

The Social Side of Finance: How Non-Financial Influencer Audiences Shape Investor Attention and Market Outcomes*

by

Zhengfa Zhang, Kevin Keasey, Danilo V. Mascia, Costas Lambrinoudakis

December 2025

Abstract

Social media networks have become integral to modern information exchange. This study develops a conceptual framework to examine how non-financial influencers, and more importantly the characteristics of their audiences, affect firms through investor attention and stock market outcomes. We find that influencers can significantly increase investor attention, especially when audiences are geographically concentrated, affluent, and socioeconomically aligned with investors, or when firm visibility and traditional information channels are weak. These attention shocks generate short-lived market effects, including narrower bid–ask spreads and transitory abnormal returns. The effects are stronger when engagement originates from audience groups that are more likely to invest or are socially aligned with active investors, such as higher-income or more educated audiences. The rapid decay of these responses indicates that attention affects markets primarily through participation and liquidity channels rather than durable information transmission. Overall, results suggest that social media exposure serves as a partial substitute for formal disclosures, underscoring the growing importance of informal attention networks in shaping information flows and market performance in the digital economy.

* Kevin Keasey, Danilo V. Mascia, and Costas Lambrinoudakis are affiliated with Leeds University Business School (LUBS), the International Banking Institute, and the Centre for Advanced Studies in Finance (CASIF). Zhengfa Zhang was previously affiliated with LUBS and CASIF and is currently affiliated with Hong Kong Metropolitan University.

1. Introduction

This paper examines how non-financial influencers - celebrities, lifestyle personalities, and other high-visibility digital figures - and the nature of their audiences affect investor attention and market activity. We argue that when such individuals mention public companies online, they generate exogenous attention shocks that alter investor trading behavior. Crucially, the magnitude and persistence of these effects depend not only on the visibility or content of the influencer's message but also on the characteristics of the audience that receives it - its demographic composition, geographic concentration, and social connectedness. Adopting an audience-centric view of investor attention, we propose that attention is mediated by the social and economic segmentation of digital networks.

We develop a conceptual framework that empirically links influencer activity to investor attention (measured through Google Trends searches), trading volume, bid-ask spreads, and short-term returns. Influencer mentions significantly increase investor attention and market activity, particularly when audiences are geographically proximate to financial centers. Geographic proximity facilitates the transmission of attention through local investor networks, magnifying the influence of informal, non-fundamental content. This effect is further amplified when audiences share demographic characteristics aligned with typical retail investor profiles—specifically, higher income, middle age, and greater education levels.

Firms lacking strong social or geographic ties to institutional investors are especially sensitive to influencer-driven attention shocks. This asymmetry indicates that influencer mentions act as stronger informational triggers for firms not already embedded in investors' informational or social networks. In line with prior research (Bushee and Leuz, 2005; Lang and Lundholm, 1996), these findings suggest that informal visibility can substitute for formal disclosure mechanisms.

Collectively, the results demonstrate that social visibility - rooted in audience composition rather than message content alone - has tangible financial consequences.

Moreover, social connectedness plays a nuanced moderating role in shaping how influencer-driven attention translates into market outcomes. Firms with weaker social and geographic ties to institutional investors exhibit significantly stronger responses to influencer engagement and sentiment, particularly in terms of bid–ask spreads and short-term return reactions. This pattern is consistent with evidence that investor behavior and price formation are conditioned by network structure and social proximity to capital (Hong et al., 2005; Kuchler et al., 2022).

For socially peripheral firms, heightened influencer engagement attracts marginal and less-informed traders, shrinking bid–ask spreads, as predicted by classic market microstructure models of heterogeneous information (Glosten and Milgrom, 1985; Bloomfield et al., 2009). At the same time, influencer sentiment provides salient directional cues that are incorporated into prices over short horizons, consistent with limited attention and sentiment-based trading theories (Hirshleifer and Teoh, 2003; Barber and Odean, 2008; Tetlock, 2007).

By contrast, firms embedded in denser investor networks—through geographic proximity to financial centers, stronger social connectedness, or sustained institutional ownership—are subject to continuous information production, monitoring, and belief updating by professional investors. For these firms, prices already reflect a steady flow of firm-specific signals originating from analyst coverage, institutional trading, and repeated interactions within investor networks. As a result, incremental attention shocks generated by non-financial social media content provide little new information and are quickly arbitrated away, limiting their impact on liquidity and returns (Merton, 1987; Bushee and Leuz, 2005).

In contrast, firms operating in thin informational environments face sporadic attention and weaker investor recognition. For such firms, non-financial influencer-driven visibility represents a meaningful deviation from traditional sources of attention, attracting new market participants and temporarily reshaping the investor base. Because attention is scarce and unevenly allocated, these visibility shocks have economic consequences not because they convey fundamental information, but because they alter who pays attention and when (Da et al., 2011).

These dynamics imply that market sensitivity to social media content depends not only on the salience or sentiment of the message itself, but also on the relational pathways through which attention diffuses across investor networks. In dense networks, information travels through formal and informal professional channels; in sparse networks, attention is more likely to propagate through social exposure (Blankespoor et al., 2014). In this sense, non-financial influencer activity functions as a partial substitute for formal disclosure and investor recognition, increasing visibility and perceived relevance for less informed firms, while playing a negligible role for firms already embedded in robust information ecosystems (Kuchler et al., 2022).

Taken together, the evidence suggests that social media engagement and sentiment produce market effects whose magnitude depends on both audience composition and the firm's social proximity to investors. However, these effects are transitory, typically persisting for only a few days. In sum, non-financial influencer mentions induce observable shifts in investor attention and market outcomes, with attention spikes translating into short-term abnormal returns and heightened volatility, especially for opaque or under-followed firms, for which sudden increases in visibility attract marginal investors and temporarily alter trading behavior in thin informational environments.

Our contribution is threefold. First, we extend limited attention theory (Hirshleifer and Teoh, 2003) by introducing heterogeneity among the recipients of attention, shifting the analytical lens from message senders to audiences. Second, we integrate insights from diffusion and social contagion research (Aral and Walker, 2012; Banerjee et al., 2013) to argue that influencer networks operate as catalysts that transmit emotionally charged, non-fundamental signals across demographically segmented groups. Third, we identify a new class of “informal disclosures” generated outside firms’ control yet capable of influencing valuation through heightened visibility. This reframing aligns with what Cookson et al. (2025) describe as a “social-media-as-lens” approach, using digital platforms not merely to observe social media effects but to illuminate the behavioral mechanisms governing investor attention. Overall, this study advances understanding of investor attention by shifting focus from the supply of information to the structure of its audience. In an information ecosystem where influencers and non-financial voices dominate visibility, market reactions depend not only on what is disclosed but on who receives and amplifies it.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant literature and develops the conceptual framework. Section 3 describes the data and empirical strategy. Section 4 presents the main results, and Section 5 concludes.

2. Literature Review and Conceptual Framework

Traditional disclosure theory assumes that firms intentionally release information to reduce asymmetry between insiders and investors (Dye, 2001; Verrecchia, 2001). Behavioral theories of attention, however, recognize that investors have finite cognitive capacity and allocate attention selectively, focusing on stimuli that are salient, emotionally engaging, or easily interpretable (Barber and Odean, 2008; Hirshleifer and Teoh, 2003; Sims, 2003). The intersection of limited attention

and salience creates conditions under which non-fundamental signals - such as media visibility, social prominence, or visual design - can influence asset prices independent of informational content. From this perspective, influencer activity represents a distinctive attention stimulus: exogenous, ‘emotionally charged’, and disseminated through networks optimized for engagement rather than analysis.

Empirical evidence underscores the importance of attention in financial markets. Antweiler and Frank (2004) show that online discussions among retail investors forecast returns, while Das and Chen (2007) quantify sentiment from message boards and demonstrate its short-term impact on prices. Da et al. (2011) formalize investor attention using search-volume indices, linking fluctuations in focus to abnormal returns and trading activity. Chen et al. (2014) and Blankespoor et al. (2014) further demonstrate that digital dissemination accelerates price discovery, connecting disclosure economics to online attention dynamics.

Subsequent research extends these insights to social platforms. Bartov et al. (2018) and Gu and Kurov (2020) find that Twitter sentiment correlates with intraday returns and trading intensity. Cookson and Niessner (2020) identify mispricing cycles on StockTwits driven by conversational momentum, while Anand and Pathak (2022) document that Reddit’s WallStreetBets channels collective attention to specific firms, triggering coordinated market reactions. Collectively, this literature establishes that social media serves both as a conduit for information diffusion and as a mechanism for attention-based market distortion, where reactions stem from salience rather than informational precision.

While this research confirms that visibility matters, it often assumes that information recipients are homogeneous. Yet audience heterogeneity is fundamental to understanding how attention translates into market behavior. Investor groups differ in analytical capacity, access to trading platforms, and susceptibility to sentiment. Behavioral finance research highlights that retail

investors, in particular, are more responsive to salient cues and emotionally charged narratives (Barber and Odean, 2008; Hirshleifer et al., 2009). Digital algorithms intensify this tendency by privileging engagement over accuracy, segmenting exposure along demographic and geographic lines.

Influencers function as high-centrality nodes that distribute salient signals through demographically distinct audiences (Hirshleifer et al., 2025). When these audiences overlap with investor populations, visibility shocks are more likely to trigger trading responses. This dynamic reflects the “social proximity” effect identified by Hong et al. (2005), where investors act more readily on information circulating within their own social or geographic networks. Consistent with Merton’s (1987) investor recognition hypothesis, firms that are informationally opaque or under-followed benefit most from incremental visibility, as attention itself conveys legitimacy and relevance.

2.1 Limited Attention and Salience

Traditional finance models assume that investors process all available information rationally and continuously. Behavioral approaches instead emphasize cognitive limitations that restrict the ability to process the vast flow of market data. Limited attention theory (Hirshleifer and Teoh, 2003; Sims, 2003) posits that attention is a scarce resource, selectively directed toward stimuli that are salient, simple, or emotionally charged. As a result, investors often respond more strongly to cues that capture attention than to those that maximize informational precision.

Digital media environments amplify these behavioral biases. Algorithms prioritize vivid, affect-laden, and shareable content, elevating attention-grabbing signals regardless of their informational substance. Empirical evidence supports this dynamic: Barber and Odean (2008) find that retail investors are net buyers of “attention-grabbing” stocks, while Da et al. (2011) demonstrate that search-volume spikes predict returns even in the absence of new information.

Non-financial influencers exemplify this form of exogenous salience. Their communications - endorsements, casual mentions, or critiques - combine emotional resonance, simplicity, and familiarity. Although unrelated to fundamentals, such content can redirect investor focus and reallocate attention within the market, producing short-term price and volume effects consistent with behavioral predictions (Hirshleifer et al., 2009).

2.2 Social Diffusion and Contagion Mechanisms

Attention must spread through social structures to affect market outcomes. Diffusion theory (Aral and Walker, 2012; Banerjee et al., 2013) and social contagion research show that information propagates unevenly across networks depending on connectivity, influence, and homophily. Highly connected individuals accelerate diffusion by linking otherwise separate audiences.

Influencers are archetypal high-centrality actors. Their posts reach large, demographically segmented audiences whose internal network structures shape how information spreads. When a non-financial influencer references a firm, the message diffuses through reposts, comments, and algorithmic amplification. The extent of this diffusion depends on the influencer's reach, audience engagement, and alignment between follower identity and investment behavior.

Social learning and informational cascade theories (Bikhchandani et al., 1992) suggest that individuals draw cues from the reactions of others as well as from content itself. When influencer posts generate high engagement, they signal collective interest that investors may interpret as latent demand. This process can generate temporary demand shocks and volatility, as observed in social trading environments such as Reddit's WallStreetBets (Anand and Pathak, 2022) and Twitter-based trading sentiment (Bartov et al., 2018; Gu and Kurov, 2020).

2.3 Audience Heterogeneity and the Attention Transmission Channel

Most prior attention research treats investors as a homogeneous audience. Yet attention transmission depends critically on the composition and connectivity of those exposed to the signal. Audience heterogeneity moderates the strength, persistence, and direction of market effects.

Demographic and socioeconomic traits influence the likelihood that attention converts into trading. Audiences with greater financial literacy, disposable income, and market access are more capable of acting on attention shocks. Younger or less financially engaged groups may exhibit interest without trading behavior.

Geographic proximity also shapes responses. Investors are more reactive to locally salient information, reflecting “home bias” and “local information advantage” (Ivković and Weisbenner, 2007). When influencer audiences are concentrated near financial centers or firms’ operations, attention shocks are more likely to generate coordinated investor actions.

Finally, network embeddedness amplifies diffusion. Densely connected audiences that frequently share or discuss content facilitate repeated exposure, strengthening attention and trading responses. Kuchler et al. (2022) find that local social connectedness predicts correlated trading and price spillovers, underscoring how network density intensifies behavioral contagion.

Together, these elements create a two-stage attention mechanism: influencer visibility generates an exogenous salience shock, and audience composition determines how that shock translates into market outcomes. Investor attention thus functions as a socially mediated, heterogeneous process rather than a uniform cognitive reaction.

2.4 Disclosure Economics and Informal Signaling

Disclosure theory traditionally focuses on firm-initiated communication intended to reduce information asymmetry (Dye, 2001; Verrecchia, 2001). Yet in digital environments, third-party communications can perform similar signaling functions. According to signaling theory (Spence, 1973), observers infer value from visible actions or associations; in low-information contexts, visibility itself becomes a proxy for legitimacy or momentum.

When a non-financial influencer references a firm, investors may interpret the association as evidence of popularity or brand strength, especially for firms with limited analyst coverage (Merton, 1987). These firms operate in “thin information markets,” where attention can temporarily substitute for formal disclosure. For highly visible or transparent firms, such effects are weaker and dissipate quickly through arbitrage. Thus, the informational environment moderates the extent to which influencer visibility affects prices.

This reasoning connects behavioral and informational economics: influencer visibility functions as a salient but informal signal that shapes valuation primarily through the attention channel. As attention drives trading and price movements, those movements, in turn, attract further attention, generating short-lived feedback loops until fundamentals reassert dominance.

2.5 Integrated Conceptual Model

Synthesizing these perspectives, the proposed framework links non-financial influencer activity to market outcomes through an **attention–diffusion–signaling mechanism** moderated by audience and firm characteristics.

Non-financial influencers create emotionally engaging, high-visibility content that captures limited investor attention - an exogenous salience shock. This salience then diffuses across social networks through reposts, comments, and algorithmic amplification, its persistence determined by

audience engagement and connectivity. The transmission of attention into market behavior depends on audience composition - financial literacy, geographic clustering, and network structure - which collectively determine whether visibility leads to observable trading activity.

