402	1.396	1.394	1.369
FF12 industry excess return$_t$	N	N	N	N	Y

This table reports the excess return and factor loadings for a dynamic trading strategy based on social media indexes constructed using information up to the preceding 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 adds momentum returns. Column 5 adds Fama-French 12 industry portfolio excess 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 preventing short-selling and leverage. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

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

This table presents tests for whether the dynamic 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. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table 8: What predicts social media sentiment and attention indexes?

	Dependent var.: Sentiment$_t(z)$		Dependent var.: 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(turnover)$_{t-1}$	-0.208**	-0.136**	0.897***	0.167***
	(0.096)	(0.057)	(0.091)	(0.055)
Ab. log(turnover)$_{t-2}$	0.024	0.039	0.016	0.058
	(0.091)	(0.055)	(0.078)	(0.050)
Ab. log(turnover)$_{t-3}$	-0.035	-0.021	0.028	-0.025
	(0.085)	(0.055)	(0.073)	(0.053)
Ab. log(turnover)$_{t-4}$	-0.035	0.015	-0.025	-0.089
	(0.105)	(0.058)	(0.077)	(0.058)
Ab. log(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

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 the firms included in the S&P 500, while columns 2 and 4 use abnormal daily turnover based on trading of SPY. All regressions include DOW, MOY, and YQ fixed effects. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$. Table A3 repeats this specification using year-month fixed effects.

Table 9: How 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

This table shows 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 separated by at least 10 days from the previous jump, leaving us with 14 negative jumps and 21 positive jumps in 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 of the jump), day +1, and days +2 through +10, and the interaction between the two. Positive jumps and days -15 through -2 are used as the reference group. All regressions control for lagged volatility (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. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$. Table **??** reports an alternative specification that includes indicators for both positive and negative jumps.

Table 10: How central-firm social media indexes relate to overall and non-central-firm indexes before vs. after ATT policy?

	Dep. var.: Overall index (z)			Dep. var.: Non-central firm index (z)		
	Day t	Day t+1	Day t+2	Day t	Day t+1	Day t+2
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Sentiment						
Post ATT × Central sentiment$_t(z)$	-0.375***	-0.496***	-0.459***	-0.514***	-0.541***	-0.545***
	(0.087)	(0.105)	(0.125)	(0.114)	(0.114)	(0.129)
Central sentiment$_t(z)$	0.910***	0.485***	0.437***	0.687***	0.445***	0.436***
	(0.070)	(0.095)	(0.106)	(0.095)	(0.102)	(0.106)
Observations	422	421	420	422	421	420
R^2	0.709	0.308	0.280	0.472	0.301	0.294
Panel B: Attention						
Post ATT × Central attention$_t(z)$	-0.093	0.030	0.145	-0.036	0.009	0.168
	(0.079)	(0.107)	(0.126)	(0.181)	(0.178)	(0.217)
Central attention$_t(z)$	0.954***	0.462***	0.154	0.330***	0.220*	0.143
	(0.050)	(0.076)	(0.105)	(0.112)	(0.115)	(0.113)
Observations	422	421	420	422	421	420
R^2	0.898	0.670	0.601	0.647	0.628	0.619
DOW FE	Y	Y	Y	Y	Y	Y
Event quarter FE	Y	Y	Y	Y	Y	Y

This table studies how central firm indexes relate to overall social media indexes and non-central firm indexes before vs. after the ATT policy change on April 26, 2021. The dependent variable in columns 1-3 is overall sentiment (panel A) and overall attention (panel B) while the dependent variable in columns 4-6 is central-firm sentiment (panel A) and central-firm attention (panel B). The sample period spans 2020 May through 2021 December; *post ATT* is an indicator for the period from May 2021 through December 2021 (the end of our data); the period from May 2020 through April 2021 is used as the reference period. Event quarter is defined in three-month intervals around the ATT policy change: event quarter 0 denotes months -2, -1, and 0 from the policy (February, March, April 2021), event quarter 1 denotes months 1, 2, 3 from the policy (May, June, July 2021), etc. All regressions control for DOW fixed effects and event-quarter fixed effects. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

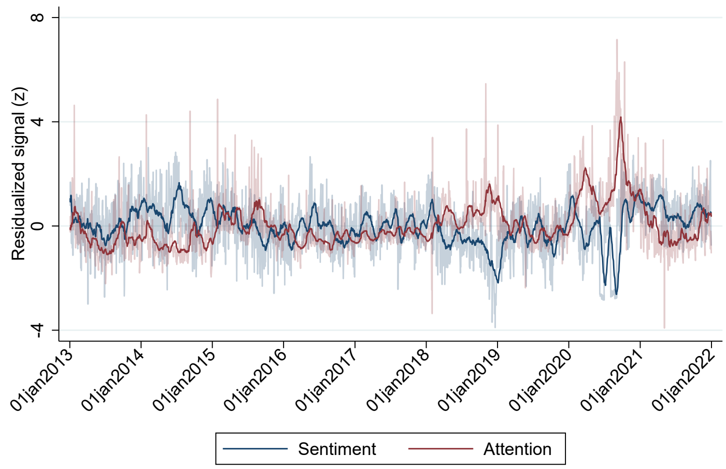
Internet Appendix:

Market Signals from Social Media

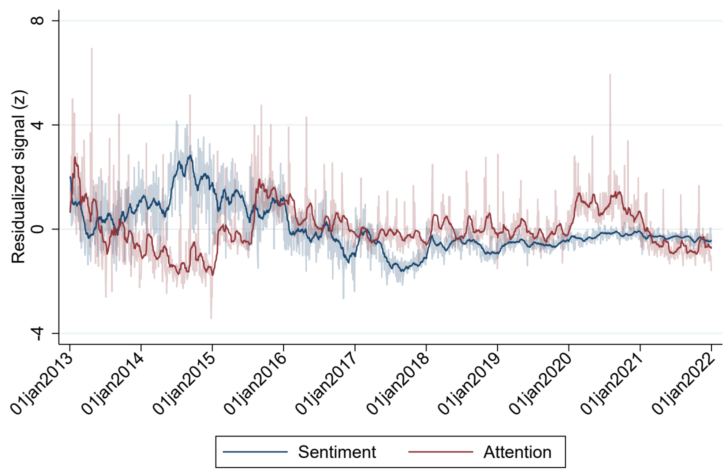
by J. Anthony Cookson, Runjing Lu, William Mullins and Marina Niessner[1]

[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)

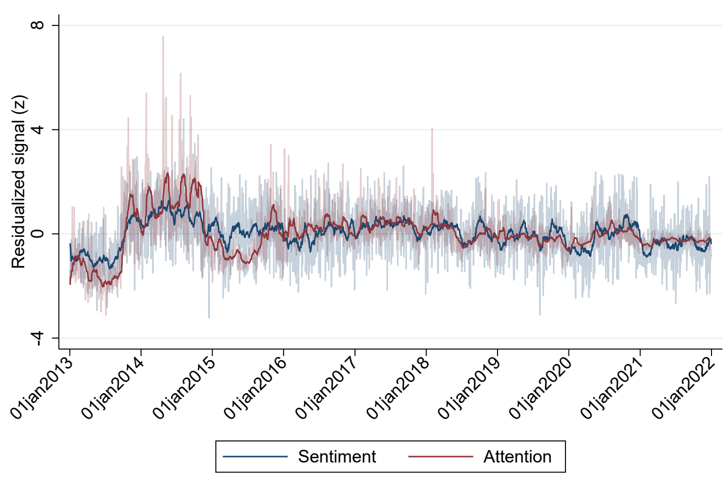
Figure A1: Time series of platform-level sentiment and attention



(a) StockTwits



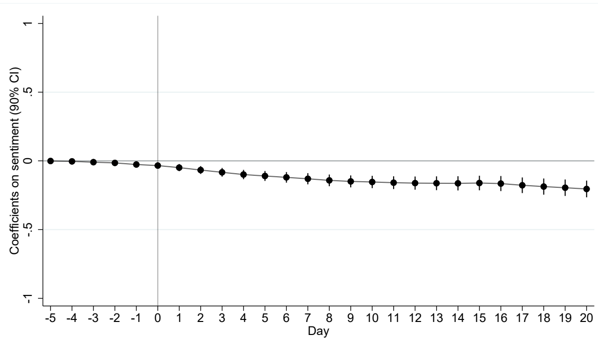
(b) Twitter



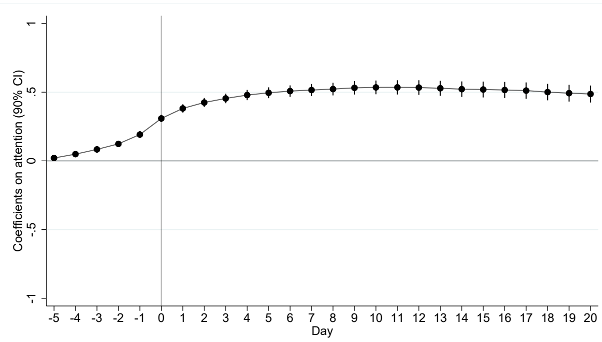
(c) Seeking Alpha

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.

