HKU BUSINESS SCHOOL
港大經管學院

HKU Jockey Club
Enterprise Sustainability
Global Research Institute

Hong Kong University Business School:
HKU Jockey Club Enterprise Sustainability
Global Research Institute Paper Series

Giving Users What they Want:
Social Media and Anomalies

Paper No. 2026/161

Joseph Engelberg
University of California, San Diego

Byoung-Hyoun Hwang
Nanyang Technological University

Runjing Lu
University of Toronto

William Mullins
University of California, San Diego

Giving Users What They Want: Social Media and Anomalies

Joseph Engelberg
UC San Diego

Byoung-Hyoun Hwang
NTU Singapore

Runjing Lu
U. of Toronto

William Mullins
*UC San Diego**

Abstract

Using StockTwits data from 2010 to 2024, we show that stocks in the "short leg" of over 200 return anomalies attract substantially more posts and more bullish sentiment than stocks in the corresponding "long leg." The differences are large: a user is around five times more likely to encounter a bullish post about a short-leg stock than about a long-leg stock. User engagement patterns point to a demand-driven explanation: bullish posts about short-leg stocks receive more likes and generate greater follower growth for their authors. Because content creators value likes and followers, this creates an incentive to amplify optimistic narratives about stocks with low expected returns.

Keywords: Social media, anomalies, social finance

JEL Classification: G12, G14, G41

*Contact: jengelberg@ucsd.edu, bh.hwang@ntu.edu.sg, runjing.lu@utoronto.ca, wmullins@ucsd.edu. The authors would like to thank David Hirshleifer, Jeff Pontiff; seminar participants at Seoul National University; and conference participants at the 2025 Financial Research Conference for their valuable suggestions and comments. The authors would also like to thank Andrew Chen and Tom Zimmermann for generously providing their anomaly data on https://www.openassetpricing.com. This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE-T2EP40122-0003).

1. Introduction

Academic finance has identified hundreds of stock characteristics that predict abnormal returns. McLean and Pontiff (2016) document that these *anomalies* do not dissipate after they are published and disseminated. At the same time, few innovations have transformed the dissemination of information as profoundly as social media, and growing evidence suggests that investor social media directly influences which stocks investors buy and sell (Ammann and Schaub, 2021, Farrell et al., 2022). In this paper, we connect these two facts and ask how the largest investor social media platform, StockTwits, disseminates information about anomalies. Does it steer investors toward or away from the stocks that the finance literature suggests they should hold?

To answer this question, we begin with the anomalies in Chen and Zimmermann (2022). For each of over 200 anomalies, we sort stocks into quintiles and define the top quintile, comprising stocks with high expected future returns, as the "long leg," and the bottom quintile, comprising stocks with low expected future returns, as the "short leg." Following the literature, in each stock-month, NET is the number of long-leg portfolios containing the stock minus the number of short-leg portfolios containing it.[1] For example, a stock that appears in 25 long-leg portfolios and 10 short-leg portfolios has a NET value of 15. Thus, the higher a stock's NET value, the more frequently it appears in long-leg portfolios relative to short-leg portfolios.

Our first result confirms the expected return pattern documented in the anomalies literature (e.g., Chen and Zimmermann, 2022, 2023): from 2010 to 2024, high-NET stocks outperform low-NET stocks. Stocks in the top NET quintile earn returns that are, on average, 2.0 percentage points (pp) higher than bottom quintile stocks over the following month, and 18.3 pp higher over the following year.

[1] Work using the NET variable includes Engelberg, McLean, and Pontiff (2018, 2020), Jiang, Liu, Peng, and Wang (2022), McLean, Pontiff, and Reilly (2025).

We next examine how the volume and sentiment of user posts vary across NET quintiles. One advantage of StockTwits data is that users can explicitly label their posts as "bullish" or "bearish," which eliminates the need to infer "sentiment" from text.[2] We document two robust patterns. First, stocks in the bottom quintile of NET (short-leg stocks) receive about five times as many posts as stocks in the top quintile (long-leg stocks) in an average month, and this gradient is monotonic. Second, sentiment is markedly more positive for short-leg stocks than for long-leg stocks. Per stock-month, bottom-NET-quintile stocks receive 472 more bullish posts (net of bearish posts) than top-quintile stocks. From the user's perspective, an individual who encounters a sentiment-labeled post at random has a 29.3% chance of seeing a bullish post about a short-leg stock, but only a 6.0% chance of seeing a bullish post about a long-leg stock. We observe the opposite pattern in traditional media: in *Dow Jones Newswires* articles, the most positive sentiment is about long-leg stocks.