For firms with low baseline visibility or limited analyst coverage, this attention functions as a substitute signal of legitimacy or popularity, influencing valuation and liquidity. Elevated attention manifests in measurable indicators such as search activity, trading volume, liquidity shifts, and short-term return fluctuations.

This integrated model predicts that the financial impact of non-financial influencer activity is **conditional and asymmetric** - most pronounced where attention is scarce, audiences are receptive, and information environments are thin, and attenuated where information is abundant or arbitrage capital rapidly neutralizes sentiment.

3. Data and Variables

This section outlines the construction of the dataset and variables. We construct a multi-level panel dataset combining influencer-generated content, audience demographics, and financial market outcomes.

3.1 Social Media Dataset

Our source of influencer-generated content is Instagram, a leading platform for lifestyle and entertainment influencers. We gather post-level data using the Supermetrics.com interface. For each post, we collect the URL, account name, caption, number of likes and comments, and the date and time of upload. Figure 2 provides an example of an Instagram post and its constituent parts.

[Figure 2 about here]

Figure 2 shows an identified post by David Beckham (Instagram account: @davidbeckham). In our sample, we collect all relevant information, including the URL, account name (@davidbeckham), number of likes (78,566) and number of comments (not shown in the figure) as of February 2025, post caption (“Play closer... it’s time for a new era of football with EA Sports FC24 ⚽ @EASportsFC #FC24”), posting date (August 29, 2023), and exact posting time (not shown in the figure).

We identify posts that refer to firms using hashtags included in each caption. A hashtag is a word or phrase preceded by the “#” symbol, typically appearing in the caption text that accompanies an image, video, or album post on Instagram. For instance, in Figure 2, we use #FC24, a game-related hashtag from Electronic Arts (EA), to identify the firm-related post and link it to the company. Hashtags serve as keywords or key phrases that categorize a post’s content and link it to specific topics, allowing users interested in those topics to find related posts. To identify hashtags associated with specific firms, we use the ChatGPT-4o-mini model via the official OpenAI Python package. These identified hashtags are then used to link posts to the corresponding firms. Hashtag identifications were initially performed by the model and subsequently verified through manual review.

3.2 Engagement Measures

To measure user engagement with a post, we use the number of comments rather than the number of likes a post was able to attract. Comments are considered a form of deep interaction, in contrast to likes, which are viewed as a form of shallow interaction (Thorgren et al., 2024), as the former are more resistant to automated manipulation and more indicative of cognitive engagement and potential further offline discussion. All engagement metrics reflect values captured in February

2025. Thorgren et al. (2024) show that Instagram posts are short-lived, with the vast majority of comments received within the first few hours after upload.¹ This means that even though we collect data long after a post has been uploaded, the number of comments typically remains relatively stable after the first day. As a result, our data still offer a reliable measure of user engagement on the day of the post's upload.

Audience characteristics, including geographic distribution, income level, age, and education, were obtained from HypeAuditor, a platform that provides AI-powered analytics and tools for influencer marketing. Like other influencer marketing platforms, HypeAuditor is primarily used to analyze the performance, authenticity, and demographics of influencers and their audiences across platforms such as Instagram, YouTube, TikTok, and others. We selected HypeAuditor over other platforms because it is widely regarded as a leader and pioneer in social media fraud detection, particularly in identifying fake follower accounts, making it one of the most trusted tools for vetting influencers and their audiences.² Given that our analysis relies on audience demographics, ensuring the authenticity of this data is essential.

Data on audience characteristics were retrieved in March 2025. While precise follower composition may evolve over time, we assume relative demographic stability across years. For example, it is unlikely that an influencer with a predominantly Spanish-speaking following in 2018 would attract a predominantly Cantonese-speaking audience by 2024. There is also empirical evidence from Koiranen et al. (2019) that socio-demographic differences among social media users, such as variations by gender, education level, and region, remain relatively stable over time. Notably, their study covers the period from 2008 to 2016, a time of rapid growth in social media

¹ Thorgren et al. (2024) report that over 50% and 75% of comments are received within the first 2 hours and 7.5 hours after a post is uploaded, respectively.

² See <https://www.favikon.com/blog/best-influencer-marketing-platforms> and <https://www.influencer-hero.com/blogs/top-influencer-marketing-analytics-tools>

platform usage. The fact that audience characteristics remained relatively stable during this early and fast growth phase offers some reassurance that they would also remain stable during our sample period, 2015–2024, when social media growth was more mature and user composition likely less volatile.

Because HypeAuditor reports location data only for the top five follower states per influencer, we restrict our geographic analysis to California (CA), New York (NY), Texas (TX), and Florida (FL). These four states were, indeed, the most frequently reported, appearing among the top five follower states in over 50% of the sampled accounts. Furthermore, California, New York, Texas, and Florida are the most populous states in the U.S., account for approximately one-third of the national population and are the most consistently represented among influencer audience data. These states also offer geographic, demographic, and cultural heterogeneity, which strengthens external validity.

We define a firm-day measure of engagement ($Eng_{i,t}$) as the total number of comments received on all posts related to firm i that were uploaded on day t :

$$Eng_{i,t} = \sum_{k \in P(i,t)} Comments_k \quad (1)$$

$$StateEng_{i,t} = \sum_{k \in P(i,t)} Comments_k \times State\ Audience_k \quad (2)$$

Here, $P(i,t)$ denotes the set of posts about firm i uploaded on day t , $Comments_k$ is the number of comments (in thousands) on post k , where post k concerns firm i , $State\ Audience_k$ is the proportion of influencer's $\mathcal{A}$ audience (who uploaded post k) located in a specific state.

3.3 Sentiment Measures

We use GPT via the official OpenAI Python package to estimate the sentiment score of each post's caption. The prompt and detailed settings used for linking posts to firms and estimating post sentiment are provided in Appendix 1 (Korinek, 2023; Lopez-Lira and Tang, 2023). We also estimate post sentiment using the dictionary-based method VADER (Hutto and Gilbert, 2014) as a robustness check. Details on the construction of the VADER sentiment score are provided in Appendix 1. The sentiment score values for a post range from -1 to $+1$, with -1 indicating the most negative sentiment and $+1$ the most positive. To aggregate sentiment at the firm-day level ($AbsSent_{i,t}$), we compute the weighted average sentiment across all posts related to firm i uploaded on day t , where the weights reflect the level of user engagement each post received. Specifically, the weight for each post is calculated by dividing the number of comments on that post by the total number of comments on all posts related to firm i uploaded on day t . The weighted average scores are then converted to absolute values to capture emotional intensity regardless of polarity.

$$AbsSent_{i,t} = \left| \frac{\sum_{k \in P(i,t)} Sentiment_k \times Comments_k}{\sum_{k \in P(i,t)} Comments_k} \right| \quad (3)$$

$$StateAbsSent_{i,t} = \left| \frac{\sum_{k \in P(i,t)} Sentiment_k \times Comments_k \times State\ Audience_k}{\sum_{k \in P(i,t)} Comments_k \times State\ Audience_k} \right| \quad (4)$$

Here, $P(i,t)$ denotes the set of posts about firm i uploaded on day t , $Sentiment_k^i$ is the sentiment score of post k , where post k concerns firm i , $Comments_k$ is the number of comments (in thousands) on post k , where post k concerns firm i , $State\ Audience_k$ is the proportion of influencer's $\mathcal{Z}$ audience (who uploaded post k) located in a specific state.

The absolute sentiment for a firm-day ($AbsSent_{i,t}$) reflects the weighted average sentiment intensity across all relevant posts, with greater weight assigned to posts that generated higher user engagement. Since sentiment scores are transformed into absolute values to capture the magnitude

of sentiment irrespective of direction, all sentiment metrics range from 0, indicating neutrality, to 1, indicating intense sentiment, either negative or positive depending on the underlying polarity.

3.4 Firm-level Variables

Our sample consists of S&P 500 constituents. We exclude firms in the finance and energy sectors to avoid contamination from sector-specific patterns of investor attention that are largely orthogonal to the economic mechanisms of interest. In the case of financial institutions, their exclusion is standard in empirical finance due to their distinct capital structures, regulatory regimes, and balance-sheet characteristics (Booth et al., 2001; Fama and French, 1992). However, in the current media environment, an additional complication arises: the attention metrics commonly used in empirical research – such as search volumes, news counts, or social media mentions – can be systematically biased for financial firms because of tagging practices on digital platforms. On Twitter, Reddit, and other social media, users frequently append hashtags or ticker symbols of major banks and other financial institutions when discussing broader economic, political, or market issues that are not directly related to those firms’ fundamentals. For example, discussions of interest rate policy, inflation, or even unrelated stock events often include hashtags like #JPM, #GS, or #HSBC alongside unrelated tickers, leading to inflated attention measures for those institutions. This tagging spillover is particularly prevalent for financial firms because their names and ticker symbols are perceived as shorthand for “the market” or “Wall Street” in general discourse (Chen et al., 2014; Cookson and Niessner, 2020). Such misattribution introduces measurement error in attention proxies, artificially correlating these measures with macro events or unrelated firm news, and risks biasing inference if these firms are retained in the sample.

Energy firms face a similar challenge during the ongoing geopolitical conflict, as investor and media attention to oil, gas, and related commodities is frequently bundled with unrelated political or market commentary, generating volatile and noisy attention signals (Caldara and

Iacoviello, 2022; Liu, 2024). Excluding both sectors mitigates the risk that sector-specific attention spillovers, rather than the intended firm-level mechanisms, drive our empirical results.

Industry classification follows the 12-industry scheme of Fama and French (1997). Daily return, volume, and volatility variables are constructed with data obtained from the Center for Research in Security Prices (CRSP). Firm headquarters information is sourced from Compustat and geographic data via Nominatim and the Geopy library.

We include the following firm-level controls: $\ln(\mathbf{MV}_{i,t-7})$, which is the natural log of firm i 's market capitalization, lagged by seven days; $\mathbf{Turn}_{i,t-7:t-36}$, that is the trading turnover of firm i 's stock over the prior four weeks; $\mathbf{RealVol}_{i,t-7:t-36}$, which is realized volatility, computed as the mean absolute return of firm i 's stock over the same period; $\mathbf{Ret}_{i,t-7:t-36}$, which is the cumulative return of firm i 's stock over the prior four weeks; and $\mathbf{EA}_{i,t}$, a dummy for whether firm i announced earnings on day t .

3.5 Dependent Variables: Attention Measures

We use Google Search Volume Index (SVI) data to measure retail investor attention, following Da et al. (2011), Madsen and Niessner (2019), and Liu et al. (2023). Searches using a firm's ticker symbol, rather than its name, serve as a standard proxy for retail investor attention. We focus on retail investors because they are the investor group typically affected by social media content (Cookson et al., 2024). We collect daily SVI data for ticker symbol queries for the four focal states, namely CA, NY, TX and FL.³ We then use such state-level values to compute our

³ We also collected and performed national level attention tests in Table OA1 in the Online Appendix.

main dependent variable, that is *FourStateSVI_{i,t}*, as the average SVI value for firm i on day t across the four focal states.

3.6 Sample and Descriptive Statistics

We begin by downloading all posts from the top 3,000 Instagram influencer accounts, ranked by follower count, covering the period from January 2015 to December 2024. Due to political restrictions and account limitations imposed by Instagram, such as restricted or protected status, we obtained a total of 10,190,241 posts from 2,563 accounts. From this dataset, we identified 31,193 posts from 1,329 accounts that included hashtags related to S&P 500 firms. After aggregating the data at the firm-day level (as discussed earlier in this section), our final sample comprises 20,534 firm-day observations across 138 firms.

Table 1 reports the summary statistics for our sample. For each variable, we present the number of observations (N), mean, standard deviation, and quartiles (p25, p50, p75). Panel A reports Instagram post characteristics at the firm-day level, including engagement (i.e. number of comments in thousands) and sentiment measures. The mean engagement value is 0.802, indicating that, on average, firm-related posts receive approximately 800 comments per event day. The number of sentiment observations is 198 fewer than that for comments because some posts contain only hashtags without any textual content, making it impossible for ChatGPT-4o-mini (via Python) to provide sentiment scores. In the empirical analysis, for non-event days (i.e., days with no Instagram posts detected), we assign a value of zero to both engagement and sentiment variables.

Panel B presents the geographic distribution of Instagram audiences, reporting the percentage of influencers' followers from the four states: California (CA), New York (NY), Florida

(FL), and Texas (TX). Data are reported at the firm-day level. On days when multiple influencers post about a specific firm, the audience percentage is calculated based on engagement-weighted averages. California has the largest average audience share at 5.66%, which is consistent with its population size and the concentration of social media influencers and firms in the state. Although Texas is the second most populous state, it ranks fourth in audience share at only 3.13%, suggesting comparatively lower Instagram or social media usage in Texas.

In terms of demographics, the age distribution shows that young audiences (ages 13–34) dominate, accounting for an average of 73.87% of followers, whereas middle-aged audiences (ages 35–54) represent only 21.91%. With respect to education, 28.19% of followers hold a degree, while 71.78% do not. The income distribution indicates that middle-income audiences (USD 50,000–150,000) comprise 30.39% of followers, compared with 59.77% in the low-income group (USD 0–50,000) and only 9.84% in the high-income group (above USD 150,000). In the empirical analysis, missing audience data (i.e., cases where CA, NY, FL, or TX are not among an influencer’s top five follower states) are coded as zero.

Panel C provides statistics for the Google Search Volume Index (SVI), capturing investor attention in the same four states. The SVI ranges from 0 to 100 and is standardized by Google Trends. A value of zero does not necessarily imply no searches; it may reflect a volume that rounds below 0.5 after standardization.

Panel D summarizes firm-level financial variables, including $\ln(\mathbf{MV}_{i,t-7})$, $\mathbf{Turn}_{i,t-7:t-36}$, $\mathbf{RealVol}_{i,t-7:t-36}$, $\mathbf{Ret}_{i,t-7:t-36}$, and $\mathbf{EA}_{i,t}$. The high average market capitalization reflects the focus on S&P 500 firms.

[Table 1 about here]

Table 2 presents the correlation matrix between the main variables used in the study, including social media engagement metrics, sentiment measures, and firm-level controls. The pairwise correlations are generally low, reducing concerns about multicollinearity in the regressions.

[Table 2 about here]

4. Empirical Results

Our empirical results on the impact of influencers are split into two main sections – investor attention and market reaction. For each of these areas, we examine where appropriate, how the impact of influencers varies with the geographic allocation of influencers’ audiences, audience characteristics, and firms’ social and physical proximity to institutional investors. We present tests that examine the longevity of the effects of posts on stock returns.