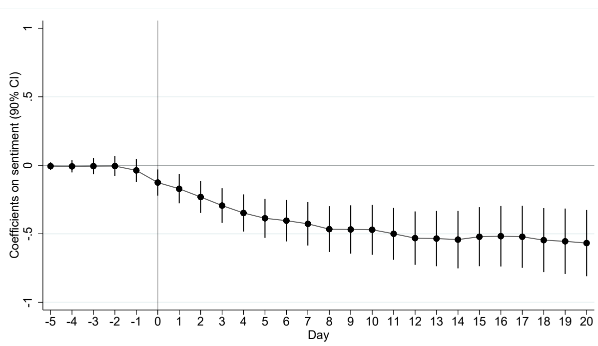
Figure A2: How does cumulative abnormal *retail* turnover relate to sentiment and attention indexes?



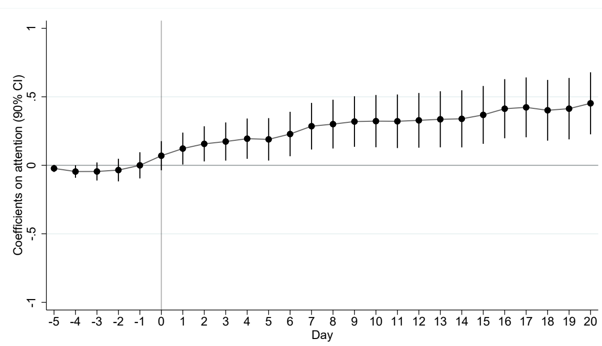
(a) Sentiment and S&P 500 turnover



(b) Attention and S&P 500 turnover



(c) Sentiment and SPY turnover

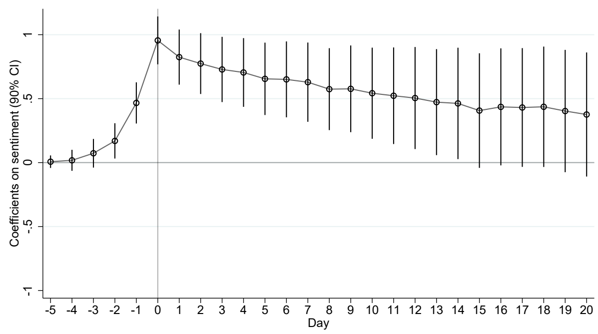


(d) Attention and SPY turnover

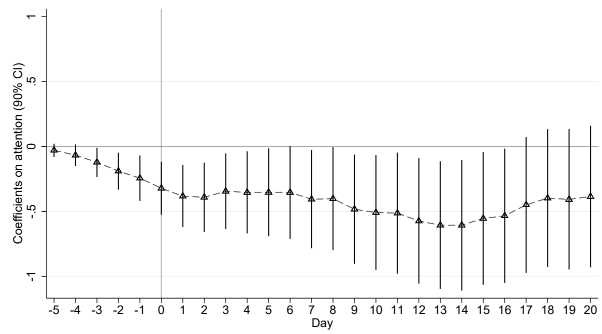
This figure plots the estimated coefficients (and 90% confidence intervals) on sentiment and attention indexes by regressing cumulative abnormal *retail* turnover starting from day $t-5$ on sentiment index, attention index, and their interactions on day 0 for an event window between days $t = -5$ and $t = +20$. 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 Figure 3. Standard errors are adjusted following Hodrick (1992).

Figure A3: How do cumulative returns and cumulative abnormal turnover relate to sentiment and attention indexes?

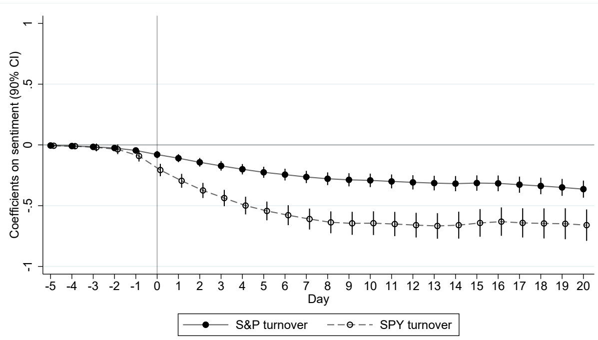
With additional controls



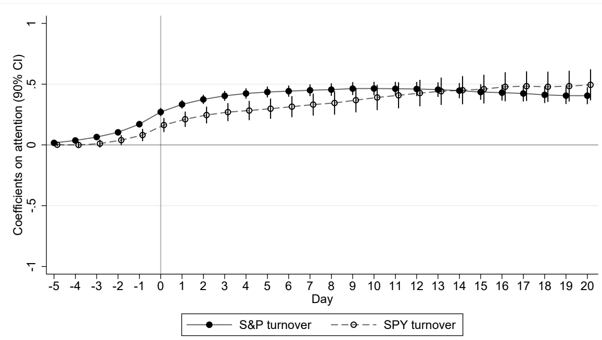
(a) Sentiment and S&P 500 return



(b) Attention and S&P 500 return



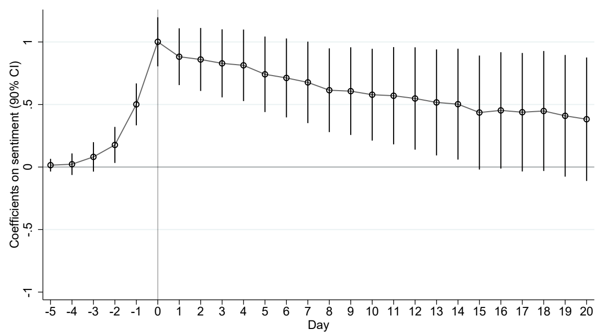
(c) Sentiment and turnover



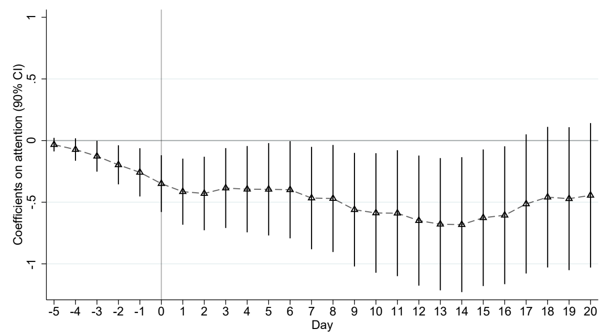
(d) Attention and turnover

This figure repeats Figure 2 and Figure 3 by additionally controlling for other attention and sentiment signals (ARA, AIA, MAI - WSJ, MAI - NYT, Twitter Economic Uncertainty, and RavenPack aggregate news sentiment). Everything else mirrors the original figures. Standard errors are adjusted following Hodrick (1992).