The short-leg tilt on StockTwits also varies systematically across anomaly types. Consistent with the social-transmission-bias framework in Han et al. (2022), this tilt is strongest for anomaly types related to lottery-like payoffs, skewness, and volatility. The pattern is concentrated among stocks with characteristics that attract attention, stimulate discussion, and support optimistic narratives.

The bullish sentiment toward short-leg stocks on social media also holds within stocks over time. In a panel regression with stock fixed effects, post volume increases and sentiment becomes more positive when a stock moves from a high-NET quintile to a low one.

These facts raise the question: what economic mechanisms could explain why StockTwits users post so positively about short-leg stocks despite their low expected returns? We conjecture that the pattern reflects posters supplying content that matches their audience's preferences, consistent with what the economics literature calls "consumer-driven slant" or

[2] "Sentiment" can refer to many dimensions of textual content. In this paper, we use the term sentiment to refer to whether a post is bullish or bearish about the stock.

"audience capture." A literature in media economics shows that outlets slant content toward consumers' priors because consumers demand confirmatory information (Mullainathan and Shleifer, 2005, Gentzkow and Shapiro, 2010).[3] A distinctive feature of our StockTwits data is that we observe not only the supply of messages but also indications of demand, such as likes and follows, which lets us test this demand-based hypothesis.

We first provide general evidence of "audience capture" by showing that influencers, defined as users in the top 1% of the StockTwits follower distribution, are sensitive to their followers' interests. Specifically, when an influencer's followers discuss a stock that the influencer did not cover in the previous month, the influencer is substantially more likely to post about that stock in the current month. In addition, influencers are more likely to continue posting about a stock when their prior posts about that stock received more likes. The evidence suggests that influencers give users what they want.

We find strong evidence that (i) bullish posts and (ii) bullish posts about short-leg stocks are precisely what users want. Bullish posts receive more than double the likes per post than bearish ones, which may explain why bullish posts outnumber bearish posts on StockTwits by 5 to 1 in our sample. Turning to anomaly stocks, bullish posts about short-leg stocks attract over 30% more likes per post than bullish posts about long-leg stocks. This result also holds in regressions examining likes per bullish post, net of likes per bearish post ($Bullish-Bearish$), and after including firm and year-month fixed effects. Moreover, when we separately examine influencers, professional users, and other users, we find that all three groups receive more likes per post for $Bullish-Bearish$ short-leg posts. However, only ordinary users (i.e., users who are neither professionals nor influencers), *give* more likes to $Bullish-Bearish$ short-leg posts compared to long-leg posts. To the extent that bullish posts about short-leg stocks

[3]Mullainathan and Shleifer (2005) model demand-driven media bias, and Gentzkow and Shapiro (2010) show that newspaper slant tracks the political composition of readers rather than owner ideology. A growing literature applies this lens to misinformation on social media (e.g., Allcott and Gentzkow, 2017, List et al., 2024, Thaler, 2024).

constitute bad information, professional users and influencers disproportionately supply this information but do not disproportionately like it.

We observe a similar pattern when we examine users' following behavior on StockTwits. In general, posting more bullishly in a given month predicts greater follower growth in the following month. However, the effect is more than three times as large for bullish posts about short-leg stocks as for bullish posts about long-leg stocks. For example, a 10 pp increase in the fraction of bullish posts about short-leg stocks predicts 2.9 new followers in the following month; for bullish posts about long-leg stocks, it predicts only 0.8 new followers.

Taken together, our evidence is consistent with a demand-driven mechanism: bullish posts about short-leg stocks earn more likes and followers, and content creators respond by supplying them.

Contributions to the literature. Our study adds to the growing literature on the role of social media in financial markets (Cookson, Mullins, and Niessner, 2024b). An increasing share of investor interactions now takes place on social media, and research on investment-oriented social media platforms shows that they can provide valuable information to investors (e.g., Chen et al., 2014, Avery et al., 2016, Da and Huang, 2020, Farrell et al., 2022, Bradley et al., 2024, Cookson et al., 2024a, Dim, 2025). Other work highlights adverse effects of social media engagement. Users favor posts and other users that confirm their prior beliefs, leading to echo chambers both on general-interest platforms (Bakshy et al., 2015, Cinelli et al., 2021) and among investors on StockTwits (Cookson, Engelberg, and Mullins, 2023). Users also disproportionately seek out and share eye-catching content, speeding the spread of false news and misinformation (Vosoughi et al., 2018, Del Vicario et al., 2016); in financial markets, social media has been used to manipulate prices through fraudulent news (Kogan, Moskowitz, and Niessner, 2023). Our study builds on this latter strand by providing evidence that social media users are disproportionately drawn to bullish posts about short-leg securities and that content creators cater to this demand. These engagement incentives favor the diffusion of biased information about overvalued stocks.