All our regressions control for firm, year, week, and day-of-the-week fixed effects, given the nature of social media activity, which varies across weekdays, weekends, and different times of the year. Week and weekday fixed effects are included to account for temporal variation throughout the year. For example, engagement and sentiment levels around New Year’s Eve may differ from those on a regular business day, and post-performance may also vary between weekends and weekdays. Comparing with prior research (e.g., Focke et al., 2020; Liu et al., 2023), we additionally include a year fixed effect to control for time trends and to account for the unique circumstances of the pandemic period. All the standard errors in the following regressions are clustered by firm.

4.1 Investors’ Attention

4.1.1 Social Media Posts and Investors’ Attention – The Role of Audience Geographic Allocation

We start our empirical investigation by examining the relationship between investors' attention and overall user engagement with posts. More specifically, following Equation (5), we regress the daily attention for firm i on day t across all four states (as proxied by $FourStateSVI_{i,t}$) on the level of user engagement with posts related to firm i published on day t . We control for the level of post sentiment, as more extreme sentiment tends to attract more attention (Knutson et al., 2024; Lim and Bentley, 2022).⁴ To address potential autocorrelation in the dependent variable, we include the first and second lags of $FourStateSVI$ as controls. Additionally, following the approach of Focke et al. (2020), we incorporate the seventh lag of SVI to account for unobserved time-varying, firm-specific mean reversion that may not be fully captured by control variables or time fixed effects. We also use a set of firm-level controls lagged by 7 days, in a similar fashion to Focke et al. (2020).⁵ A dummy to account for earnings announcements (EA) is also added to the models.

$$FourStateSVI_{i,t} = \alpha + \beta Eng_{i,t} + \gamma AbsSent_{i,t} + \sum_{j=1}^2 \delta_j FourStateSVI_{i,t-j} + \zeta FourStateSVI_{i,t-7} + \theta Controls_{i,t-7} + \xi EA_{i,t} + \epsilon_{i,t} \quad (5)$$

Next, we examine the relationship between state-adjusted engagement and investor attention. Following Equation (6), we regress the daily investor attention for firm i on day t , aggregated across all four states, on the level of engagement with posts related to firm i published on day t , where the engagement with these posts was likely generated by audiences from the four focal states, as well as on the sentiment scores of these posts.⁶ $FourStateEng_{i,t}$ represents the sum of

⁴ To address potential sentiment measurement bias, a robustness test using the VADER sentiment detection method (i.e., a dictionary-based approach tailored for social media content), has been included in Table OA2 in the Online Appendix. Results remain consistent.

⁵ Following Focke et al. (2020), we lag the control variables by 7 days to avoid endogeneity with the engagement and sentiment variables; engagement and sentiment might directly affect these variables.

⁶ $FourStateEng_{i,t}$ is the number of comments on all posts about firm i uploaded on day t by influencers whose audience is predominantly located in the four states, and $FourStateAbsSent_{i,t}$ is the absolute value of the average sentiment score of these posts.

$StateEng_{i,t}$ values from all four states. $FourStateAbsSent_{i,t}$ is calculated as the absolute value of the average sentiment across all four states.

$$FourStateSVI_{i,t} = \alpha + \beta FourStateEng_{i,t} + \gamma FourStateAbsSent_{i,t} + \sum_{j=1}^2 \delta_j FourStateSVI_{i,t-j} + \zeta FourStateSVI_{i,t-7} + \theta Controls_{i,t-7} + \xi EA_{i,t} + \epsilon_{i,t} \quad (6)$$

Table 3 reports the results. In columns (1), (3), and (5), we use equation (5), while in columns (2), (4), and (6), we use equation (6). Columns (1) and (2) present results where the dependent variable is measured on the same day as the post upload (t), while columns (3) and (4), and (5) and (6), report the same set of results with the dependent variable measured one day and two days after the post upload, respectively.

[Table 3 about here]

The results in Table 3 show that the level of engagement has a positive effect on investor attention across the board, which persists for at least two days following the day the post was uploaded. This indicates that the effect not only materializes immediately but also endures for at least two days. Notably, this effect becomes stronger, both in magnitude and statistical significance, when engagement is adjusted for the geographical distribution of the audience. For instance, at the overall engagement level, one thousand comments are associated with a 0.011 increase in SVI at the 5% significance level, whereas one thousand comments generated by audiences from the targeted states (CA, NY, TX, FL) are associated with a 0.096 increase in SVI at the 1% significance level, which is more than nine times higher. The results from t-tests comparing the coefficients on overall engagement and engagement from the audiences in the four states also support that the latter has a significantly greater impact on SVI. This pattern holds consistently across all three time-lag specifications.

These results suggest that a social media signal that can potentially attract the attention of a particular group of investors will be transmitted more quickly and efficiently when it is received by individuals who are geographically close to those investors. This may be driven by geographical proximity and/or shared cultural or informational contexts that facilitate the diffusion of information. All in all, this evidence is consistent with audience-centric attention, which posits that the market relevance of a social media signal is shaped not only by who posts it, but also by who receives it. Geographic proximity facilitates faster attention transmission through local investor networks, amplifying the salience of otherwise informal content.

[Table 4 about here]

Table 4 presents additional results, based on separate regressions for each state, where we use each state’s SVI individually rather than the aggregate SVI across the four states, and state-specific engagement – computed as per equation (2) – rather than aggregated engagement across the four states. The results are consistent with those in Table 3. Using California, the state with the highest number of influencers, as an example, standard engagement has a positive coefficient of 0.012 on California’s SVI, significant at the 10% level. In contrast, state-specific engagement exhibits a stronger and more statistically significant effect, with a coefficient of 0.412. This state-level evidence reinforces one of the central claims of our analysis, namely that attention spillovers are conditioned by where the audience is.

4.1.2 Audience Demographics and Attention

We next explore in Table 5 whether the demographic composition of an influencer’s audience moderates the spillover of attention. Following equation (7), we run three separate

regressions, one for each demographic classification: audience age, income, and education. In each case, we regress the *FourStateSVI* on measures of *FourStateEng* and *FourStateAbsSent* adjusted by group within the demographic classification. The different groups (i.e., age, income, education) are denoted by g , and we have n different groups in each demographic classification, where $n = 2, 3$, and 2 for audience age, income, and education, respectively.⁷ More specifically, in column (1) we distinguish between engagement from ‘young’ individuals (aged 13–34) and ‘middle-aged’ individuals (35–54). In column (2), engagement is broken down into income groups: low (\$0–50,000 per year), middle (\$50,000–150,000), and high (above \$150,000). Finally, column (3) differentiates engagement between users with and without a degree.

$$\begin{aligned}
\text{FourStateSVI}_{i,t} = & \alpha + \sum_{g=1}^n \beta_g \text{FourStateEng}_{i,t}^g \\
& + \sum_{g=1}^n \beta_g \text{FourStateAbsSent}_{i,t}^g + \sum_{j=1}^2 \delta_j \text{FourStateSVI}_{i,t-j} \\
& + \zeta \text{FourStateSVI}_{i,t-7} + \theta \text{Controls}_{i,t-7} + \xi \text{EA}_{i,t} + \epsilon_{i,t}
\end{aligned} \tag{7}$$

[Tables 5 about here]

The results, reported in Table 5, reveal significant heterogeneity. Engagement from middle-aged and higher-income audiences exhibits markedly larger spillover effects than that from younger or lower-income groups. For instance, the coefficient for high-income audiences is 0.798,

⁷ When calculating $\text{FourStateEng}_{i,t}^g$ and $\text{FourStateAbsSent}_{i,t}^g$, additional weights are applied to the number of comments and the sentiment score associated with each post. These weights account for the proportion of each influencer’s audience belonging to a specific audience group, complementing the existing weights that capture the share of the audience originating from the four states. For instance, assuming we focus on the group classification g corresponding to audience age, then $\text{FourStateEng}_{i,t}^g$ captures user engagement with posts by influencers whose followers include both individuals from the four focal states and whose age is either 13–34 or 35–54, therefore excluding posts by influencers with minimal or no following in the four states or whose following is outside those age brackets. Consequently, a higher value of this variable indicates a greater likelihood that the observed engagement originates from young or middle-aged people residing in the four states. Similar reasoning applies for the $\text{FourStateAbsSent}_{i,t}^g$ variable.

nearly four times that of low-income audiences (0.205), with an F-test confirming the difference at the 1% level. Similarly, audiences with a college degree generate stronger effects than those without.

These findings suggest that audience sophistication, proxied by working experience, income and education, amplifies the transmission of attention shocks. The greater the socioeconomic alignment with potential investors, the more likely engagement will lead to action, such as search activity or trading behavior. They also directly support the audience-centric view of attention, which argues that signals are more likely to produce investor responses when they reach demographically aligned, financially capable audiences. The stronger effects from middle-aged, affluent, and educated followers show that audience structure drives salience conversion into investor action.

4.1.3 Social and Physical Proximity to Capital

To assess how firm-level proximity to investors moderates attention spillovers, we partition firms based on three measures: Social Proximity of Capital (SPC), Physical Proximity of Capital (PPC), and Social Connectedness Index (SCI) to the Tri-State financial hub. Please refer to Appendix 3 for a thorough explanation of these three measures.

Following Kuchler et al. (2022), we set a firm's social proximity to institutional capital as equal to the SPC value of the firm's headquarters location county. Similarly, we set a firm's physical proximity to institutional capital as equal to the PPC value of the firm's headquarters location county. For SCI, we set a firm's social proximity to institutional capital as equal to the value of the SCI to the Tri-State Area of the firm's headquarters location county. For each measure, we use the

median value of the latest available SPC and PPC value to divide firms into groups of high and low proximity investors and then estimate equation (11) separately for each group.

[Table 6 about here]

As shown in Table 6, in all three dimensions of connectedness, the coefficient on *Four_State Engagement_{it}* is significantly positive for firms with low social connectedness (columns 2, 4, and 6), but statistically insignificant for firms in the high connectedness groups (columns 1, 3, and 5). This pattern suggests that firms lacking strong social or geographic ties to institutional investors are more sensitive to influencer-driven attention shocks. This asymmetric effect suggests that mentions from influencers serve as a more potent informational shock for firms that are not already embedded in investors' informational or social networks.

These results show that the amplification effect of audience-driven salience is strongest for firms lacking traditional investor-facing channels. When firm visibility is low, the social media audience can compensate by serving as an alternative conduit of market attention. Furthermore, this finding supports the intuition drawn from Merton's (1987) investor recognition hypothesis: firms that are less visible to investors derive greater benefit from reductions in information asymmetry when exposed through public signals, such as social media posts. These results underscore the asymmetric influence of social media based on firm location and network embeddedness.

Regarding potential concerns about changes in social connectedness between counties, prior studies have shown that social connectedness is persistent and does not change significantly over short time periods, as discussed by Bailey et al. (2018a, 2018b), Chetty et al. (2022), and Hirshleifer et al. (2025). We also conduct robustness checks by using prior mean values to validate the

consistency of results that use the last available value to split the groups. By using the mean value from 2013 to 2016 of SPC and PPC to separate the group, instead of relying on the latest available value as we did in Table 6, we conduct another set of regressions to demonstrate the robustness of our results. As shown in Table OA3 in the Online Appendix, the main results remain consistent with those in Table 6. These robustness tests confirm the findings obtained using the original grouping method.

4.2 Financial Market Reactions

The previous results demonstrate that non-financial posts can generate spillover effects on investor attention, and that these effects are significantly shaped by audience geographic and demographic characteristics as well as the firm's connectedness with investors. The next step is to examine whether these attention effects translate into actions by investors. To assess financial market reactions to non-financial posts, we first analyse the overall effects on market variables and then conduct subsample tests based on firms' social connectedness in Section 4.2.1. We then incorporate audience demographic characteristics in Sections 4.2.2 and 4.2.3 to further explore their influence on bid-ask spreads, trading turnover and abnormal returns.

4.2.1 Overall Market Reaction and the Social Connectedness of Firms

In this section, we examine whether the spillover effect on investors' attention, moderated by audience characteristics, also translates into actual investor behaviour. We focus on three dimensions of financial behaviour: bid-ask spreads, trading turnover, and abnormal returns, adjusted using the Fama-French-Carhart four-factor model. To more precisely assess the impact on financial behaviour, particularly abnormal returns, we rely on sentiment values rather than their absolute values.

Motivated by our finding that the effect of posts on investor attention is stronger for firms with weaker ties to institutional investors, we further examine whether posts influence financial variables differently across firms with varying degrees of connectedness to institutional investors. The low and high social connectedness groups are defined by the first and last deciles of the Social Connectedness Index (SCI) to the Tri-State area. The low group consists of the least connected counties (lowest decile), while the high group includes the most connected counties (highest decile). The imbalance in observation counts arises from the uneven distribution of firms across counties.

Table 7 presents panel regression estimates evaluating the effect of non-financial influencer activity on firms' financial market variables over the period January 2015 to December 2024. The analysis is stratified by the level of firm-investor social connectedness (SCI), with separate results for firms in the highest and lowest decile for SCI of Tri-State area groups. All regressions control for firm, year, week, and weekday fixed effects, and cluster standard errors at the firm level.

[Table 7 about here]

In columns (1)–(3), we assess the impact on bid-ask spreads. While the baseline regression for the full sample (column 1) shows no significant relationship, we observe a statistically significant negative effect of engagement on spreads among low-SCI firms, indicating decreased market frictions following heightened engagement with influencer posts. Conversely, sentiment exhibits a significant positive relationship with spreads in the same subsample, suggesting that more extreme sentiment increases the frictions. These effects are absent in the high-SCI group (column 3).

Columns (4)–(6) report results for trading turnover. Across all specifications, the coefficients on both engagement and sentiment are statistically insignificant, implying that trading volume does not respond significantly to influencer post characteristics, regardless of the connectedness with investors.

The results on abnormal returns are reported in columns (7)–(9). Sentiment is positively associated with abnormal returns in the full sample, with stronger effects among firms in the low SCI group. This indicates that investors in less connected firms are more reactive to the tone of influencer posts, potentially due to greater reliance on external sentiment signals in the absence of robust informational ties. No significant effects are observed among high SCI firms (column 9), reinforcing the hypothesis that social connectedness buffers firms against exogenous attention shocks.

Overall, these results underscore the nuanced role of social connectedness in moderating financial market reactions to non-financial influencer content. Firms with weaker investor ties appear more sensitive to both engagement and sentiment metrics, particularly in terms of spread and return effects. From an audience-centric perspective, these results demonstrate that financial market responses to attention shocks are not automatic but contingent. Investors in weakly connected firms rely more heavily on external, demographically filtered signals, especially when influencer engagement comes from audiences aligned with investor characteristics.

4.2.2 Spread, Social Connectedness and Audience Group

Table 8 extends the test on spreads by including the interaction of audience groups and social connectedness group. The table stratifies results by firm-level social connectedness with investors (SCI) and audience demographics (Age, Income, and Education level).