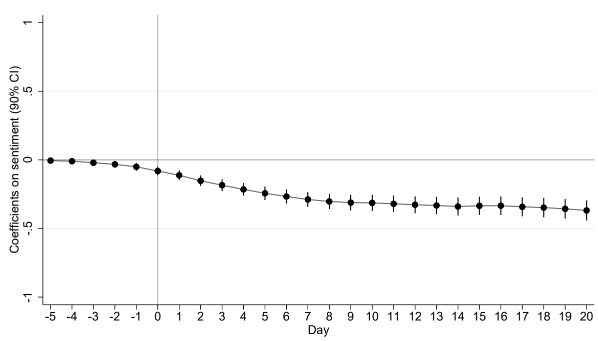
Figure A4: How do cumulative returns and cumulative abnormal turnover relate to sentiment and attention indexes?
Alternative sample



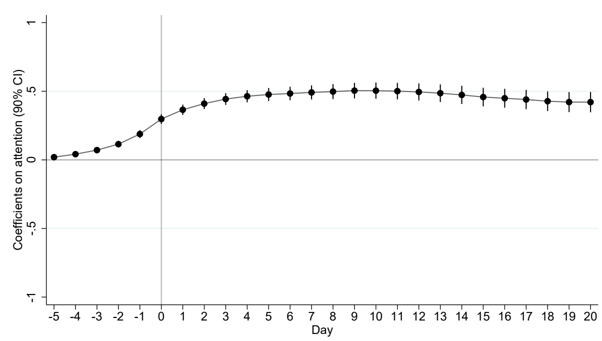
(a) Sentiment



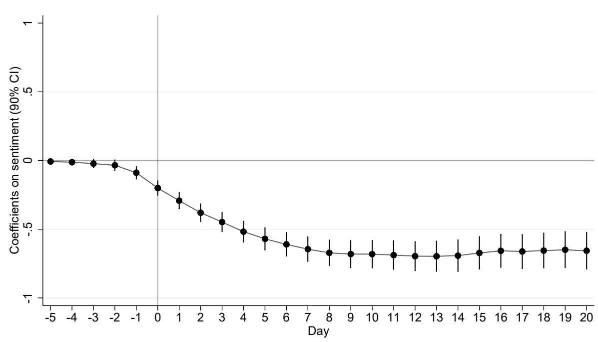
(b) Attention



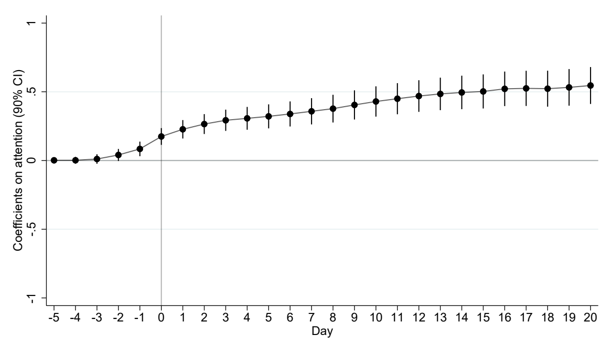
(c) Sentiment and S&P 500 turnover



(d) Attention and S&P 500 turnover



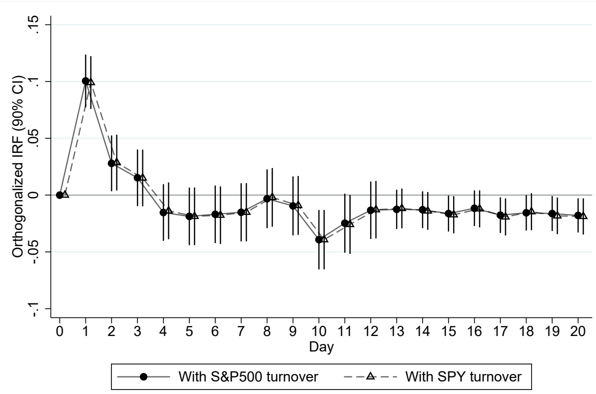
(e) Sentiment and SPY turnover



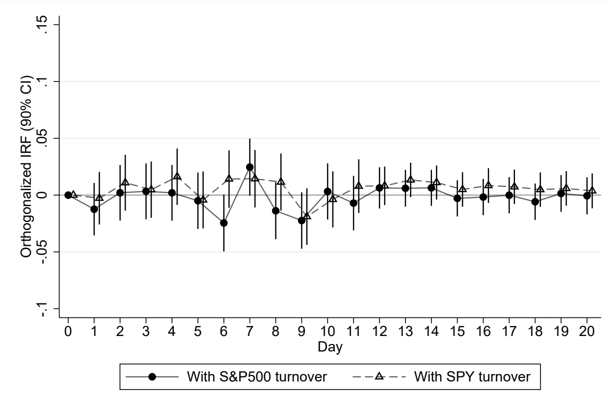
(f) Attention and SPY turnover

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 the corresponding tables. Standard errors are adjusted following Hodrick (1992).

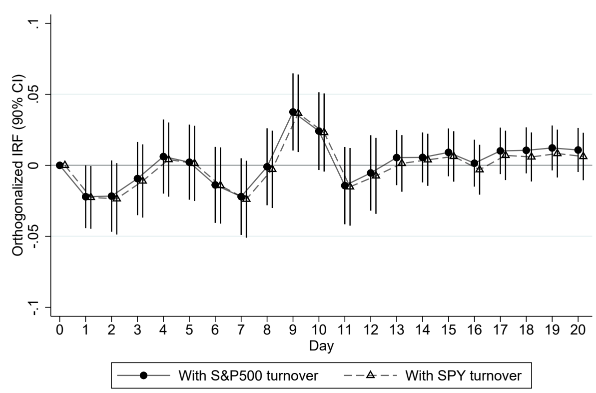
Figure A5: What predicts sentiment and attention indexes?
Impulse-response function from a VAR model
With additional attention controls



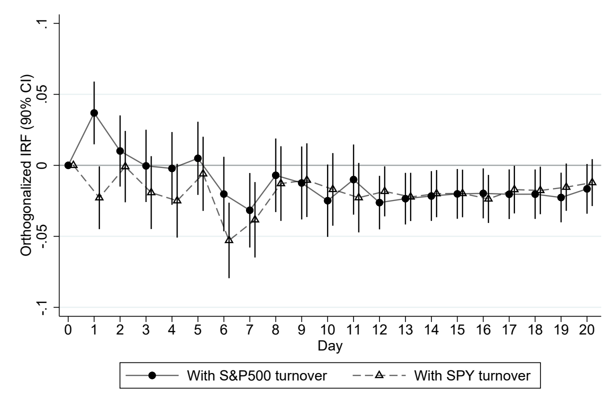
(a) Sentiment response to return



(b) Sentiment response to turnover



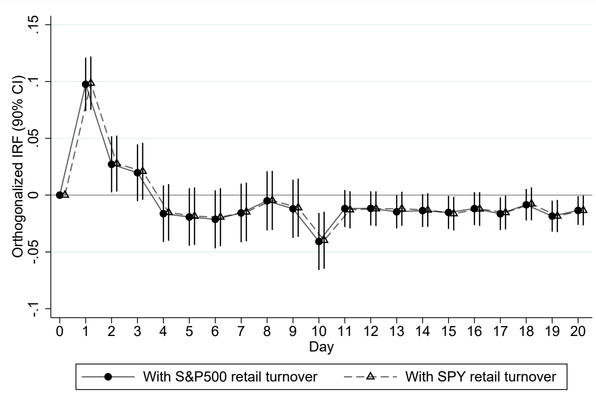
(c) Attention response to return



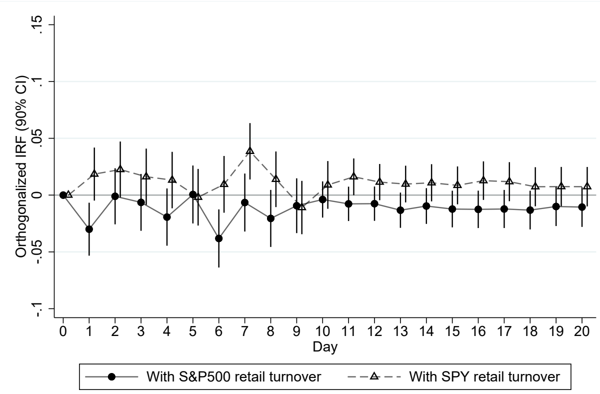
(d) Attention response to turnover

This figure repeats the VAR models in Figure 5 while further including ARA and AIA. Everything else mirrors Figure 5.