Our findings also speak to the literature on social finance (Kuchler and Stroebel, 2021). Recent research recognizes that investors turn to each other for advice and that social interactions can influence investment decisions. Through "social transmission biases," these interactions can systematically tilt toward certain types of stocks and affect asset prices (e.g., Hirshleifer, 2020, Han, Hirshleifer, and Walden, 2022).[4] Our finding that social media systematically skews investor conversations toward short-leg stocks, particularly those with lottery-like characteristics, is consistent with this view.

Finally, our paper contributes to the large literature on stock return anomalies. A growing body of research shows that the profitability of many anomaly strategies comes primarily from the short leg rather than the long leg, and that limits to arbitrage, such as short-sale constraints and borrowing costs, can prevent arbitrageurs from correcting this overpricing (e.g., Stambaugh et al., 2012, Chu et al., 2020, Stambaugh and Yuan, 2017).[5] McLean, Pontiff, and Reilly (2025) link this overpricing to retail demand, showing that retail investors systematically accumulate low-forecasted-return (short-leg) stocks and that anomaly mispricing is stronger where retail trading is more intense. Our findings complement this literature by highlighting a previously underexplored driver of that demand: the appeal of optimistic narratives about short-leg stocks on social media.

[4]To illustrate, suppose that investors (i) have a preference for discussing stocks with certain features, (ii) tend to purchase stocks that they become aware of (Barber and Odean, 2008), and (iii) face short-sale constraints. Social interactions and synchronized purchases can then generate overpricing for stocks with these features. Han, Hirshleifer, and Walden (2022) identify two characteristics toward which investor conversations are likely to be biased: high volatility and high skewness. The intuition is that investors prefer to discuss high-return experiences, and stocks with high volatility and skewness are more likely to have generated them.

[5]Possible reasons short-leg stocks may become overpriced include disagreement (Diether, Malloy, and Scherbina, 2002), extrapolation (La Porta, 1996, Barberis, Greenwood, Jin, and Shleifer, 2018), analyst optimism (Engelberg, McLean, and Pontiff, 2020), and a preference for lottery-like stocks (Barberis and Huang, 2008, Kumar, 2009, Bali, Cakici, and Whitelaw, 2011, Chen, Hwang, and Peng, 2025).

2. Data

This paper studies the relationship between social media activity and anomaly-based mispricing. Our main analysis combines StockTwits data with 209 documented stock return anomalies. In additional analyses, we compare the patterns on StockTwits with those in traditional media coverage, specifically from *Dow Jones Newswires*.

2.1 Social media

Our main data on social media activity come from StockTwits, which has been widely used in academic studies (e.g. Cookson et al., 2024a). The data are obtained via the StockTwits API and comprise detailed post- and interaction-level records from August 2010 through December 2024. The sample consists of the 1,500 most discussed stocks on StockTwits during 2010 to 2021 as identified in Cookson et al. (2024a), and the 1,500 most discussed stocks during 2022 to 2024, yielding 1,796 unique stocks.

A key advantage of StockTwits is that users can self-label their posts as "bullish" or "bearish," allowing us to observe sentiment directly and avoid the noise associated with inferring sentiment using natural language processing (NLP). We restrict our analysis to posts that reference a single ticker through a "cashtag," that is, a dollar sign followed by a ticker symbol, ensuring that sentiment is unambiguously associated with a single stock. Another appealing feature of our data is that they contain detailed information on user interactions, including whether users *like* other users' posts, as well as whether and when they *follow* other users.

We measure the volume of discussion for a given stock using the total number of StockTwits posts that mention the stock in a given month (*Posts*). To account for variation in overall platform activity, we also compute the fraction of posts referring to a given stock relative to all posts in that month (*%Posts*).