[Table 8 about here]

For firms with low social connectedness, social media engagement is associated with significantly narrower bid–ask spreads, indicating an improvement in market liquidity rather than an increase in trading frictions. This effect is present across age groups and is stronger among middle-aged audiences than younger audiences, suggesting that engagement from more economically active demographic groups contributes more meaningfully to information incorporation. The magnitude of this liquidity-enhancing effect is largest among high-income audiences, consistent with the view that engagement originating from groups with greater investment capacity is more informative to the market.

A similar pattern emerges across educational groups. Engagement from audiences with a college degree is associated with a larger reduction in bid–ask spreads compared to engagement from audiences without a degree. This evidence supports the interpretation that who engages with the information—rather than engagement volume alone—determines how effectively attention is translated into market liquidity. Demographically aligned audiences with greater proximity to investors appear to facilitate faster information diffusion, thereby reducing informational frictions.

In contrast, sentiment effects operate in the opposite direction. For firms with low social connectedness, positive sentiment is associated with wider bid–ask spreads across all demographic groups, indicating heightened disagreement or increased trading activity driven by sentiment-based attention rather than fundamental information. The consistency of this positive sentiment–spread relationship across age, income, and education groups suggests that sentiment primarily captures non-fundamental attention shocks, which increase short-term uncertainty and trading costs.

These effects are substantially muted or absent for firms with high social connectedness. In these firms, neither engagement nor sentiment exhibits a robust association with bid–ask spreads across demographic groups. This pattern is consistent with the view that well-connected firms benefit from stronger direct information channels and a more informed investor base, reducing their sensitivity to indirect or non-fundamental signals originating from social media.

Overall, the results indicate that the liquidity impact of social media activity depends jointly on audience composition and firm-level social proximity to investors. Engagement improves liquidity most strongly when firms are less socially connected and when attention originates from investor-relevant demographic groups. By contrast, sentiment-driven attention tends to increase trading frictions, particularly in environments where traditional information channels are weaker. Together, these findings highlight that attention shocks affect markets not only through their intensity but through the identity of the audience transmitting them, shaping how information is processed and priced.

4.2.3 Abnormal Returns and Short-Term Reversals

Next, we examine whether the increase in investor attention leads to abnormal price movements, and if so, whether these movements persist or reverse. Table 9 investigates how non-financial social media posts affect firms’ abnormal stock returns, with a particular emphasis on how the demographic composition of the engaged audience mediates investor response. The analysis is segmented by age, income, and education of the audience and differentiated by firms’ social connectedness with investors (SCI).

[Table 9 about here]

For firms with low social connectedness, post sentiment plays a significant role in generating abnormal returns. Remarkably, for the high-income audience group, engagement is also highly significant, and sentiment maintains its positive and significant effect. This suggests that market participants interpret influencer activity engaging economically strong audiences as having the highest signal value, potentially due to perceived credibility or alignment with investor profiles.

Among high-SCI firms, neither engagement nor sentiment generates significant return effects across age, income and education groups, indicating a muted investor reaction due to stronger internal information channels or diminished reliance on external narratives.

These findings support the idea of cross-domain attention spillovers: when non-financial social media content gains traction among certain audience demographics, it triggers investor responses in financial markets, as reflected in abnormal return movements. The effects are strongest in firms with lower social connectedness, indicating that investors rely more on public sentiment when firm-level relational trust or transparency is weak.

[Table 10 about here]

Table 10 investigates whether the initial return effects are persistent or prone to short-term reversal. Using cumulative abnormal returns (CAR), adjusted by the Fama-French-Carhart four-factor model, over 3-day, 1-week, and 2-week windows, the analysis focuses on the firms with the lowest level of social connectedness, where prior evidence suggested heightened sensitivity to non-financial sentiment shocks.

The results indicate a clear pattern of short-term reversal. After three days, no statistically significant abnormal return is observed, and there are no further significant changes in CAR over

the two-week horizon. The absence of sustained significance in sentiment effects across all time windows reinforces the view that investor reactions to non-financial sentiment are largely transient. While sentiment may trigger an initial revaluation, as demonstrated in Table 10 these effects do not persist in subsequent trading days, suggesting a correction or fading of initial belief revision. The transience of these return effects aligns with theories of limited attention and narrative-driven behavior. However, our contribution is to show that this behavior is filtered through the composition of the audience, amplifying the immediate effect but also contributing to its short-lived nature as market participants reassess the informational quality of the source.

In summary our results show that non-financial influencer mentions cause measurable shifts in both investor attention and market outcomes. Attention spikes translate into abnormal returns and heightened volatility, especially for opaque or under-followed firms. These effects are amplified when audiences are economically aligned with potential investors and when influencer sentiment is strongly positive.

5. Discussion, Conclusions and Implications

The results of this paper document a novel causal chain from informal digital visibility to investor attention and financial market outcomes. We show that unsolicited mentions of public companies by non-financial influencers (individuals outside the traditional domain of finance) can generate measurable spikes in investor attention, followed by immediate market responses. These include narrower bid–ask spreads and short-term abnormal returns, reflecting lower trading frictions and price pressure following attention shocks. Crucially, these effects occur in the absence of fundamental news or formal disclosure, challenging prevailing assumptions about what constitutes “financially relevant” information and reinforcing the view that salience and social context are core drivers of investor behavior (Drake et al., 2012; Hirshleifer and Teoh, 2003).

More importantly, our results show that not all influencer mentions produce the same financial consequences. Rather, the market impact of these informal signals depends critically on the demographic and social connectivity of the influencer's audience. Mentions by influencers with economically engaged audiences (older, higher-income, well-educated followers embedded in urban investor hubs) generate more pronounced attention effects and more significant trading responses. This finding underscores a key point: attention does not travel uniformly. It is mediated through structured social networks and amplified when recipient groups are more likely to act on the information.

This aligns with recent evidence that investor responses depend not just on the information content but on attention allocation and audience receptivity (e.g., Hirshleifer et al., 2009; Hong et al., 2005). In particular, the role of audience connectivity extends this literature by treating social media followers as economically differentiated information intermediaries, agents whose characteristics influence whether attention turns into capital flow. More connected and demographically aligned audiences, in effect, serve as nodes in investor networks, increasing the likelihood of peer discussion, mimicry, and information cascades.

On the firm side, the effects of influencer-driven attention are most salient for companies that are peripheral to formal financial communication channels—those with weak institutional ownership. These are firms that operate in high-information-asymmetry environments. For these companies, incidental social media visibility can provide an important, if noisy, substitute for traditional investor outreach. This mechanism resonates with Merton's (1987) investor recognition hypothesis: when a stock is brought to the attention of new investors, it experiences increased demand and improved liquidity. Our results show that non-financial influencers act as unexpected

vehicles of investor recognition, broadening the visibility of firms that might otherwise remain unnoticed by the broader investment community.

These insights have several implications. First, for individual investors, the findings offer a cautionary note. The digital age has created an environment in which culturally salient but informationally diffuse signals can drive asset prices. A casual mention on Instagram or TikTok may lead to heightened volatility or short-term overvaluation, particularly for thinly traded stocks. Investors relying on social cues should exercise scepticism and complement narrative-driven insights with fundamental analysis. Yet at the same time, the very fact that social narratives can move prices implies they are economically relevant, if only temporarily, and thus worthy of monitoring.

Second, for firms, especially small and mid-cap companies, influencer-driven visibility may present an underutilized strategic lever. Mentions by non-financial influencers can enhance firm recognition, attract new retail investors, and improve liquidity. However, these effects are largely exogenous and unpredictable. Firms cannot fully control the messaging or audience reception, and ill-timed or negative mentions may introduce reputational risk or volatility. As such, firms should treat social visibility as part of a broader investor relations strategy, one that includes real-time monitoring, contingency messaging, and an understanding of how informal discourse intersects with capital markets.

Third, our results raise policy considerations. Influencers with financially active audiences are increasingly functioning as informal information gatekeepers, yet they remain largely outside the scope of financial disclosure regulation. As the SEC and international regulators (e.g., ESMA, FCA) grapple with the proliferation of social trading and digital financial content, our findings suggest that attention should be paid not only to “finfluencers” but to a broader class of cultural

influencers whose content spills over into market behavior. Clear guidelines on promotional disclosures, the scope of permissible financial commentary, and real-time surveillance mechanisms may become increasingly necessary.

Ultimately, this paper contributes to a broader re-evaluation of the investor information environment. Traditional models assume that material information is delivered through formal, regulated channels; in contrast, we show that informal, decentralized signals, mediated through culturally embedded and demographically structured networks, can yield comparable market effects. The rise of influencer-driven attention shocks represents a new class of exogenous information events, demanding integration into our models of investor behavior, price formation, and information economics.

We conclude by suggesting that attention, like capital, is subject to allocation frictions. In a media ecosystem that is not just data-rich but narrative-saturated, financial actors must compete not only for informational precision but for cognitive and social salience. Markets are increasingly shaped not only by what is known but by what is noticed, and by whom.

Bibliography

- Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).
- Bailey, M., Cao, R., Kuchler, T., Stroebe, J. and Wong, A., 2018a. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), pp.259-280.
- Bailey, M., Cao, R., Kuchler, T. and Stroebe, J., 2018b. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6), pp.2224-2276.
- Bailey, M., Dávila, E., Kuchler, T. and Stroebe, J., 2019. House price beliefs and mortgage leverage choice. *The Review of Economic Studies*, 86(6), pp.2403-2452.
- Bailey, M., Johnston, D.M., Koenen, M., Kuchler, T., Russel, D. and Stroebe, J., 2020. Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing During the COVID-19 Pandemic (No. w28234). National Bureau of Economic Research.
- Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R. and Stroebe, J., 2021. International trade and social connectedness. *Journal of International Economics*, 129, p.103418.
- Bailey, M., Johnston, D., Kuchler, T., Stroebe, J. and Wong, A., 2022. Peer effects in product adoption. *American Economic Journal: Applied Economics*, 14(3), pp.488-526.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2), pp.785-818.
- Barber, B.M., Huang, X., Odean, T. and Schwarz, C., 2022. Attention-induced trading and returns: Evidence from Robinhood users. *The Journal of Finance*, 77(6), pp.3141-3190.
- Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5), 992–1026.

- Blankespoor, E., Miller, G.S., White, H. D., 2014. The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter. *Accounting Review*, 89(1), 79–112.
- Bloomfield, R., O'Hara, M., Saar, G., 2009. How Noise Trading Affects Markets: An Experimental Analysis. *Review of Financial Studies* 22 (6), 2275–2302.
- Bushee, B.J. and Leuz, C., 2005. Economic consequences of SEC disclosure regulation: evidence from the OTC bulletin board. *Journal of Accounting and Economics*, 39(2), pp.233-264.
- Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5), 1367–1403.
- Clark-Joseph, A.D., Ye, M. and Zi, C., 2017. Designated market makers still matter: Evidence from two natural experiments. *Journal of Financial Economics*, 126(3), pp.652-667.
- Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6), 2045-2073.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461–1499.
- Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. *The Journal of Finance*, 46(4), 1325–1359.
- Engelberg, J., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67–97.
- Focke, F., Ruenzi, S. and Ungeheuer, M., 2020. Advertising, attention, and financial markets. *The Review of Financial Studies*, 33(10), pp.4676-4720.
- Giglio, S., Kuchler, T., Stroebel, J. and Zeng, X., 2023. *Biodiversity risk* (No. w31137). National bureau of economic research.
- Glosten, R., Milgrom, P.R., 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14(1), 71–100.

- Hirshleifer, D., Peng, L. and Wang, Q., 2025. News diffusion in social networks and stock market reactions. *The Review of Financial Studies*, 38(3), pp.883-937.
- Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1-3), 337–386.
- Hong, H., Kubik, J.D. and Stein, J.C., 2005. Thy Neighbor's Portfolio: Word-of-mouth Effects in the Holdings and Trades of Money Managers. *The Journal of Finance*, 60(6), pp.2801–2824.
- Hu, Z., 2022. Social interactions and households' flood insurance decisions. *Journal of Financial Economics*, 144(2), pp.414-432.
- Hutto, C. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).
- Ivković, Z., and Weisbenner, S., 2007. Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices. *The Review of Financial Studies*, 20(4), 1327–1357
- Knutson, B., Hsu, T.W., Ko, M. and Tsai, J.L., 2024. News source bias and sentiment on social media. *PloS one*, 19(10), p.e0305148.
- Korinek, A., 2023. Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4), pp.1281-1317.
- Kuchler, T., Li, Y., Peng, L., Stroebel, J. and Zhou, D., 2022. Social proximity to capital: Implications for investors and firms. *The Review of Financial Studies*, 35(6), pp.2743-2789.
- Lang, M., & Lundholm, R. (1996). Corporate disclosure policy and analyst behavior. *The Accounting Review*, 71(4), 467–492.
- Lim, S.L. and Bentley, P.J., 2022. Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, 12(1), p.18131.

- Liu, H., Peng, L. and Tang, Y., 2023. Retail attention, institutional attention. *Journal of Financial and Quantitative Analysis*, 58(3), pp.1005-1038.
- Lopez-Lira, A. and Tang, Y., 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*.
- Madsen, J. and Niessner, M., 2019. Is investor attention for sale? The role of advertising in financial markets. *Journal of Accounting Research*, 57(3), pp.763-795.
- Rehbein, O. and Rother, S., 2025. Social connectedness in bank lending. *The Review of Financial Studies*, p.hhaf014.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.
- Thorgren, E., Mohammadinodooshan, A. and Carlsson, N., 2024, May. Temporal dynamics of user engagement on Instagram: A comparative analysis of album, photo, and video interactions. In: *Proceedings of the 16th ACM Web Science Conference* (pp. 224-234).