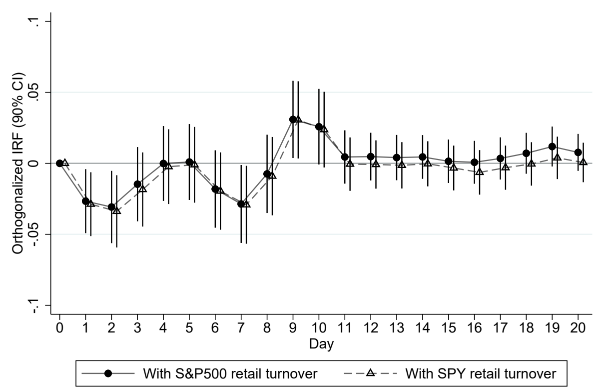
Figure A6: What predicts sentiment and attention indexes?
Impulse-response function from a VAR model
Turnover based on retail trading



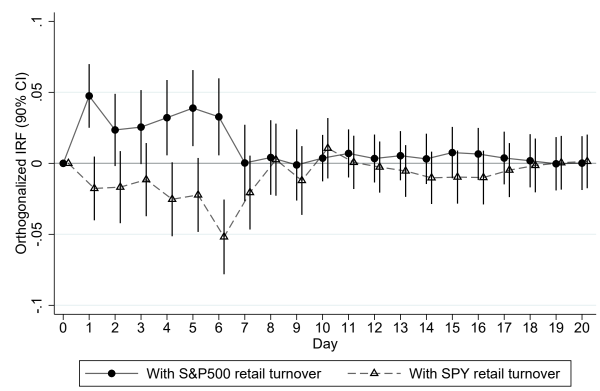
(a) Sentiment response to return



(b) Sentiment response to turnover



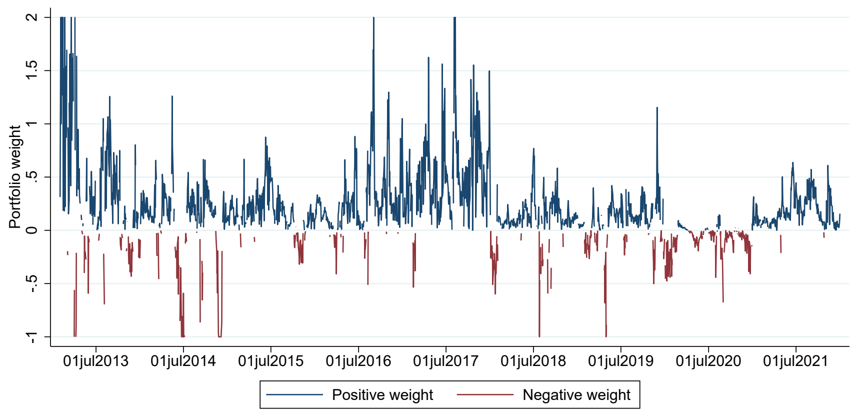
(c) Attention response to return



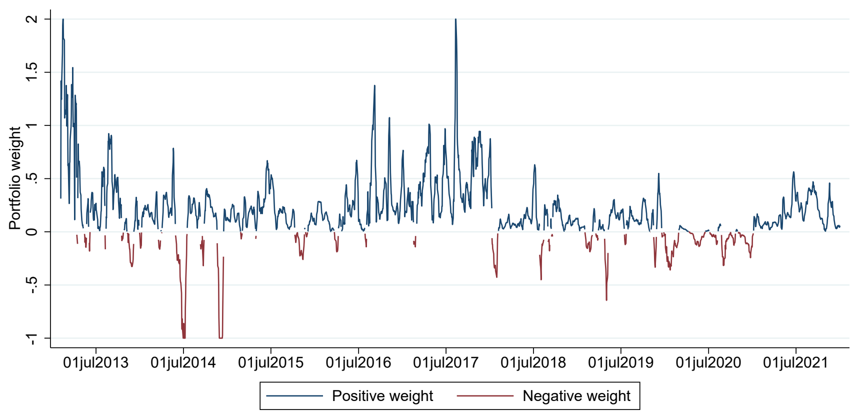
(d) Attention response to turnover

This figure repeats the VAR models in Figure 5 while replacing total turnover with respective retail turnover. Everything else mirrors Figure 5.

Figure A7: Dynamic trading strategy based on social media indexes
Time series of portfolio weights



(a) Baseline portfolio weight

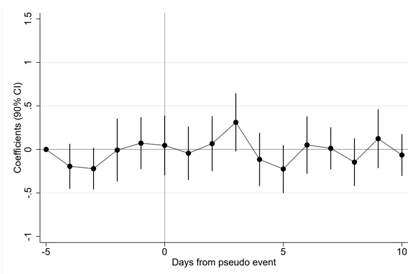


(b) Rolling 5-day avg. weight

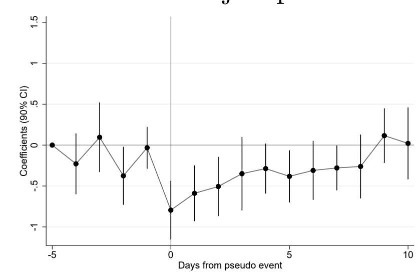
This figure plots the portfolio weights for a dynamic trading strategy based on social media indexes constructed using information up to last month (see Section 3.3). Panel A plots the daily weight while Panel B plots the 5-day rolling average. On average, 76% (78%) of days put positive weights on the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

Figure A8: How do sentiment and attention indexes change around return jumps
Alternative definitions of return jumps

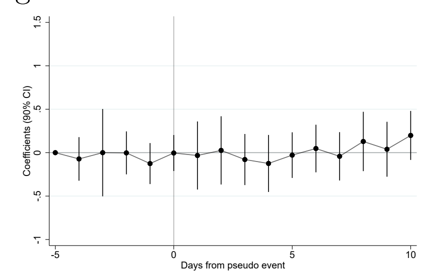
Exclude return jumps coinciding with FOMC announcements