To capture the average sentiment across these posts, we compute the difference between the number of bullish and bearish posts at the stock-month level (*NetBullishPosts*). For example, a value of 100 in month t indicates that the stock receives 100 more bullish posts than bearish ones. We also construct a scaled measure as the difference between the fraction of bullish and bearish posts (*NetBullishShare*).

We measure post popularity using the number of likes a post receives within 30 days of posting and aggregate these counts to the stock-month level (*Likes*). To facilitate comparisons across stocks with different levels of post volume, we also compute the average number of likes per post (*LikesPerPost*). To capture the popularity gap between bullish and bearish posts, we compute the difference between likes on bullish and bearish posts (*NetBullishLikes*) and its per-post analogue (*NetBullishLikesPerPost*).

Finally, we consider user following behavior as an additional measure of popularity. Unlike *Likes*, follows are not tied to individual posts. We measure following activity at the user-month level as the total number of new followers a user receives in a given month (*NewFollower*) and link this measure to the user's posting behavior in the prior month.

2.2 Anomalies

Our anomaly variable is based on 209 firm characteristics documented in the literature, as summarized by Chen and Zimmermann (2022), and made publicly available on Open Source Asset Pricing (OSAP, October 2025 release).[6] These firm characteristics are based on studies published in peer-reviewed finance, accounting, and economics journals and have been shown to predict the cross section of average stock returns.

To construct our stock-month anomaly measure, we first sort stocks each month on each signed firm characteristic so that higher values correspond to higher expected returns. We

[6] Chen and Zimmermann (2022) report 212 anomalies in total. Three governance-related anomalies—*Activism 1*, *Activism 2*, and *Governance*—are unavailable during our sample period, leaving 209 anomalies for the analysis.

define the long and short legs of each anomaly strategy as the top and bottom quintiles of the stock universe in Chen and Zimmermann (2022). For anomalies based on indicator variables, such as credit rating downgrades, there is only a long or short side based on the binary value of the indicator.

Following Engelberg et al. (2018, 2020), we construct an anomaly index, NET, to summarize a stock's relative tilt toward long-leg versus short-leg anomaly portfolios. For each stock and month, NET is defined as the difference between the number of anomaly portfolios in which appears in the long leg and the number in which it resides in the short leg. For example, a NET value of ten in month t indicates that the stock belongs to ten more anomaly long-leg portfolios than short-leg portfolios in that month.

2.3 Other data

To benchmark our social media results, we study the relationship between anomalies and traditional news coverage. Specifically, we obtain article-level data for *Dow Jones Newswires* from Ravenpack 1.0. We keep all articles with a relevance score above 75 and use Ravenpack's Event Sentiment Score (ESS) as our sentiment measure. We then aggregate the data to the stock-month level using the same procedure described above.

2.4 Sample

Table 1 reports summary statistics for the StockTwits sample at the stock-month level. The sample spans August 2010 to December 2024 for posts and sentiment, and September 2012 to December 2024 for likes. As shown in Panel A, the average stock-month in the full StockTwits sample has 917 posts, with substantial cross-sectional dispersion (standard deviation of 7,359). StockTwits posts are more often bullish than bearish: the average stock-month has 383 bullish posts and 75 bearish posts, a 5:1 bullish:bearish ratio. Among stock-months with at least one post, the average stock receives 2,082 likes and 0.9 likes per post.

Panels B and C report summary statistics separately for stock-month observations belonging to the bottom quintile of NET (NET 1 or "short-leg stocks") and the top quintile (NET 5 or "long-leg stocks") in the prior month, respectively. Two patterns stand out. First, NET 1 stocks attract substantially more attention on StockTwits. In an average month, stocks in NET 1 receive roughly five times as many posts as stocks in NET 5. They also receive more engagement: seven times as many likes and nearly twice as many likes per post. Second, sentiment is markedly more bullish for NET 1 stocks than for NET 5 stocks (0.05% of posts are bullish, versus 0.01%) . In an average month, NET 1 stocks have 472 more bullish posts, after subtracting bearish posts, than NET 5 stocks, a variable we call *Bullish minus Bearish* (or $Bullish - Bearish$). That is, StockTwits posters are substantially more bullish about NET 1 than about NET 5 stocks, even net of bearish posts.

3. Univariate analyses

We begin by characterizing how NET relates to future stock returns and social media activity. Each month, we sort stocks into quintile portfolios based on their prior-month NET value and track the number and share of *Posts* and their net sentiment, *Bullish minus Bearish*.