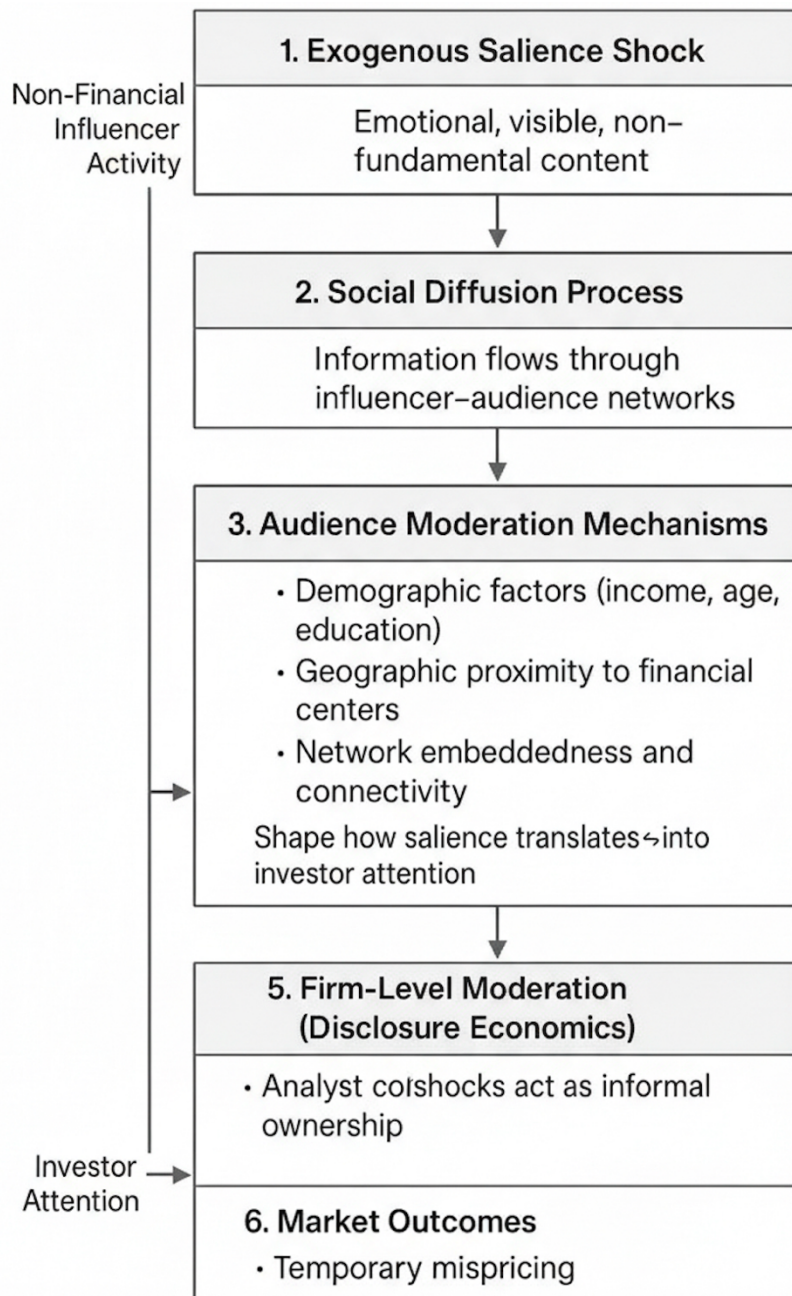


Figure 1. Conceptual Model of Influencer-Driven Investor Attention

This figure illustrates how non-financial influencer activity affects capital markets through a multi-stage attention mechanism. Influencer communications generate **exogenous salience shocks**, which diffuse through social networks and reach heterogeneous audiences. The **composition and structure of these audiences**—defined by demographic, geographic, and network characteristics—moderate the transmission of attention into investor behavior. Investor attention manifests in heightened search activity, trading, and short-term price movements. At the firm level, the effect is conditioned by **informational opacity** and **baseline visibility**: attention shocks function as informal signals where formal disclosures are limited. Together, these mechanisms link non-fundamental visibility to measurable market outcomes.

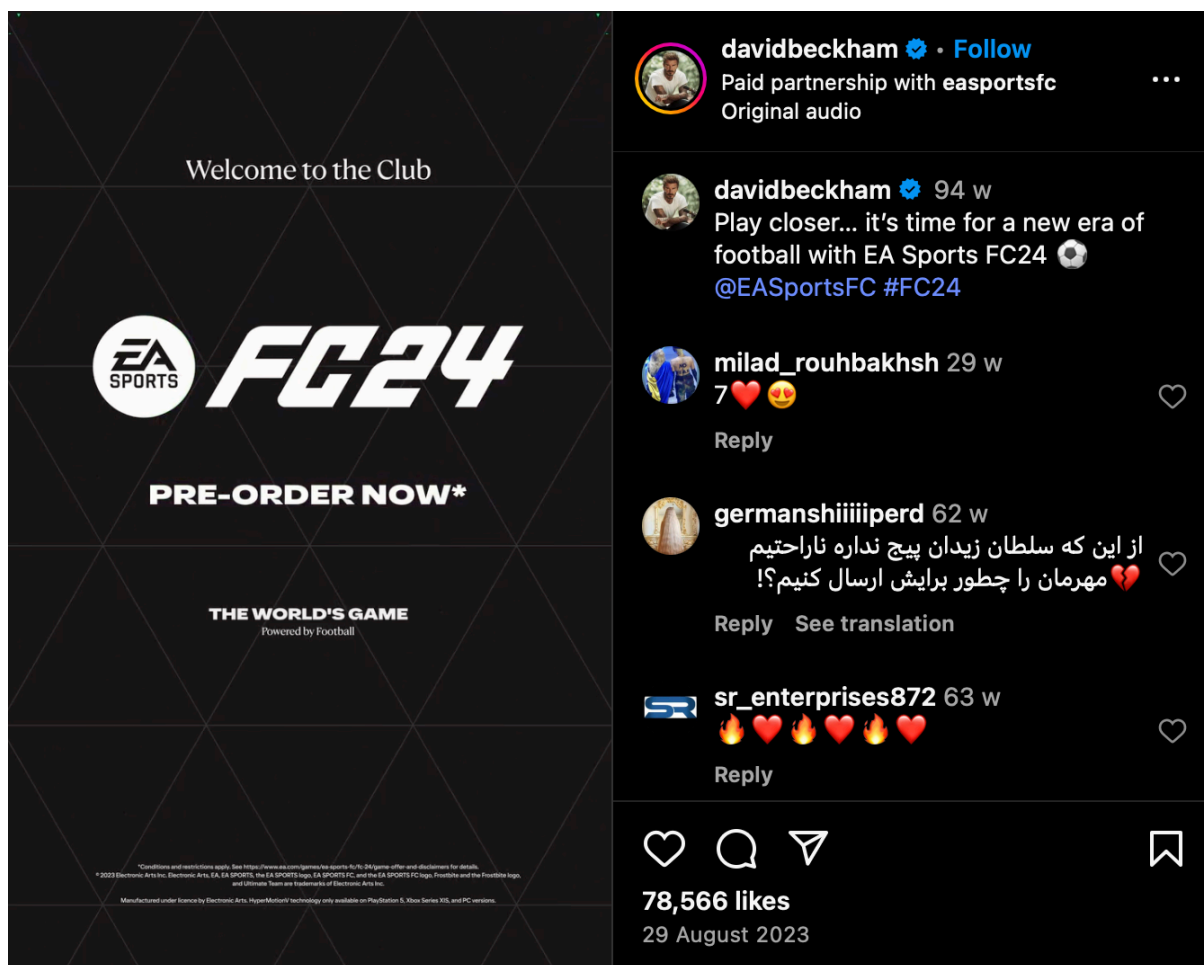


Figure 2: Example of Instagram Posting

This figure shows a sample post identified from David Beckham (Instagram account: @davidbeckham) dated August 29, 2023. The screenshot was taken on June 24, 2025. The post is accessible at <https://www.instagram.com/reel/Cwh2MSJoaNK/>. It was identified using the hashtag #FC24, referring to a game by EA Sports, and linked to the firm **Electronic Arts** (ticker: EA).

Table 1: Summary Statistics

This table presents summary statistics for key variables from January 2015 to December 2024. Panel A reports Instagram post characteristics at the firm-day level, specifically engagement ($Eng_{i,t}$) and sentiment ($AbsSent_{i,t}$) measures. Panel B displays the distribution of audience characteristics, including geography, age, education, and income, for Instagram audiences. Panel C provides statistics for Google Volume Index (SVI) data, measuring investor attention across the same four states. Panel D summarizes firm-level financial variables: market capitalization ($\ln(MV)$), turnover ($Turn$), realized volatility ($RealVol$), returns (Ret), and earnings announcement indicators (EA). Standard deviations, quartiles (p25, p50, p75), and observation counts (N) are reported for each variable.

	N	Mean	Std. Dev.	p25	Median	p75
Panel A: Instagram Post Data						
$Eng_{i,t}$	20,534	0.802	8.426	0.047	0.136	0.398
$AbsSent_{i,t}$	20,336	0.681	0.221	0.500	0.700	0.800
Panel B: Instagram Audience Data						
CA Audience (%)	13,942	5.662	2.543	3.730	5.580	8.170
NY Audience (%)	13,819	4.333	1.867	2.840	3.953	6.240
FL Audience (%)	13,844	3.456	1.469	2.310	3.610	4.960
TX Audience (%)	13,749	3.130	1.255	2.095	3.250	4.370
Age: Middle Age (%)	20,024	21.905	5.479	17.110	21.560	25.220
Age: Young (%)	20,024	73.872	7.397	69.260	74.070	79.890
Edu: No Degree (%)	20,024	71.786	7.620	66.590	68.590	77.370
Edu: Degree (%)	20,024	28.187	7.667	22.610	31.410	33.410
Income: High (%)	20,024	9.848	5.854	3.820	12.040	13.670
Income: Middle (%)	20,024	30.385	14.043	16.120	37.330	42.080
Income: Low (%)	20,024	59.766	19.603	44.250	49.250	80.190
Panel C: Google Volume Index (SVI) Data						
CA SVI	496,817	28.978	31.808	0.000	17.000	58.000
NY SVI	496,953	24.864	30.796	0.000	0.000	53.000
FL SVI	496,953	24.976	31.719	0.000	0.000	55.000
TX SVI	496,953	26.352	32.018	0.000	0.000	56.000
Panel D: Firm Data						
$\ln(MV_{i,t-7})$	498,299	17.594	1.339	16.631	17.506	18.572
$Turn_{i,t-7;t-36}$ (%)	499,715	4.158	12.396	0.759	1.723	3.770
$RealVol_{i,t-7;t-36}$	499,537	0.014	0.008	0.009	0.012	0.016
$Ret_{i,t-7;t-36}$ (%)	499,393	1.019	0.109	0.959	1.017	1.075
$EA_{i,t}$	500,449	0.012	0.107	0.000	0.000	0.000

Table 2: Correlation Matrix

This table reports the pairwise correlations.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) $Eng_{i,t}$	1.000						
(2) $AbsSent_{i,t}$	0.080	1.000					
(3) $\ln(MV_{i,t-7})$	0.017	0.088	1.000				
(4) $Turn_{i,t-7;t-36}$	-0.003	-0.021	-0.206	1.000			
(5) $RealVol_{i,t-7;t-36}$	0.000	0.007	-0.170	0.402	1.000		
(6) $Ret_{i,t-7;t-36}$	0.000	0.008	0.046	-0.005	-0.078	1.000	
(7) $EA_{i,t}$	0.001	0.002	0.001	-0.004	-0.009	-0.002	1.000

Table 3: Social Media Posts and Investor Attention: Main Results

This table reports how investor attention relates to audience engagement with social media posts about firms. The dependent variable ($\text{FourStateSVI}_{i,t}$) is the average Google search volume for firm i on day t across the four states in the sample (CA, TX, NY, and FL). Engagement ($\text{Eng}_{i,t}$) is the number of comments on all posts about firm i uploaded on day t , and absolute sentiment ($\text{AbsSent}_{i,t}$) is the absolute value of the average sentiment score of these posts. $\text{FourStateEng}_{i,t}$ is the number of comments on all posts about firm i uploaded on day t by influencers whose audience is predominantly located in the four states, and $\text{FourStateAbsSent}_{i,t}$ is the absolute value of the average sentiment score of these posts. $\ln(\text{MV})_{i,t-7}$ is the natural log of firm i 's market capitalization, lagged by seven days. $\text{Turn}_{i,t-7:t-36}$ is the trading turnover over the prior four weeks; $\text{RealVol}_{i,t-7:t-36}$ is realized volatility, computed as the mean absolute return over the same period. $\text{Ret}_{i,t-7:t-36}$ is the cumulative return over the prior four weeks. $\text{EA}_{i,t}$ a dummy for whether firm i announced earnings on day t . Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	$\text{FourStateSVI}_{i,t}$		$\text{FourStateSVI}_{i,t+1}$		$\text{FourStateSVI}_{i,t+2}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Eng}_{i,t}$	0.011**		0.008**		0.019**	
	(0.006)		(0.004)		(0.008)	
$\text{AbsSent}_{i,t}$	0.461**		0.197		-0.526	
	(0.209)		(0.280)		(0.378)	
$\text{FourStateEng}_{i,t}$		0.096***		0.120***		0.319***
		(0.029)		(0.029)		(0.034)
$\text{FourStateAbsSent}_{i,t}$		0.400*		0.088		-0.475
		(0.240)		(0.274)		(0.346)
$\text{FourStateSVI}_{i,t-1}$	0.386***	0.386***	0.156***	0.156***	0.147***	0.147***
	(0.014)	(0.014)	(0.016)	(0.016)	(0.015)	(0.015)
$\text{FourStateSVI}_{i,t-2}$	0.032**	0.032**	0.128***	0.128***	0.170***	0.170***
	(0.015)	(0.015)	(0.013)	(0.013)	(0.014)	(0.014)
$\text{FourStateSVI}_{i,t-7}$	0.367***	0.367***	0.324***	0.324***	0.181***	0.181***
	(0.018)	(0.018)	(0.016)	(0.016)	(0.012)	(0.012)
$\ln(\text{MV}_{i,t-7})$	0.023	0.023	0.035	0.035	0.042	0.042
	(0.034)	(0.034)	(0.065)	(0.065)	(0.087)	(0.087)
$\text{Turn}_{i,t-7:t-36}$	0.014***	0.014***	0.027***	0.027***	0.035***	0.035***
	(0.005)	(0.005)	(0.009)	(0.009)	(0.011)	(0.011)
$\text{RealVol}_{i,t-7:t-36}$	-1.543	-1.574	0.376	0.351	2.623	2.647
	(5.157)	(5.162)	(9.429)	(9.433)	(12.187)	(12.183)
$\text{Ret}_{i,t-7:t-36}$	-0.281**	-0.280**	-0.408*	-0.407*	-0.435*	-0.436*
	(0.128)	(0.128)	(0.210)	(0.210)	(0.260)	(0.260)
$\text{EA}_{i,t}$	13.176***	13.177***	9.977***	9.978***	2.083***	2.083***
	(1.409)	(1.409)	(1.355)	(1.355)	(0.422)	(0.422)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y
N	495,890	495,890	495,745	495,745	495,600	495,600
R-sq	0.915	0.915	0.884	0.884	0.869	0.869
t-test	Chi-sq = 11.40		Eng = FourStateEng		Chi-sq = 99.86	
	P-value = 0.001		P-value = 0.000		P-value = 0.000	

Table 4: Social Media Posts and Investor Attention: Audience Geographic Allocation