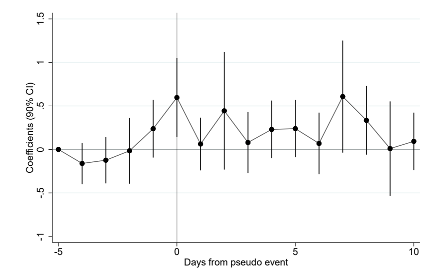
(a) Sentiment, positive jumps



(b) Sentiment, negative jumps

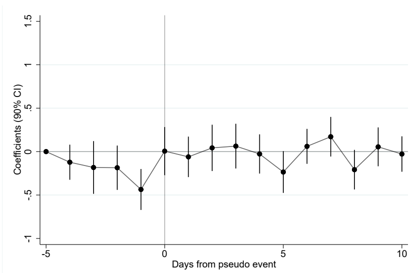


(c) Attention, positive jumps

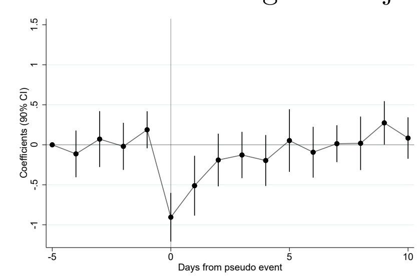


(d) Attention, negative jumps

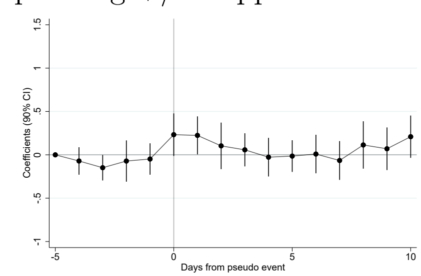
Defining return jumps using +/- 1.5pp



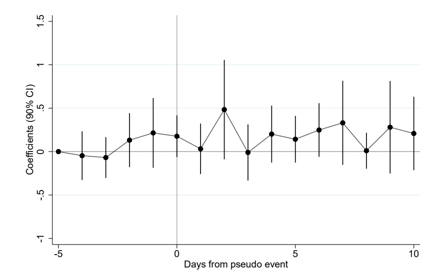
(e) Sentiment, positive jumps



(f) Sentiment, negative jumps



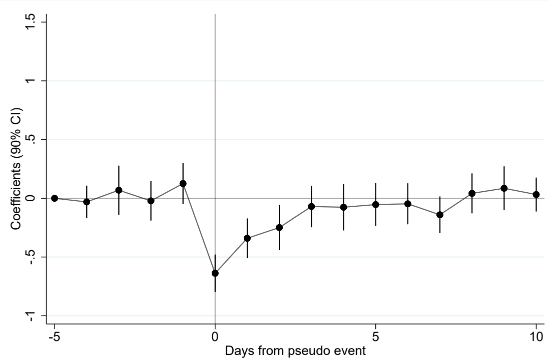
(g) Attention, positive jumps



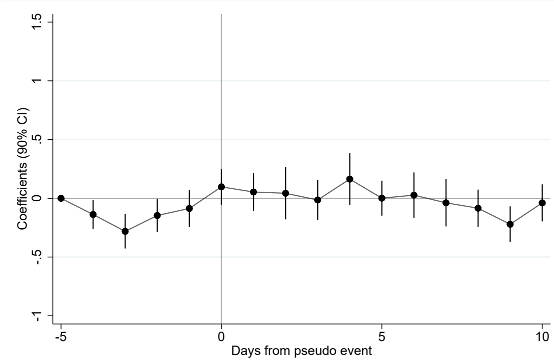
(h) Attention, negative jumps

This figure repeats Figure 6 using alternative definitions of return jumps. The first row excludes return jumps on the same day of a FOMC meeting, and the 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 Figure 6.

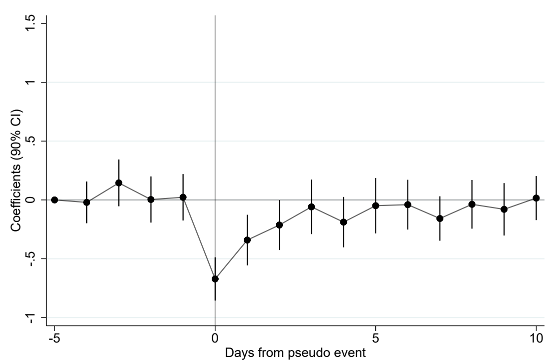
Figure A9: How do sentiment and attention indexes change around jumps in the VIX?
Exclude VIX jumps coinciding with FOMC announcements



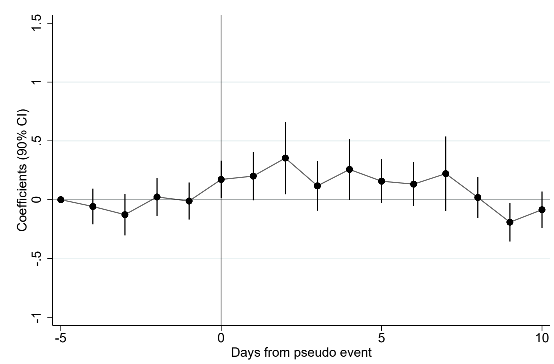
(a) Sentiment, high-Δ Vix, 15pp



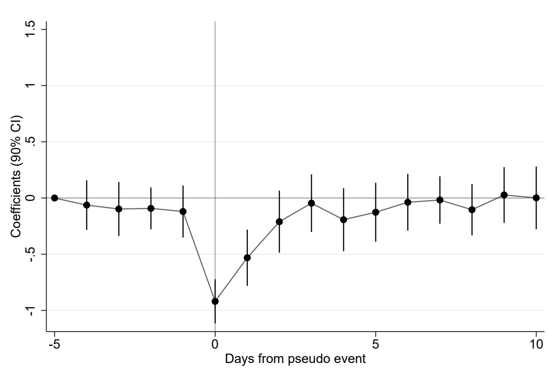
(b) Attention, high-Δ Vix, 15pp



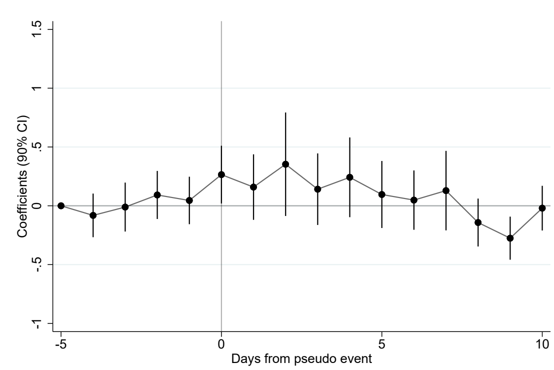
(c) Sentiment, high-Δ Vix, 20pp



(d) Attention, high-Δ Vix, 20pp



(e) Sentiment, high-Δ Vix, 25pp



(f) Attention, high-Δ Vix, 25pp

This figure repeats Figure 7 while excluding VIX jumps that occur on the same day as an FOMC meeting. Everything else mirrors Figure 7.

Table A1: How social media sentiment and attention indexes relate to other sentiment and attention indexes *Robustness*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Sentiment$_t$ (z)												
$\text{ARA}_t(z)$	-0.068**					-0.079***	0.003					0.021
	(0.030)					(0.030)	(0.027)					(0.026)
$\text{AIA}_t(z)$		0.100***				0.134***		-0.039				-0.032
		(0.030)				(0.032)		(0.025)				(0.025)
MAI $(\text{WSJ})_t(z)$			-0.045*			-0.051**			-0.028			-0.022
			(0.026)			(0.026)			(0.019)			(0.019)
MAI $(\text{NYT})_t(z)$			0.030			0.047*			-0.028*			-0.026
			(0.025)			(0.025)			(0.017)			(0.017)
Twitter $\text{EU}_t(z)$				-0.076**		-0.078***				-0.054**		-0.048**
				(0.032)		(0.030)				(0.024)		(0.023)
RavenPack $\text{news}_t(z)$					-0.011	-0.035					0.023	0.021
					(0.028)	(0.030)					(0.020)	(0.020)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.005	0.010	0.002	0.006	0.000	0.028	0.504	0.505	0.506	0.507	0.505	0.509
Panel B: Attention$_t$ (z)												
$\text{ARA}_t(z)$	0.378***					0.342***	0.356***					0.315***
	(0.057)					(0.059)	(0.043)					(0.043)
$\text{AIA}_t(z)$		0.133***				0.073**		0.248***				0.162***
		(0.033)				(0.032)		(0.026)				(0.026)
MAI $(\text{WSJ})_t(z)$			-0.148***			-0.161***			0.002			-0.017
			(0.040)			(0.036)			(0.019)			(0.015)
MAI $(\text{NYT})_t(z)$			0.103***			0.059**			0.058***			0.023
			(0.027)			(0.023)			(0.018)			(0.016)
Twitter $\text{EU}_t(z)$				0.168**		0.110**				0.060**		0.004
				(0.070)		(0.054)				(0.025)		(0.017)
RavenPack $\text{news}_t(z)$					0.044	0.021					-0.012	-0.016
					(0.032)	(0.030)					(0.021)	(0.019)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.143	0.018	0.026	0.028	0.002	0.182	0.575	0.530	0.499	0.499	0.496	0.589
DOW, MOY, YQ FE	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y