Figure 1 plots equal-weighted average returns for each NET quintile over the subsequent month, quarter, and year. Consistent with the literature, there is a strong and monotonic relationship between a stock's net anomaly exposure and its subsequent performance. Stocks in the highest NET quintile bin ("long-leg stocks") earn substantially higher returns than those in the lowest NET quintile bin ("short-leg stocks"), and the spread widens considerably as the horizon extends from one month to one year. The return differential between the top and bottom NET bins is 2.0 pp at the one-month horizon and 18.3 pp at the one-year horizon. Appendix Figure A1 shows that this return predictability holds in almost every year of the sample. These patterns are consistent with the anomaly literature (e.g. McLean and Pontiff, 2016, Chen and Zimmermann, 2022) and confirm that our NET measure captures cross-sectional differences in expected returns over our sample period.

Panel (a) of Figure 2 reports the distribution of posts across NET quintiles, separately for posts that users label as bullish or bearish, and for unlabeled posts. Three patterns emerge. First, posts are disproportionately concentrated in short-leg stocks: the bottom quintile accounts for nearly five times the post share of the top quintile, even though the two groups contain similar numbers of stocks by construction. Second, the gradient is monotonic across quintiles. Third, the skew is strongest for bullish posts, whereas bearish posts are more evenly distributed across quintiles. In Appendix Figure A2, we restrict the sample to posts with self-labeled sentiment and find an even sharper contrast: a StockTwits user who encounters a sentiment-labeled post at random has a 29.3% chance of seeing a bullish post about a short-leg stock, but only a 6.0% chance of seeing a bullish post about a long-leg stock. By contrast, bearish posts exhibit a much flatter distribution, with only a 5.3% difference between the short- and long-leg quintiles.

Panel (b) of Figure 2 focuses on a measure of net bullishness, *Bullish minus Bearish*, which is the difference between the fraction of bullish and bearish posts. *Bullish minus Bearish* declines monotonically from the lowest to the highest NET quintile. For short-leg stocks, it is 9.0 pp, while for long-leg stocks it is only 1.7 pp. This pattern indicates that the average post encountered by a StockTwits user is not only more likely to be about stocks with poor future returns but also more likely to be optimistic about these stocks.

Appendix Figure A3 replicates these findings using decile sorts and shows that the monotonic relationships in post volume and sentiment are not driven by the use of quintile bins.

4. Regression analyses

We next examine these patterns in a regression framework:

$$Y_{i,t} = \alpha + \beta NET_{i,t-1} + \gamma' X_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t} \tag{1}$$

where the outcome variable $Y_{i,t}$ captures the attention and engagement that stock i attracts in month t. $NET_{i,t-1}$ represents the quintile of stock i's net anomaly exposure in the prior month, while NET quintile 3 is the omitted category. α_i and α_t are the firm and year-month fixed effects, respectively, which absorb time-invariant stock characteristics and time-varying common shocks, isolating within-stock variation in NET quintile and social media activity. In some specifications, we further control for time-varying characteristics of stock i, $X_{i,t}$, such as the presence of 8-K filings and earnings announcements from month $t-1$ through $t+1$.

4.1 Number of posts

Table 2 reports estimates from regressions of the number of posts (*#Posts*) or the fraction of posts (*%Posts*) in a given stock-month on NET quintile indicators, with NET quintile 3 serving as the baseline category. Columns 1 and 2 include only year-month fixed effects, while Columns 3 and 4 add stock fixed effects.

The magnitude of the estimates is substantial. In the specification with only year-month fixed effects, short-leg stocks (bottom NET quintile) receive 623 more posts per month relative to stocks in the middle quintile, while long-leg stocks (top NET quintile) receive 621 fewer posts per month. When we add stock fixed effects, the estimates remain statistically significant and economically large. Short-leg stocks receive 521 more posts per month than the middle quintile, while long-leg stocks receive 312 fewer. The within-stock spread between the short- and long-leg stocks is 833 posts per month, or approximately 90% of the mean.

We find a similar pattern for fraction of posts – number of stock-month posts scaled by total posts in a month. In column 4, short-leg stocks account for 0.05 pp more of total posts than middle-quintile stocks, while long-leg stocks account for 0.02 pp fewer. These estimates incorporate stock fixed effects, so they reflect within-stock variation over time: when a stock moves toward the short leg in terms of net anomaly exposure, posting activity increases, and when it moves toward the long leg, posting activity declines. The stability of the pattern across specifications with and without stock fixed effects further suggests that the tilt toward short-leg stocks is not driven by time-invariant stock characteristics such as size, industry, or name recognition.