This table reports how investor attention relates to audience engagement with social media posts about firms. The dependent variable is the Google search volume for firm i on day t in each of the four states in the sample (CA $SVI_{i,t}$, TX $SVI_{i,t}$, NY $SVI_{i,t}$, and FL $SVI_{i,t}$). Engagement ($Eng_{i,t}$) is the number of comments on all posts about firm i uploaded on day t , and absolute sentiment ($AbsSent_{i,t}$) is the absolute value of the average sentiment score of these posts. XX State $Eng_{i,t}$ is the number of comments on all posts about firm i uploaded on day t by influencers whose audience is predominantly located in state XX, and XX State $AbsSent_{i,t}$ is the absolute value of the average sentiment score of these posts. $DepVar_{i,t-n}$ is the n -period lag of the dependent variable. $Ln(MV)_{i,t-7}$ is the natural log of firm i 's market capitalization, lagged by seven days. $Turn_{i,t-7:t-36}$ is the trading turnover over the prior four weeks; $RealVol_{i,t-7:t-36}$ is realized volatility, computed as the mean absolute return over the same period. $Ret_{i,t-7:t-36}$ is the cumulative return over the prior four weeks. $EA_{i,t}$ a dummy for whether firm i announced earnings on day t . Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	CA $SVI_{i,t}$		TX $SVI_{i,t}$		NY $SVI_{i,t}$		FL $SVI_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Eng_{i,t}$	0.012* (0.007)		0.019*** (0.005)		0.015** (0.007)		0.011* (0.006)	
$AbsSent_{i,t}$	0.519* (0.293)		0.507* (0.292)		0.482* (0.286)		0.747*** (0.284)	
CA State$Eng_{i,t}$		0.412*** (0.147)						
CA State $AbsSent_{i,t}$		0.450 (0.352)						
TX State$Eng_{i,t}$				0.694*** (0.070)				
TX State $AbsSent_{i,t}$				0.313 (0.328)				
NY State$Eng_{i,t}$						0.650** (0.303)		
NY State $AbsSent_{i,t}$						0.419 (0.356)		
FL State$Eng_{i,t}$								0.284** (0.135)
FL State $AbsSent_{i,t}$								0.716** (0.331)
$DepVar_{i,t-1}$	0.298*** (0.014)	0.298*** (0.014)	0.268*** (0.013)	0.268*** (0.013)	0.257*** (0.014)	0.257*** (0.014)	0.246*** (0.012)	0.246*** (0.012)
$DepVar_{i,t-2}$	0.087*** (0.011)	0.087*** (0.011)	0.098*** (0.013)	0.098*** (0.013)	0.097*** (0.011)	0.097*** (0.011)	0.108*** (0.012)	0.108*** (0.012)
$DepVar_{i,t-7}$	0.276*** (0.015)	0.276*** (0.015)	0.281*** (0.018)	0.281*** (0.018)	0.259*** (0.017)	0.259*** (0.017)	0.267*** (0.016)	0.267*** (0.016)
$ln(MV_{i,t-7})$	0.048 (0.050)	0.048 (0.050)	0.097 (0.085)	0.097 (0.085)	0.065 (0.077)	0.065 (0.077)	-0.040 (0.056)	-0.039 (0.056)
$Turn_{i,t-7:t-36}$	0.023***	0.023***	0.026***	0.026***	0.028***	0.028***	0.025**	0.025**

	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)
RealVol _{<i>i,t-7:t-36</i>}	2.325	2.291	1.745	1.692	1.826	1.791	14.938	14.891
	(8.983)	(8.987)	(9.246)	(9.254)	(9.977)	(9.981)	(10.676)	(10.682)
Ret _{<i>i,t-7:t-36</i>}	-0.284	-0.284	-0.274	-0.273	-0.338	-0.338	0.123	0.123
	(0.219)	(0.219)	(0.240)	(0.240)	(0.213)	(0.213)	(0.209)	(0.209)
EA _{<i>i,t</i>}	16.546***	16.547***	10.563***	10.564***	16.277***	16.278***	9.506***	9.507***
	(1.753)	(1.753)	(1.352)	(1.352)	(1.648)	(1.648)	(1.169)	(1.169)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y	Y	Y
N	495,747	495,747	495,890	495,890	495,890	495,890	495,890	495,890
R-sq	0.815	0.815	0.816	0.816	0.804	0.804	0.803	0.803
t-tests	Eng = CA Eng		Eng = TX Eng		Eng = NY Eng		Eng = FL Eng	
	Chi-sq = 7.72		Chi-sq = 90.78		Chi-sq = 7.72		Chi-sq = 90.78	
	P-value = 0.006		P-value = 0.000		P-value = 0.006		P-value = 0.000	

Table 5: Social Media Posts and Investor Attention: Audience Demographics

This table reports how investor attention relates to audience engagement with social media posts about firms, conditional on audience demographic composition. The table presents the results from three separate regressions, one for each demographic classification: age, income, and education. The dependent variable ($\text{FourStateSVI}_{i,t}$) is the average Google search volume for firm i on day t across the four states in the sample (CA, TX, NY, and FL). $\text{FourStateEng}_{i,t}^{XX}$ is the number of comments on all posts about firm i uploaded on day t by influencers whose audience is predominantly located in the four states and belongs to demographic group XX , and $\text{FourStateAbsSent}_{i,t}^{XX}$ (not reported) is the absolute value of the average sentiment score of these posts. Lags of the dependent variable and firm-level control variables are the same as in Tables 3 and 4. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	$\text{FourStateSVI}_{i,t}$	$\text{FourStateSVI}_{i,t}$	$\text{FourStateSVI}_{i,t}$
$\text{FourStateEng}_{i,t}^{\text{Young}}$	0.134*** (0.039)		
$\text{FourStateEng}_{i,t}^{\text{Middle Age}}$	0.376*** (0.113)		
$\text{FourStateEng}_{i,t}^{\text{Low income}}$		0.205*** (0.067)	
$\text{FourStateEng}_{i,t}^{\text{Middle income}}$		0.290*** (0.079)	
$\text{FourStateEng}_{i,t}^{\text{High income}}$		0.798*** (0.229)	
$\text{FourStateEng}_{i,t}^{\text{No Degree}}$			0.135*** (0.041)
$\text{FourStateEng}_{i,t}^{\text{Degree}}$			0.336*** (0.096)
Sentiment Control	Y	Y	Y
Lags of FourStateSVI	Y	Y	Y
Firm Controls	Y	Y	Y
Firm FE	Y	Y	Y
Year FE	Y	Y	Y
Week FE	Y	Y	Y
Weekday FE	Y	Y	Y
N	495,890	495,890	495,890
R-sq	0.915	0.915	0.915
t-test	$\text{FourStateEng}_{i,t}^{\text{Young}}$ = $\text{FourStateEng}_{i,t}^{\text{Middle Age}}$ Chi-sq = 10.14 P-value = 0.002	$\text{FourStateEng}_{i,t}^{\text{Low income}}$ = $\text{FourStateEng}_{i,t}^{\text{High income}}$ Chi-sq = 10.86 P-value = 0.001	$\text{FourStateEng}_{i,t}^{\text{No Degree}}$ = $\text{FourStateEng}_{i,t}^{\text{Degree}}$ Chi-sq = 12.57 P-value = 0.000

Table 6: Social Media Posts and Investor Attention – Firm Proximity to Capital

This table reports how investor attention relates to audience engagement with social media posts about firms, conditional on firm-level proximity to institutional capital. We use three alternative measures to assess a firm's proximity to institutional capital: Social Proximity of Capital (SPC), Physical Proximity of Capital (PPC), and the Social Connectedness Index (SCI) to the Tri-State financial hub. For each measure, we assign a firm the value of the county in which its headquarters is located. We then divide firms into high- and low-proximity groups based on the median of the latest available year of data and estimate the regression separately for each group. The dependent variable ($\text{FourStateSVI}_{i,t}$) is the average Google search volume for firm i on day t across the four states in the sample (CA, TX, NY, and FL). $\text{FourStateEng}_{i,t}$ is the number of comments on all posts about firm i uploaded on day t by influencers whose audience is predominantly located in the four states, and $\text{FourStateAbsSent}_{i,t}$ is the absolute value of the average sentiment score of these posts. $\ln(\text{MV})_{i,t-7}$ is the natural log of firm i 's market capitalization, lagged by seven days. $\text{Turn}_{i,t-7:t-36}$ is the trading turnover over the prior four weeks; $\text{RealVol}_{i,t-7:t-36}$ is realized volatility, computed as the mean absolute return over the same period. $\text{Ret}_{i,t-7:t-36}$ is the cumulative return over the prior four weeks. $\text{EA}_{i,t}$ a dummy for whether firm i announced earnings on day t . Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: $\text{FourStateSVI}_{i,t}$						
	SPC Groups		PPC Groups		SCI to Tri-state Groups	
	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{FourStateEng}_{i,t}$	-0.154	0.081***	0.001	0.081***	0.866	0.068***
	(0.345)	(0.030)	(0.345)	(0.027)	(0.540)	(0.023)
$\text{FourStateAbsSent}_{i,t}$	0.386	0.548	0.383	0.539	-0.009	0.649*
	(0.518)	(0.397)	(0.435)	(0.458)	(0.557)	(0.388)
$\text{FourStateSVI}_{i,t-1}$	0.346***	0.422***	0.359***	0.405***	0.346***	0.418***
	(0.018)	(0.023)	(0.019)	(0.024)	(0.019)	(0.022)
$\text{FourStateSVI}_{i,t-2}$	0.063***	-0.031	0.050**	-0.016	0.055**	-0.020
	(0.021)	(0.023)	(0.021)	(0.022)	(0.021)	(0.023)
$\text{FourStateSVI}_{i,t-7}$	0.360***	0.364***	0.366***	0.360***	0.373***	0.354***
	(0.022)	(0.033)	(0.022)	(0.032)	(0.023)	(0.032)
$\ln(\text{MV})_{i,t-7}$	0.341	0.000	0.226	0.008	0.011	-0.013
	(0.212)	(0.017)	(0.197)	(0.021)	(0.021)	(0.201)
$\text{Turn}_{i,t-7:t-36}$	0.033***	0.000	0.033***	0.000	-0.029	0.008*
	(0.009)	(0.006)	(0.009)	(0.007)	(0.018)	(0.005)
$\text{RealVol}_{i,t-7:t-36}$	-5.408	-9.717	-10.477	-6.403	2.239	-5.243
	(8.667)	(6.164)	(8.400)	(6.252)	(7.486)	(8.151)
$\text{Ret}_{i,t-7:t-36}$	-0.535**	-0.146	-0.424*	-0.204	-0.181	-0.354
	(0.228)	(0.182)	(0.246)	(0.163)	(0.139)	(0.236)
$\text{EA}_{i,t}$	11.455***	15.912***	12.142***	15.295***	10.057***	17.513***
	(1.661)	(2.653)	(1.887)	(2.493)	(1.583)	(2.666)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y
N	204,544	213,102	209,987	207,659	212,951	204,695
R-sq	0.903	0.931	0.910	0.925	0.918	0.918

Table 7: Financial Reactions and Firms' Social Connectedness with Investors

This table reports the regression results of financial market reactions—measured by daily closing spread, turnover, and abnormal return—to non-financial social media posts from January 2015 to December 2024. Firms are grouped based on the decile distribution of their Social Connectedness Index (SCI) to the Tri-State area, where Low refers to firms in the bottom decile (least connected) and High refers to those in the top decile (most connected). Engagement is the number of social media engagement in thousands, and Sentiment is the GPT-based sentiment score. Abnormal return is adjusted using the Fama-French-Carhart four-factor model. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	Spread _{<i>i,t</i>}			Turnover _{<i>i,t</i>}			Abnormal Ret _{<i>i,t</i>}		
SCI Group	(1) All	(2) Low	(3) High	(4) All	(5) Low	(6) High	(7) All	(8) Low	(9) High
Eng _{<i>i,t</i>}	-0.000 (0.000)			0.000 (0.000)			0.001 (0.001)		
Sent _{<i>i,t</i>}	-0.007 (0.008)			-0.003 (0.005)			0.059** (0.030)		
Eng _{<i>i,t</i>}		-0.001*** (0.000)	0.000 (0.000)		0.003 (0.004)	0.000 (0.000)		0.008 (0.008)	0.020 (0.014)
Sent _{<i>i,t</i>}		0.002** (0.001)	-0.001 (0.001)		-0.009 (0.010)	-0.007 (0.010)		0.084** (0.025)	0.005 (0.102)
Lagged Dependent _{<i>i,t-7</i>}	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	288,415	17,654	23,973	288,415	17,654	23,973	286,953	17,521	23,973
R-sq	0.341	0.686	0.488	0.657	0.819	0.722	0.002	0.009	0.005

Table 8: Spread, Social Connectedness and Audience Groups

This table presents regression results on the relationship between closing bid-ask spread and non-financial social media information, with an emphasis on heterogeneity across audience demographic groups from Jan 2015 to Dec 2024. Firms are classified based on their Social Connectedness Index (SCI) to the Tri-State area, where Low and High indicate the bottom and top deciles, respectively. Audience groups are split by age (young, mid-age), income (low, middle, high), and education (no degree vs. degree holders). Engagement is the number of social media engagement in thousands, and Sentiment is the GPT-based sentiment score. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: Spread _{i,t}									
	Age	(1)	(2)	Income	(3)	(4)	Education	(5)	(6)
		Low SCI	High SCI		Low SCI	High SCI		Low SCI	High SCI
Eng _{i,t} ^g	Young	-0.001***	0.000*	Low	-0.001***	0.000*	No Degree	-0.001***	0.000*
		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
Sent _{i,t} ^g		0.002**	-0.001		0.002**	-0.001		0.002**	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Eng _{i,t} ^g		-0.004***	0.001		-0.002***	0.002		-0.003***	0.001
		(0.000)	(0.001)		(0.000)	(0.002)		(0.000)	(0.001)
Sent _{i,t} ^g	Mid Age	0.002***	-0.001	Middle	0.002**	-0.001	Degree	0.002**	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Eng _{i,t} ^g		-0.006***	0.003		-0.006***	0.003		-0.006***	0.003
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Sent _{i,t} ^g		0.002**	-0.001		0.002**	-0.001		0.002**	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Eng _{i,t} ^g	High	-0.006***	0.003	High	-0.006***	0.003		-0.006***	0.003
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Sent _{i,t} ^g		0.002**	-0.001		0.002**	-0.001		0.002**	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Eng _{i,t-7}		Y	Y		Y	Y		Y	Y
		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
Firm Controls		Y	Y		Y	Y		Y	Y
Firm FE		Y	Y		Y	Y		Y	Y
Year FE		Y	Y		Y	Y		Y	Y
Week FE		Y	Y		Y	Y		Y	Y
Weekday FE		Y	Y		Y	Y		Y	Y
N		17,654	23,973		17,654	23,973		17,654	23,973