This table regresses social media sentiment and attention indexes on other daily attention and sentiment indexes. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A2: How social media sentiment and attention indexes relate to other sentiment and attention indexes *Adding macroeconomic indexes*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Sentiment$_t$ (z)												
$\text{ARA}_t(z)$	-0.068**					-0.079***	0.003					0.021
	(0.030)					(0.030)	(0.027)					(0.026)
$\text{AIA}_t(z)$		0.100***				0.134***		-0.039				-0.032
		(0.030)				(0.032)		(0.025)				(0.025)
MAI (WSJ)$_t(z)$			-0.045*			-0.051**			-0.028			-0.022
			(0.026)			(0.026)			(0.019)			(0.019)
MAI (NYT)$_t(z)$			0.030			0.047*			-0.028*			-0.026
			(0.025)			(0.025)			(0.017)			(0.017)
Twitter EU$_t(z)$				-0.076**		-0.078***				-0.054**		-0.048**
				(0.032)		(0.030)				(0.024)		(0.023)
RavenPack news$_t(z)$					-0.011	-0.035					0.023	0.021
					(0.028)	(0.030)					(0.020)	(0.020)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.005	0.010	0.002	0.006	0.000	0.028	0.504	0.505	0.506	0.507	0.505	0.509
Panel B: Attention$_t$ (z)												
$\text{ARA}_t(z)$	0.378***					0.342***	0.356***					0.315***
	(0.057)					(0.059)	(0.043)					(0.043)
$\text{AIA}_t(z)$		0.133***				0.073**		0.248***				0.162***
		(0.033)				(0.032)		(0.026)				(0.026)
MAI (WSJ)$_t(z)$			-0.148***			-0.161***			0.002			-0.017
			(0.040)			(0.036)			(0.019)			(0.015)
MAI (NYT)$_t(z)$			0.103***			0.059**			0.058***			0.023
			(0.027)			(0.023)			(0.018)			(0.016)
Twitter EU$_t(z)$				0.168**		0.110**				0.060**		0.004
				(0.070)		(0.054)				(0.025)		(0.017)
RavenPack news$_t(z)$					0.044	0.021					-0.012	-0.016
					(0.032)	(0.030)					(0.021)	(0.019)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.143	0.018	0.026	0.028	0.002	0.182	0.575	0.530	0.499	0.499	0.496	0.589
DOW, MOY, YQ FE	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y

This table regresses social media sentiment and attention indexes on other daily attention and sentiment indexes. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A3: 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

This table repeats Table 8 while replacing MOY and YQ fixed effects with YM (year-month) fixed effects. Everything else mirrors Table 8.Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A4: Do sentiment and attention indexes predict returns and turnover?
Cumulative outcomes over various horizons

	(1) Day 1-5	(2) Day 1-5	(3) Day 1-10	(4) Day 1-10	(5) Day 1-15	(6) Day 1-15	(7) Day 1-20	(8) Day 1-20
Panel A: S&P return								
Sentiment$_t(z)$	-0.213**	-0.213**	-0.274**	-0.279**	-0.383**	-0.382**	-0.377*	-0.372*
	(0.084)	(0.089)	(0.131)	(0.136)	(0.174)	(0.183)	(0.214)	(0.225)
Attention$_t(z)$	-0.075	-0.075	-0.259	-0.258	-0.335	-0.335	-0.213	-0.213
	(0.104)	(0.104)	(0.160)	(0.161)	(0.206)	(0.207)	(0.242)	(0.243)
Sentiment × Attention$_t(z)$		0.003		-0.044		0.008		0.039
		(0.094)		(0.145)		(0.191)		(0.217)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.157	0.157	0.240	0.241	0.326	0.326	0.415	0.415
Panel B: S&P turnover								
Sentiment$_t(z)$	-0.106***	-0.110***	-0.153***	-0.159***	-0.173***	-0.178***	-0.209***	-0.217***
	(0.017)	(0.017)	(0.024)	(0.025)	(0.033)	(0.033)	(0.040)	(0.040)
Attention$_t(z)$	0.113***	0.113***	0.136***	0.136***	0.120***	0.121***	0.099***	0.100***
	(0.015)	(0.015)	(0.024)	(0.024)	(0.033)	(0.033)	(0.038)	(0.038)
Sentiment × Attention$_t(z)$		-0.029**		-0.046**		-0.043		-0.067**
		(0.014)		(0.021)		(0.028)		(0.034)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.593	0.594	0.673	0.673	0.724	0.724	0.762	0.762
Panel C: SPY turnover								
Sentiment$_t(z)$	-0.225***	-0.229***	-0.280***	-0.281***	-0.266***	-0.264***	-0.264***	-0.264***
	(0.029)	(0.029)	(0.043)	(0.044)	(0.059)	(0.060)	(0.072)	(0.074)
Attention$_t(z)$	0.050*	0.051*	0.106**	0.106**	0.156**	0.156**	0.186***	0.186***
	(0.030)	(0.030)	(0.049)	(0.049)	(0.063)	(0.063)	(0.072)	(0.072)
Sentiment × Attention$_t(z)$		-0.034		-0.013		0.010		-0.003
		(0.024)		(0.041)		(0.053)		(0.062)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.607	0.607	0.653	0.653	0.679	0.679	0.708	0.708
DOW FE	Y	Y	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y	Y	Y

This table repeats Table 4 and Table 5 while using cumulative return and turnover with various horizons as outcomes. Standard errors are adjusted following Hodrick (1992) in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A5: 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+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P retail turnover						
Sentiment$_t(z)$	0.000	-0.000	-0.007	-0.008	-0.072**	-0.082***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.029)	(0.029)
Attention$_t(z)$	0.085***	0.085***	0.052***	0.052***	0.091***	0.092***
	(0.007)	(0.006)	(0.005)	(0.005)	(0.029)	(0.029)
Sentiment × Attention$_t(z)$		-0.007*		-0.010**		-0.079***
		(0.004)		(0.004)		(0.026)
Observations	2,267	2,267	2,266	2,266	2,252	2,252
R^2	0.742	0.743	0.611	0.612	0.799	0.800
Panel B: SPY retail turnover						
Sentiment$_t(z)$	-0.063***	-0.065***	-0.020	-0.021	-0.213**	-0.212**
	(0.017)	(0.017)	(0.018)	(0.018)	(0.106)	(0.106)
Attention$_t(z)$	0.051***	0.051***	0.036**	0.036**	0.163	0.162
	(0.016)	(0.016)	(0.018)	(0.018)	(0.108)	(0.108)
Sentiment × Attention$_t(z)$		-0.016		-0.008		0.014
		(0.013)		(0.013)		(0.089)
Observations	2,267	2,267	2,266	2,266	2,252	2,252
R^2	0.352	0.352	0.323	0.323	0.724	0.724
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

This table reports how sentiment and attention indexes predict day t, day t+1, and day t+1∼t+15 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 5. Newey-West standard errors with 6 lags in parentheses in columns 1-4 and following Hodrick (1992) in columns 5-6; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A6: 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+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P return						
Sentiment$_t(z)$	0.583***	0.609***	-0.142***	-0.145***	-0.609***	-0.619***
	(0.043)	(0.045)	(0.037)	(0.040)	(0.106)	(0.114)
Attention$_t(z)$	-0.077***	-0.076***	-0.033	-0.033	-0.030	-0.030
	(0.026)	(0.029)	(0.032)	(0.032)	(0.149)	(0.149)
Sentiment × Attention$_t(z)$		0.190***		-0.021		-0.075
		(0.040)		(0.036)		(0.145)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.222	0.245	0.085	0.086	0.588	0.588
Panel B: S&P turnover						
Sentiment$_t(z)$	-0.021***	-0.022***	-0.016***	-0.018***	-0.100***	-0.109***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.020)	(0.020)
Attention$_t(z)$	0.075***	0.075***	0.040***	0.040***	-0.037*	-0.038*
	(0.008)	(0.007)	(0.006)	(0.005)	(0.022)	(0.022)
Sentiment × Attention$_t(z)$		-0.010*		-0.012**		-0.068***
		(0.005)		(0.005)		(0.019)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.616	0.617	0.517	0.519	0.861	0.861
Panel C: SPY turnover						
Sentiment$_t(z)$	-0.087***	-0.090***	-0.060***	-0.062***	-0.211***	-0.225***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.037)	(0.038)
Attention$_t(z)$	0.059***	0.059***	0.020**	0.020**	-0.137***	-0.137***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.049)	(0.049)
Sentiment × Attention$_t(z)$		-0.021***		-0.022**		-0.099***
		(0.007)		(0.009)		(0.036)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.653	0.654	0.576	0.578	0.847	0.848
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
YM FE	Y	Y	Y	Y	Y	Y

This table reports how sentiment and attention indexes predict day t, day t+1, and day t+1∼t+15 S&P 500 cumulative return and cumulative abnormal turnover. Panel A follows Table 4 and panels B and C follow Table 5 panels A and B, except that we replace MOY and YQ fixed effects with YM fixed effects. Newey-West standard errors with 6 lags in parentheses in columns 1-4 and following Hodrick (1992) in columns 5-6; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A7: Do sentiment and attention indexes predict returns and turnover? *Alternative index: market capitalization-weighted raw signals*

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P return						
Sentiment$_t(z)$	0.570***	0.589***	-0.114***	-0.118***	-0.361*	-0.334*
	(0.045)	(0.049)	(0.037)	(0.039)	(0.190)	(0.200)
Attention$_t(z)$	-0.102***	-0.121***	-0.075**	-0.071*	-0.264	-0.292
	(0.031)	(0.033)	(0.035)	(0.038)	(0.206)	(0.222)
Sentiment × Attention$_t(z)$		0.111***		-0.023		0.157
		(0.036)		(0.032)		(0.200)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.171	0.176	0.035	0.036	0.323	0.324
Panel B: S&P turnover						
Sentiment$_t(z)$	-0.020***	-0.022***	-0.016***	-0.018***	-0.212***	-0.210***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.034)	(0.034)
Attention$_t(z)$	0.054***	0.056***	0.037***	0.039***	0.132***	0.130***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.036)	(0.037)
Sentiment × Attention$_t(z)$		-0.014***		-0.007		0.009
		(0.005)		(0.005)		(0.033)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.572	0.573	0.476	0.476	0.724	0.724
Panel C: SPY turnover						
Sentiment$_t(z)$	-0.085***	-0.088***	-0.061***	-0.062***	-0.311***	-0.304***
	(0.010)	(0.010)	(0.011)	(0.012)	(0.062)	(0.063)
Attention$_t(z)$	0.031***	0.034***	0.019	0.020	0.218***	0.211***
	(0.009)	(0.009)	(0.011)	(0.012)	(0.070)	(0.072)
Sentiment × Attention$_t(z)$		-0.016*		-0.006		0.043
		(0.008)		(0.010)		(0.062)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.620	0.620	0.531	0.531	0.680	0.680
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

This table reports how sentiment and attention indexes constructed using an alternative procedure predict day t, day t+1, and day t+1∼t+15 S&P 500 cumulative returns and cumulative abnormal turnover. Sentiment and attention indexes are constructed by first market-capitalization weighting *raw* firm-day level signals to platform-day level and then obtaining the first principal component from a principal component analysis on platform-day level signal across StockTwits, Twitter, and SeekingAlpha. Each panel represents a different outcome: panel A S&P 500 returns, panel B S&P 500 cumulative abnormal turnover, and panel C SPY cumulative abnormal turnover. Everything else mirrors Table 4 and Table 5. Newey-West standard errors with 6 lags in parentheses in columns 1-4 and following Hodrick (1992) in columns 5-6; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A8: Do sentiment and attention indexes predict returns and turnover? *alternative index: Equal-weighted residualized signals*

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P return						
$\text{Sentiment}_t(z)$	0.514***	0.521***	-0.052	-0.052	-0.417*	-0.421*
	(0.044)	(0.045)	(0.036)	(0.037)	(0.215)	(0.220)
$\text{Attention}_t(z)$	-0.022	-0.042	-0.073*	-0.074*	-0.609	-0.597
	(0.046)	(0.049)	(0.039)	(0.040)	(0.406)	(0.423)
Sentiment $\times$ $\text{Attention}_t(z)$		0.090***		0.006		-0.052
		(0.032)		(0.024)		(0.181)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.154	0.159	0.031	0.031	0.329	0.329
Panel B: S&P turnover						
$\text{Sentiment}_t(z)$	-0.029***	-0.029***	-0.027***	-0.027***	-0.182***	-0.177***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.039)	(0.039)
$\text{Attention}_t(z)$	0.047***	0.048***	0.019**	0.020***	0.230***	0.217***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.072)	(0.073)
Sentiment $\times$ $\text{Attention}_t(z)$		-0.004		-0.002		0.058**
		(0.004)		(0.005)		(0.029)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.564	0.564	0.471	0.471	0.724	0.725
Panel C: SPY turnover						
$\text{Sentiment}_t(z)$	-0.092***	-0.093***	-0.076***	-0.075***	-0.363***	-0.347***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.073)	(0.072)
$\text{Attention}_t(z)$	0.017	0.018	-0.004	-0.005	0.032	-0.017
	(0.012)	(0.013)	(0.015)	(0.015)	(0.128)	(0.129)
Sentiment $\times$ $\text{Attention}_t(z)$		-0.005		0.006		0.212***
		(0.007)		(0.009)		(0.053)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.622	0.622	0.536	0.536	0.680	0.681
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

This table reports how sentiment and attention indexes constructed using an alternative procedure predict day t, day t+1, and day t+1∼t+15 S&P 500 cumulative return and cumulative abnormal turnover. Sentiment and attention indexes are constructed by first equal weighting residualized firm-day level signals to platform-day level and then obtaining the first principal component from a principal component analysis on platform-day level signal across StockTwits, Twitter, and SeekingAlpha. Each panel represents a different outcome: panel A S&P 500 returns, panel B S&P 500 cumulative abnormal turnover, and panel C SPY cumulative abnormal turnover. Everything else mirrors Table 4 and Table 5.Newey-West standard errors with 6 lags in parentheses in columns 1-4 and following Hodrick (1992) in columns 5-6; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A9: Do sentiment and attention indexes predict returns and turnover?
With additional controls

	(1)	(2)	(3)	(4)	(5)
Panel A: Return$_t$					
Sentiment$_t(z)$	0.544***	0.542***	0.543***	0.543***	0.537***
	(0.042)	(0.043)	(0.042)	(0.042)	(0.042)
Attention$_t(z)$	-0.097***	-0.110***	-0.064**	-0.100***	-0.091***
	(0.030)	(0.032)	(0.030)	(0.030)	(0.030)
Sentiment × Attention$_t(z)$	0.158***	0.161***	0.153***	0.156***	0.158***
	(0.038)	(0.038)	(0.037)	(0.038)	(0.038)
ARA$_t(z)$		0.031			
		(0.038)			
AIA$_t(z)$			-0.125**		
			(0.055)		
MAI (WSJ)$_t(z)$				-0.046*	
				(0.028)	
MAI (NYT)$_t(z)$				0.055	
				(0.036)	
Twitter EU$_t(z)$					-0.078***
					(0.028)
RavenPack news$_t(z)$					0.093***
					(0.034)
Observations	2,267	2,267	2,267	2,267	2,267
R^2	0.192	0.192	0.199	0.195	0.202
Panel B: S&P turnover$_t$					
Sentiment$_t(z)$	-0.021***	-0.023***	-0.021***	-0.021***	-0.020***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
Attention$_t(z)$	0.071***	0.057***	0.055***	0.071***	0.069***
	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Sentiment × Attention$_t(z)$	-0.007	-0.003	-0.004	-0.006	-0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
ARA$_t(z)$		0.034***			
		(0.008)			
AIA$_t(z)$			0.071***		
			(0.006)		
MAI (WSJ)$_t(z)$				0.016***	
				(0.004)	
MAI (NYT)$_t(z)$				0.007*	
				(0.004)	
Twitter EU$_t(z)$					0.027***
					(0.003)
RavenPack news$_t(z)$					-0.002
					(0.005)
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$_t$					
Sentiment$_t(z)$	-0.080***	-0.081***	-0.080***	-0.079***	-0.077***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
Attention$_t(z)$	0.054***	0.043***	0.033***	0.054***	0.051***
	(0.009)	(0.010)	(0.008)	(0.009)	(0.009)
Sentiment × Attention$_t(z)$	-0.010	-0.007	-0.006	-0.009	-0.010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ARA$_t(z)$		0.025**			
		(0.010)			
AIA$_t(z)$			0.082***		
			(0.009)		
MAI (WSJ)$_t(z)$				0.028***	
				(0.007)	
MAI (NYT)$_t(z)$				0.001	
				(0.007)	
Twitter EU$_t(z)$					0.049***
					(0.007)
RavenPack news$_t(z)$					-0.025***
					(0.008)
Observations	2,267	2,267	2,267	2,267	2,267
R^2	0.627	0.629	0.645	0.631	0.639
Controls and DOW, MOY, YQ FE	Y	Y	Y	Y	Y