Table 9: Abnormal Return, Social Connectedness and Audience Groups

This table presents regression results on the relationship between abnormal return and non-financial social media information, with an emphasis on heterogeneity across audience demographic groups from Jan 2015 to Dec 2024. Firms are classified based on their Social Connectedness Index (SCI) to the Tri-State area, where Low and High indicate the bottom and top deciles, respectively. Audience groups are split by age (young, mid-age), income (low, middle, high), and education (no degree vs. degree holders). Engagement is the number of social media engagement in thousands, and Sentiment is the GPT-based sentiment score. Abnormal return is adjusted using the Fama-French-Carhart four-factor model. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: Abnormal Return _{<i>i,t</i>}									
	Age	(1)	(2)	Income	(3)	(4)	Education	(5)	(6)
		Low SCI	High SCI		Low SCI	High SCI		Low SCI	High SCI
Eng _{<i>i,t</i>} ^g	Young	0.010	0.025	Low	0.014	0.018	No Degree	0.012	0.024
		(0.010)	(0.019)		(0.022)	(0.012)		(0.013)	(0.017)
Sent _{<i>i,t</i>} ^g		0.076**	-0.003		0.076**	0.004		0.075**	-0.001
		(0.026)	(0.110)		(0.026)	(0.106)		(0.026)	(0.109)
Eng _{<i>i,t</i>} ^g	Mid Age	0.056	0.112	Middle	0.023*	0.241	Degree	0.027	0.117
		(0.045)	(0.084)		(0.011)	(0.253)		(0.018)	(0.104)
Sent _{<i>i,t</i>} ^g		0.073**	-0.004		0.075**	-0.020		0.075**	-0.010
		(0.026)	(0.111)		(0.026)	(0.137)		(0.026)	(0.116)
Eng _{<i>i,t</i>} ^g				High	0.074**	0.689			
					(0.031)	(0.682)			
Sent _{<i>i,t</i>} ^g					0.073**	-0.054			
					(0.025)	(0.119)			
Abnormal Return _{<i>i,t-7</i>}		Y	Y		Y	Y		Y	Y
Firm Controls		Y	Y		Y	Y		Y	Y
Firm FE		Y	Y		Y	Y		Y	Y
Year FE		Y	Y		Y	Y		Y	Y
Week FE		Y	Y		Y	Y		Y	Y
Weekday FE		Y	Y		Y	Y		Y	Y
N		17,521	23,973		17,521	23,973		17,521	23,973

Table 10: Short-Term Reverse of Abnormal Return in Low Connected Area

This table reports regression results examining the short-term reversal of abnormal returns following non-financial social media activities, focusing on firms located in areas with low social connectedness to investors from Jan 2015 to Dec 2024. The dependent variable is the cumulative abnormal return (CAR) over 3-day, 1-week, and 2-week windows. Abnormal returns are adjusted using the Fama-French-Carhart four-factor model. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: Cumulative Abnormal Return						
	3 Days		1 Week		2 Weeks	
	(1)	(2)	(3)	(4)	(5)	(6)
$Eng_{i,t}$	-0.057 (0.033)	-0.056 (0.032)	-0.022 (0.047)	-0.020 (0.048)	-0.029 (0.056)	-0.028 (0.059)
$Sent_{i,t}$	0.117 (0.097)	0.098 (0.090)	0.177 (0.205)	0.148 (0.182)	-0.043 (0.356)	-0.101 (0.294)
Abnormal Ret $_{i,t-5}$	N	Y	N	Y	N	Y
Firm Controls	N	Y	N	Y	N	Y
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y
N	17,510	17,505	17,494	17,489	17,454	17,449
R-sq	0.009	0.013	0.011	0.016	0.016	0.026

APPENDIX

Appendix 1: ChatGPT Settings

This paper uses ChatGPT model 4o-mini¹ through the official Python package OpenAI to do the sentiment detection and assistant on the firm hashtag's detection. The prompt settings are as follows:

GPT Settings for Hashtags Identification

Prompt_system = "You are a helpful assistant specialized in social media trends."

Prompt = f"The company '{company_name}' is a well-known entity. Analyze its brand identity, products, services, and social media presence. "

f"Based on this analysis:\n"

f"1. Exclude generic or industry-wide hashtags (e.g., #Technology, #Innovation).\n"

f"2. Exclude common trends that are not directly associated with this company.\n"

f"3. Prioritize hashtags that are ****specific, unique, and strongly tied to this company's brand identity, marketing campaigns, slogans, or key product names****. "

f"These should set this company apart from its competitors and reflect its distinct characteristics.\n\n"

f"Important: Ensure you are not making up information or providing inaccurate hashtags. If you can find fewer than 5 hashtags, report only the ones you can find. "

f"If you can identify more than 5, include all relevant ones.\n\n"

f"Provide the hashtags in the following format:\n"

f"- #Hashtag1\n"

f"- #Hashtag2\n"

f"- #Hashtag3\n"

GPT Settings for Sentiment Detection

Prompt_system = "You are an assistant that provides sentiment analysis on a scale from -1 (negative) to 1 (positive)."

Prompt = f"Analyse the sentiment of this text: '{text}'. Provide a score between -1 and 1."

f"Provide the sentiment score in the following format: Sentiment Score : Score"

f"If you cannot determine the sentiment or the content is blank, provide: 'Not Applicable'"

¹ <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>

Appendix 2: VADER Sentiment Scores

We replicate our core results using VADER (Valence Aware Dictionary for sEntiment Reasoning) sentiment scores developed by Hutto and Gilbert (2014) as an alternative to GPT-based sentiment classification. Unlike the GPT sentiment testing, the VADER method provides a detailed lexicon and a rule-based approach, specifically designed for analyzing social media content. It is preferred over traditional financial dictionaries—such as the LM dictionary by Loughran and McDonald (2011) and the **HL dictionary** by Hu and Liu (2004)—due to its suitability for informal online text.

Each word in the VADER lexicon is assigned a sentiment score ranging from -4 (extremely negative) to 4 (extremely positive). The method evaluates sentiment at the sentence level and is capable of detecting slang, emojis, informal grammar, punctuation, and capitalization patterns. For instance, it can convert “😊” into “smile face” and assign it a positive sentiment. It can also distinguish between “The food here is good.” and “The food here is GOOOOOOD!!!!,” with the latter clearly expressing a stronger positive sentiment. The full list of sentiment words and testing rules can be found on VADER’s GitHub page.²

After estimating the sentiment score for each post using VADER, we construct the firm-day sentiment variables following the same approach used for processing raw GPT sentiment, as described in Section 3. Specifically, we take the absolute values of sentiment scores and adjust them for user engagement, as well as for geographic and audience group segmentation. Table OA2 in the Online Appendix confirms that the positive effect of States Engagements remains statistically significant at the 1% or 5% level, even when using VADER sentiment. This reinforces

² <https://github.com/cjhutto/vaderSentiment>

the robustness of our findings to alternate sentiment detection approaches and supports the conclusion that engagement itself—rather than content tone alone—is a key driver of investor attention.

Appendix 3: Social and Capital Proximity

To evaluate how information diffusion is moderated by social and institutional connectedness, we incorporate three proximity measures: the Social Connectedness Index (SCI), Social Proximity to Capital (SPC), and Physical Proximity to Capital (PPC).

Bailey et al. (2018a) introduced the Social Connectedness Index (SCI), developed by Meta (formerly Facebook), which uses Facebook friendship links to estimate the degree of social connectedness between US counties. Given the widespread use of Facebook, these friendship links are considered reflective of real-life social ties. The SCI is defined as the number of Facebook friendship links between individuals in two counties divided by the product of the number of Facebook users in each county.:

$$SCI_{ij} = \frac{\text{Facebook Connections}_{ij}}{\text{Facebook Users}_i \times \text{Facebook Users}_j} \quad (\text{a1})$$

Facebook Users i and Facebook Users j are the number of Facebook users in counties i and j , and Facebook Connections $_{ij}$ is the total number of Facebook friendship between individuals in county i and county j . Essentially, SCI_{ij} measures the probability that a person from county i is acquainted with a person in county j .

Since Bailey et al. (2018a) introduced the SCI, it has become a key measure of “online distance” between US counties based on Facebook friendship ties. This index has been used in various studies, such as housing markets (Bailey et al., 2018b), mortgage decisions (Bailey et al., 2019), pandemic behavior (Bailey et al., 2022), international trade (Bailey et al., 2021), product adoption (Bailey et al., 2022), and migration (Bailey et al., 2024). The core finding is that individuals’

behaviors and decisions are significantly shaped by their social networks, namely by individuals in areas that are highly connected to them. For example, institutional investors are more likely to invest in firms located in areas where they have stronger social ties (Kuchler et al., 2022), and lending barriers decline when bank and borrower regions are highly connected (Rehbein & Rother, 2025).

Beyond physical distance, SCI captures connections across cultural, political, and socioeconomic lines (Bailey et al., 2018a). This is illustrated in Figure A1, which presents a heatmap of SCI measures using New York County, NY (Panel A) and Austin County, TX (Panel B) as reference points. A higher value (darker green) indicates a stronger social connection with the target county, while a lower value (lighter green) indicates weaker ties. We observe that New York County (Manhattan) has the strongest connections with other counties in the Northeast, particularly within New York State, as well as scattered urban centres across the US. In contrast, Austin shows broader, more diffuse connections, with relatively strong ties across many parts of Texas and the South. In both cases, stronger SCI values tend to correlate with shared social, economic, and cultural characteristics. Compared to using simple physical distance, leveraging social connectedness provides a more nuanced understanding of the hidden relationships between non-financial audiences and their ties to investors.

For each county, following Hirshleifer et al. (2025), we use the Tri-State Area (New York, New Jersey, and Connecticut), which hosts the headquarters of many financial institutions, as a proxy for a region highly concentrated with financial investors. We then calculate each county's SCI with the Tri-State Area to proxy for its social connections with investors. Higher SCI values to the tri-state area indicate stronger social integration with the investor core. Then, following Kuchler et al. (2022), we set a firm's social proximity to institutional capital as equal to the value of the SCI to the Tri-State Area of the firm's headquarters location county.

The second measure we use is the Social Proximity to Capital (SPC), developed by Kuchler et al (2022). SPC captures how socially linked a county is to institutional investors:

$$SPC_{i,t} = \sum AUM_{j,t} \times SCI_{ij} \quad (a2)$$

$AUM_{j,t}$ is the total assets under management by institutions headquartered in county j in quarter t , and SCI_{ij} is the Social Connectedness Index between counties i and j . A higher SPC means that county i is more connected to institutional investors at time t . Following Kuchler et al. (2022), we set a firm's social proximity to institutional capital as equal to the SPC value of the firm's headquarters location county.

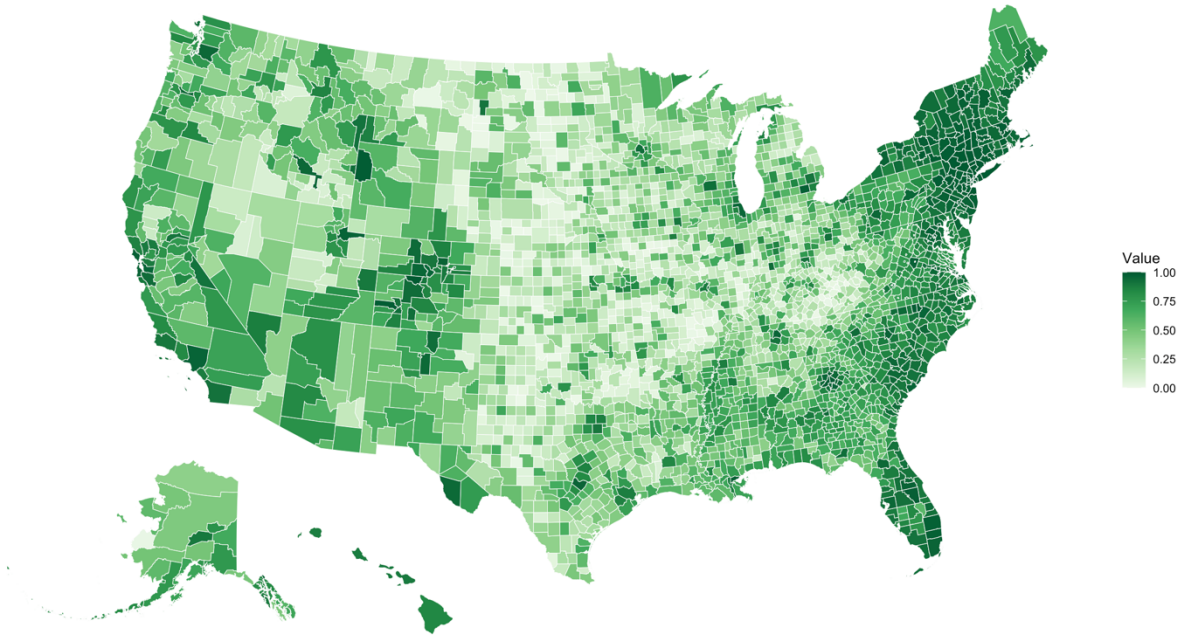
The third measure we use is Physical Proximity to Capital (PPC), also developed by Kuchler et al (2022). PPC captures the geographic closeness of county i to counties j with high AUM:

$$PPC_{i,t} = \frac{\sum AUM_{j,t}}{1 + \text{Distance}_{ij}} \quad (a3)$$

Distance_{ij} is the physical distance between counties i and j in miles, so here proximity is purely physical rather than social. A higher PPC indicates that county i is geographically closer to the locations where institutional investors operate. We set a firm's physical proximity to institutional capital as equal to the PPC value of the firm's headquarters location county. As updated values of SPC and PPC are unavailable beyond 2016, we use the last available cross-section for classification purposes. Despite this limitation, the average percentage differences indicate that these variables exhibit approximately 30% cross-time variation from the first available date (2007) to the most recent date (2016). This relatively moderate variation, coupled with their

stability over time, supports their suitability as stratification instruments. Such consistency helps ensure that any observed effects are less likely to be driven by short-term fluctuations, thereby strengthening the reliability of the analysis.