This table repeats Table 4 and Table 5 while including additional attention and sentiment measures: ARA, AIA, MAI (WSJ), MAI (NYT), Twitter EU index, and RavenPack aggregate news sentiment. Everything else follows column 2 of the corresponding tables. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A10: Do sentiment and attention indexes predict returns and turnover?
Heterogeneity by sentiment range

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P return						
Above-mean sentiment$_t(z)$	0.332***	0.387***	-0.067*	-0.061	-0.235	-0.243
	(0.039)	(0.043)	(0.038)	(0.045)	(0.229)	(0.250)
Below-mean sentiment$_t(z)$	0.724***	0.659***	-0.145**	-0.134**	-0.537	-0.562*
	(0.080)	(0.063)	(0.066)	(0.058)	(0.327)	(0.288)
Attention$_t(z)$	-0.094***	-0.031	-0.068**	-0.102**	-0.335	-0.269
	(0.029)	(0.049)	(0.033)	(0.047)	(0.206)	(0.230)
Above-mean sentiment × Attention$_t(z)$		0.060		0.024		-0.040
		(0.041)		(0.036)		(0.211)
Below-mean sentiment × Attention$_t(z)$		0.228**		-0.065		0.133
		(0.104)		(0.080)		(0.425)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.183	0.198	0.036	0.037	0.327	0.327
Panel B: S&P turnover						
Above-mean sentiment$_t(z)$	-0.017**	-0.017**	-0.026***	-0.030***	-0.323***	-0.321***
	(0.007)	(0.008)	(0.009)	(0.009)	(0.050)	(0.049)
Below-mean sentiment$_t(z)$	-0.023***	-0.019**	-0.009	-0.004	-0.017	0.031
	(0.008)	(0.008)	(0.009)	(0.009)	(0.055)	(0.051)
Attention$_t(z)$	0.070***	0.060***	0.042***	0.036***	0.121***	0.018
	(0.007)	(0.008)	(0.006)	(0.007)	(0.033)	(0.041)
Above-mean sentiment × Attention$_t(z)$		0.006		-0.003		0.047
		(0.007)		(0.007)		(0.045)
Below-mean sentiment × Attention$_t(z)$		-0.022**		-0.019**		-0.228***
		(0.010)		(0.009)		(0.047)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.596	0.598	0.483	0.485	0.725	0.726
Panel C: SPY turnover						
Above-mean sentiment$_t(z)$	-0.062***	-0.063***	-0.072***	-0.079***	-0.640***	-0.618***
	(0.012)	(0.013)	(0.015)	(0.016)	(0.093)	(0.090)
Below-mean sentiment$_t(z)$	-0.095***	-0.091***	-0.041**	-0.036**	0.124	0.170*
	(0.015)	(0.015)	(0.016)	(0.016)	(0.095)	(0.090)
Attention$_t(z)$	0.054***	0.046***	0.024**	0.023*	0.157**	0.027
	(0.009)	(0.011)	(0.010)	(0.012)	(0.063)	(0.079)
Above-mean sentiment × Attention$_t(z)$		0.001		-0.010		0.089
		(0.010)		(0.013)		(0.083)
Below-mean sentiment × Attention$_t(z)$		-0.018		-0.014		-0.251***
		(0.014)		(0.016)		(0.094)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.627	0.628	0.533	0.534	0.681	0.682
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
YM FE	Y	Y	Y	Y	Y	Y

This table repeats Table 4 and Table 5 while replacing $Sentiment_t(z)$ with $Above-meansentiment_t(z)$ (the same as $Sentiment_t(z)$ if sentiment index is positive and zero otherwise) and $Below-meansentiment_t(z)$ (same as $Sentiment_t(z)$ if the sentiment index is negative and zero otherwise). Everything else follows the corresponding tables. Newey-West standard errors with 6 lags in parentheses in columns 1-4 and following Hodrick (1992) in columns 5-6; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A11: 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

This table presents robustness checks for Table 6 Panel A by using an alternative outcome and portfolio weight. Panel A winsorizes forecast returns at the 10% level before calculating portfolio weight. Panel B replaces portfolio weights with a rolling 5-day average. Everything else follows Table 6 Panel A.

Table A12: Dynamic trading strategy based on sentiment or attention index *alone*

	Dependent var.: Portfolio excess return$_{t+1}$(%)				
	(1)	(2)	(3)	(4)	(5)
Panel B: Predict with sentiment alone					
Alpha	0.012***	0.013***	0.013***	0.013***	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Market excess return$_t$		-0.009**	-0.010**	-0.010**	-0.163
		(0.005)	(0.005)	(0.005)	(0.100)
SMB$_t$			0.011	0.010	0.006
			(0.008)	(0.008)	(0.009)
HML$_t$			-0.007*	-0.011*	-0.019
			(0.004)	(0.007)	(0.015)
MOM$_t$				-0.005	-0.003
				(0.005)	(0.005)
Observations	2,246	2,246	2,246	2,246	2,246
R^2	–	0.002	0.003	0.003	0.009
Alpha (annualized)	3.095***	3.241***	3.223***	3.213***	3.076***
	(1.077)	(1.108)	(1.110)	(1.111)	(1.151)
Information ratio (annualized)	0.963	0.980	0.972	0.969	0.895
Panel C: Predict with attention alone					
Alpha	0.008**	0.008**	0.008**	0.008**	0.008**
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Market excess return$_t$		-0.007**	-0.008**	-0.008**	-0.153*
		(0.003)	(0.004)	(0.003)	(0.081)
SMB$_t$			0.010	0.010	0.006
			(0.007)	(0.007)	(0.007)
HML$_t$			-0.004	-0.005	-0.022*
			(0.004)	(0.006)	(0.012)
MOM$_t$				-0.001	0.002
				(0.004)	(0.005)
Observations	2,246	2,246	2,246	2,246	2,246
R^2	–	0.002	0.003	0.003	0.009
Alpha (annualized)	1.977**	2.095**	2.090**	2.088**	2.019**
	(0.877)	(0.886)	(0.882)	(0.884)	(0.894)
Information ratio (annualized)	0.755	0.792	0.794	0.791	0.756
FF12 industry excess return$_t$	N	N	N	N	Y

This table repeats Table 6 panel A while using either day t sentiment index alone (panel A), or day t attention index alone to predict day $t+1$ return. Everything else follows Table 6. Newey-West standard errors with 6 lags in parentheses; * $p < .1$; ** $p < .05$; *** $p < .01$.

Table A13: Correlation between overall and (non-)central firm social media indexes

Panel A: Correlation among Sentiment Indexes

	Central firm sentiment		Non-central firm sentiment	
	Pre-ATT	Post-ATT	Pre-ATT	Post-ATT
Overall sentiment	0.856	0.683	0.946	0.824

Panel B: Correlation among Attention Indexes

	Central firm attention		Non-central firm attention	
	Pre-ATT	Post-ATT	Pre-ATT	Post-ATT
Overall attention	0.896	0.783	0.803	0.769

This table reports the correlations between overall social media indexes and (non-)central-firm indexes before vs. after ATT policy change on April 26, 2021. Firms are defined as central if they rank in the top 20 based on between-ness, closeness, or eigenvector centrality in 2019; the complement is defined as non-central. The Pre-ATT period is defined as from May 2020 through April 2021 and the post-ATT period from May 2021 through December 2021 (end of the available data).