A: SCI ranking by percentage link to New York County, NY



B: SCI ranking by percentage link to Austin County, TX

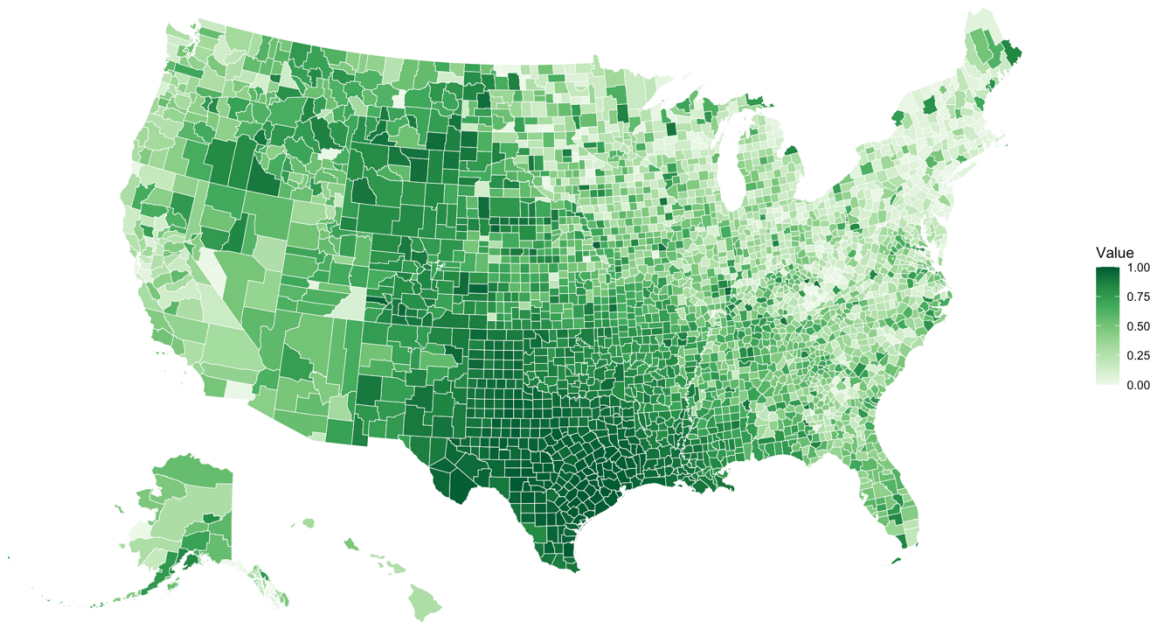


Figure A1: Heat Map examples of SCI of New York, NY and Austin, TX

This figure demonstrates the Social Connectedness Index rank by percentile, 1 is the most connected county with target and 0 is the least connected county. Panel A shows the SCI ranks link to New York county in New York state. And Panel B shows the SCI ranks link to Austin county in Texas State.

ONLINE APPENDIX

Table OA1: US SVI with General Audience and US Audiences' Engagement

This table reports the regression results of US SVI of all engagements and US audience engagements from Jan 2015 to Dec 2024, respectively. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	US SVI _{<i>i,t</i>}					
	(1)	(2)	(3)	(4)	(5)	(6)
Eng_{<i>i,t</i>}	0.007 (0.008)		0.006 (0.007)		0.005 (0.005)	
AbsSent _{<i>i,t</i>}	0.337 (0.554)		0.497** (0.214)		0.459** (0.206)	
US Eng_{<i>i,t</i>}		0.023 (0.020)		0.019* (0.012)		0.023** (0.010)
Abs US Sent _{<i>i,t</i>}		0.205 (0.567)		0.469** (0.223)		0.432** (0.216)
US SVI _{<i>i,t-1</i>}			0.371*** (0.017)	0.371*** (0.017)	0.367*** (0.016)	0.367*** (0.016)
US SVI _{<i>i,t-2</i>}			0.054*** (0.013)	0.054*** (0.013)	0.055*** (0.013)	0.055*** (0.013)
US SVI _{<i>i,t-7</i>}			0.341*** (0.017)	0.341*** (0.017)	0.344*** (0.017)	0.344*** (0.017)
ln(MV _{<i>i,t-7</i>})					0.043 (0.050)	0.043 (0.050)
Turn _{<i>i,t-7:t-36</i>}					0.008 (0.009)	0.008 (0.009)
RealVol _{<i>i,t-7:t-36</i>}					-18.186*** (6.916)	-18.194*** (6.918)
Ret _{<i>i,t-7:t-36</i>}					-0.452** (0.198)	-0.452** (0.198)
EA _{<i>i,t</i>}					23.250*** (1.830)	23.250*** (1.830)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y
N	500,449	500,449	500,302	500,302	499,386	499,386
R-sq	0.684	0.684	0.811	0.811	0.818	0.818

Table OA2: Robustness for VADER Sentiment, a Dictionary Method

This table reports the regression results of states' audience on States SVI by using the VADER method to test the sentiment from Jan 2015 to Dec 2024. States SVI is the average number of four US states, which are CA, TX, NY and FL. The States engagements is the sum of all four States' number of engagements. And the Abs States Sentiment is the absolute value of the weighted average of GPT sentiment score of all the post for firm-day. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	States SVI _{it}	CA SVI _{it}	TX SVI _{it}	NY SVI _{it}	FL SVI _{it}
	(1)	(2)	(3)	(4)	(5)
FourStateEng_{it}	0.100***	0.448***	0.697***	0.668**	0.284**
	(0.031)	(0.157)	(0.070)	(0.317)	(0.139)
VADER FourStateAbsSent _{it}	0.366	0.363	0.335	0.386	0.816**
	(0.261)	(0.342)	(0.365)	(0.379)	(0.344)
Dependent SVI _{it-1}	0.386***	0.298***	0.268***	0.257***	0.246***
	(0.014)	(0.014)	(0.013)	(0.014)	(0.012)
Dependent SVI _{it-2}	0.032**	0.087***	0.098***	0.097***	0.108***
	(0.015)	(0.011)	(0.013)	(0.011)	(0.012)
Dependent SVI _{it-7}	0.367***	0.276***	0.281***	0.259***	0.267***
	(0.018)	(0.015)	(0.018)	(0.017)	(0.016)
ln(MV _{it-7})	0.023	0.048	0.097	0.065	-0.039
	(0.034)	(0.050)	(0.085)	(0.077)	(0.056)
Turn _{it-7,t-36}	0.014***	0.022***	0.026***	0.028***	0.025**
	(0.005)	(0.006)	(0.009)	(0.009)	(0.011)
RealVol _{it-7,t-36}	-1.591	2.267	1.682	1.774	14.874
	(5.165)	(8.991)	(9.256)	(9.985)	(10.687)
Return _{it-7,t-36}	-0.281**	-0.284	-0.273	-0.338	0.123
	(0.128)	(0.219)	(0.240)	(0.213)	(0.209)
EA _{it}	13.177***	16.547***	10.564***	16.277***	9.507***
	(1.409)	(1.753)	(1.352)	(1.648)	(1.169)
Firm FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y
N	495,890	495,747	495,890	495,890	495,890
R-sq	0.915	0.815	0.816	0.804	0.803

Table OA3: Different Benchmarks of PPC and SPC

This table reports the regression results of states' audience on States SVI by subsamples of firms based on their connectedness to investors from Jan 2015 to Dec 2024, groups are separated by mean value of SPC and PPC from 2013 to 2016. States SVI is the average number of four US states, which are CA, TX, NY and FL. The States engagement is the sum of all four States' number of engagements. And the Abs States Sentiment is the absolute value of the weighted average of GPT sentiment score of all the post for firm-day. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: States SVI _{<i>it</i>}				
	SPC Group		PPC Group	
	High (1)	Low (2)	High (3)	Low (4)
FourStateEng_{<i>it</i>}	-0.170	0.085***	0.231	0.077***
	(0.347)	(0.031)	(0.406)	(0.022)
FourStateAbsSent _{<i>it</i>}	0.337	0.597	0.278	0.642
	(0.488)	(0.411)	(0.416)	(0.485)
States SVI _{<i>it-1</i>}	0.347***	0.424***	0.361***	0.401***
	(0.018)	(0.023)	(0.019)	(0.024)
States SVI _{<i>it-2</i>}	0.053**	-0.023	0.040*	-0.005
	(0.021)	(0.022)	(0.023)	(0.022)
States SVI _{<i>it-7</i>}	0.368***	0.357***	0.372***	0.353***
	(0.023)	(0.031)	(0.025)	(0.032)
ln(MV _{<i>it-7</i>})	0.078	0.002	0.249	0.007
	(0.078)	(0.022)	(0.191)	(0.021)
Turn _{<i>it-7:t-36</i>}	0.026**	0.000	0.035***	0.000
	(0.011)	(0.006)	(0.008)	(0.007)
RealVol _{<i>it-7:t-36</i>}	-5.586	-9.685	-9.648	-7.433
	(8.695)	(6.360)	(8.400)	(6.279)
Return _{<i>it-7:t-36</i>}	-0.412**	-0.162	-0.445*	-0.200
	(0.203)	(0.184)	(0.248)	(0.161)
EA _{<i>it</i>}	11.689***	15.796***	12.710***	14.687***
	(1.675)	(2.614)	(2.054)	(2.405)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Week FE	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y
N	210,619	207,027	209,895	207,751
R-sq	0.903	0.932	0.910	0.925

Table OA4: Influencers' Occupation and Robustness for Attention Grabbing

This section tests the robustness of the results by excluding accounts associated with occupations in media and official firm accounts. As shown in Table OA4.1, most influencer accounts represent real individuals, but approximately 20% are associated with firms, politicians, and news organizations. These accounts are excluded to mitigate the potential influence of firm announcements, news reporting, policy updates, and similar external signals. Table OA4.2 presents the regression results after removing these accounts.

The results in Table OA4.1 indicate that, even after excluding these alternative information channels, the spillover attention effect remains robust. The spillover impact in targeted locations continues to be significant.

Table OA4.1: Top 10 Influencers' Occupation

This table reports the top 10 occupation of the targeted influencers.

Occupation	Number of Influencers	Percentage
Influencer	745	24.84
Singer	387	12.9
Company	267	8.9
Football Player	251	8.37
Actress/Actor	200	6.67
Sportsman	155	5.17
Musician	97	3.23
Comedian	89	2.97
Politician	85	2.83
Journal	83	2.77

Table OA4.2: Excluding News and Official Accounts

This table reports the regression results of the audience engagement on US and States SVI after excluding posts from media and firms' official accounts from Jan 2015 to Dec 2024. States SVI is the average number of four US states, which are CA, TX, NY and FL. The States engagements is the sum of all four States' number of engagements. And the Abs States Sentiment is the absolute value of the weighted average of GPT sentiment score of all the post for firm-day. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

	US SVI _{<i>i,t</i>}		States SVI _{<i>i,t</i>}	
	(1)	(2)	(3)	(4)
Eng_{<i>i,t</i>}	0.006 (0.005)		0.012*** (0.004)	
AbsSent _{<i>i,t</i>}	0.224 (0.184)		0.187 (0.136)	
US Eng_{<i>i,t</i>}		0.023** (0.010)		
US AbsSent _{<i>i,t</i>}		0.205 (0.193)		
FourStateEng_{<i>i,t</i>}				0.085*** (0.017)
FourStateAbsSent _{<i>i,t</i>}				0.238 (0.170)
Lagged SVI _{<i>i,t-1</i>}	0.366*** (0.016)	0.366*** (0.016)	0.391*** (0.014)	0.391*** (0.014)
Lagged SVI _{<i>i,t-2</i>}	0.055*** (0.013)	0.055*** (0.013)	0.029* (0.017)	0.029* (0.017)
Lagged SVI _{<i>i,t-7</i>}	0.344*** (0.017)	0.344*** (0.017)	0.377*** (0.020)	0.377*** (0.020)
ln(MV _{<i>i,t-7</i>})	0.198 (0.131)	0.198 (0.131)	0.101 (0.106)	0.101 (0.106)
Turn _{<i>i,t-7;t-36</i>}	0.010 (0.008)	0.010 (0.008)	0.013*** (0.004)	0.013*** (0.004)
RealVol _{<i>i,t-7;t-36</i>}	-17.831** (7.059)	-17.837** (7.059)	0.165 (4.669)	0.168 (4.668)
Return _{<i>i,t-7;t-36</i>}	-0.514** (0.206)	-0.514** (0.206)	-0.253** (0.123)	-0.253** (0.123)
EA _{<i>i,t</i>}	23.400*** (1.822)	23.400*** (1.822)	11.779*** (1.273)	11.779*** (1.273)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Week FE	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y
N	497,378	497,378	493,887	493,887
R-sq	0.817	0.817	0.923	0.923

Table OA5: Manufacturing Industry

This table reports the regression results of states' audience on States SVI by subsamples of firms based on firms' industries from Jan 2015 to Dec 2024. States SVI is the average number of four US states, which are CA, TX, NY and FL. The States engagements is the sum of all four States' number of engagements. And the Abs States Sentiment is the absolute value of the weighted average of GPT sentiment score of all the post for firm-day. Variable definitions are consistent with those presented in Table 1. Standard errors in all regressions are clustered by firm and are reported in parentheses. Statistical significance at the 10%, 5%, or 1% levels is denoted by *, **, or ***, respectively.

Dependent Variable: States SVI _{it}						
	Manufacturing Industry			Non-Manufacturing Industries		
	(1)	(2)	(3)	(4)	(5)	(6)
FourStateEng_{it}	0.129***	0.077***	0.081***	0.916	0.367	0.361
	(0.017)	(0.015)	(0.015)	(0.617)	(0.335)	(0.308)
FourStateAbsSent _{it}	-1.095	0.080	0.073	1.264*	0.572*	0.569*
	(0.993)	(0.378)	(0.376)	(0.753)	(0.310)	(0.297)
States SVI _{it-1}		0.379***	0.377***		0.394***	0.390***
		(0.020)	(0.019)		(0.019)	(0.019)
States SVI _{it-2}		0.026	0.027		0.036**	0.035*
		(0.026)	(0.026)		(0.018)	(0.018)
States SVI _{it-7}		0.379***	0.380***		0.356***	0.358***
		(0.033)	(0.033)		(0.021)	(0.021)
ln(MV _{it-7})			0.009			-0.138
			(0.019)			(0.193)
Turn _{it-7:t-36}			0.004			0.018***
			(0.002)			(0.005)
RealVol _{it-7:t-36}			-1.085			-4.652
			(8.083)			(6.432)
Return _{it-7:t-36}			-0.090			-0.353**
			(0.216)			(0.166)
EA _{it}			9.889***			15.849***
			(1.860)			(1.963)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Week FE	Y	Y	Y	Y	Y	Y
Weekday FE	Y	Y	Y	Y	Y	Y
N	215,328	215,293	215,116	281,625	281,513	280,774
R-sq	0.869	0.926	0.927	0.820	0.899	0.903

In Table OA5, we examine heterogeneity across industries, focusing on manufacturing vs. non-manufacturing firms. The spillover effect of States Engagements is both economically and statistically stronger among manufacturing firms.