The estimated relationship between NET quintile and the number of posts is also monotonic. In Column 3, with stock and year-month fixed effects, quintile 2 stocks receive 294 more posts per month than middle-quintile stocks, while quintile 4 stocks receive 200 fewer posts per month. This monotonicity, which mirrors the univariate patterns in Figure 2, provides additional evidence that the association between anomaly exposure and social media attention is not driven by a few outliers.

The post volume results remain very similar when we exclude stock-months with analyst coverage and when we control for firm news (Appendix Tables A1 and A2, Columns 1 and 2). These results are therefore not driven by firm news or analyst coverage.

4.2 Sentiment of posts

Table 3 turns to the sentiment of posts and again presents results consistent with the univariate analyses. The outcome variables are the number and share of *Bullish minus Bearish* posts, which captures net bullishness.

In the specification with only year-month fixed effects (Column 1), short-leg stocks receive 322 more bullish posts (net of bearish posts) per month than those in the middle quintile, and long-leg stocks receive 149 fewer. Both are statistically significant at the 1% level.

The estimates again exhibit a monotonic pattern. Quintile 2 stocks receive 140 more net bullish posts per month than middle-quintile stocks, while quintile 4 stocks receive 61 fewer. Adding stock fixed effects (column 3) reduces the magnitudes, but the estimates remain large and significant. short-leg stocks receive 197 more net bullish posts per month than middle-quintile stocks, while long-leg stocks receive 138 fewer.

Results are similar for the net bullish fraction of posts (*Bullish minus Bearish %Posts*) in columns 2 and 4. Whether or not we include stock fixed effects, the net bullish share is higher for bottom-NET-quintile stocks than for the middle quintile and lower for top quintile stocks.

Results are also similar when we exclude stock-months with analyst coverage and when we control for firm news (Appendix Tables A1 and A2, Columns 3 and 4.)

The regression results indicate that the bullish tilt toward short-leg stocks documented in the univariate analysis is not driven by composition effects or time-invariant stock characteristics. Rather, when a stock's net anomaly exposure moves toward the short leg, social media sentiment becomes more optimistic. This within-stock pattern is the opposite of what one would expect if social media efficiently processed information about expected returns.

4.3 User heterogeneity in posting behavior

Figure 3 decomposes posting patterns by user type. The figure plots coefficient estimates from the within-stock specification, estimated separately for professional users, influencers (users in the top 1% of follower counts), and ordinary users (all remaining users), with NET quintile 3 serving as the comparison group.

Panel (a) plots results for the fraction of posts. All user types post more about short-leg stocks relative to the middle quintile, but the gradient is steepest among ordinary users - those who are neither professionals nor influencers. Ordinary users exhibit larger positive coefficients for bottom-quintile stocks and more negative coefficients for top-quintile stocks, although the differences are not always statistically significant.

Panel (b) reports results for post sentiment. All user types are more bullish about short-leg stocks and more bearish about long-leg stocks compared to middle quintile stocks. The spread between short- and long-leg stocks is again most pronounced among ordinary users. However, professionals and influencers also exhibit a systematic tilt toward short-leg stocks. The magnitude is smaller than for other users, but even these likely more-sophisticated participants disproportionately discuss stocks with low future returns and express more bullish views about them. This posting behavior is consistent with the demand-driven explanation we develop below: if content creators are rewarded with likes and follows for bullish posts about short-leg stocks, they have an incentive to supply such content regardless of their private beliefs about expected returns.

5. Traditional news coverage

Table 4 examines whether the attention and sentiment patterns on social media also appear in traditional media, specifically *Dow Jones Newswires*. We distinguish between corporate disclosure-related articles (columns 1–4) and other firm news (columns 5–8).

We first examine article volume and then turn to article sentiment. The volume results for traditional media largely mirror those for social media. Among corporate disclosures, stocks in the short leg receive 0.5 more articles per month than middle NET quintile stocks. For other news, short-leg stocks receive 2 more articles per month. However, unlike on social media, where long-leg stocks also receive substantially fewer posts, long-leg stocks do not see systematically lower coverage on traditional media relative to middle quintile stocks.

The starker difference between social media and traditional media lies in sentiment. On StockTwits, net bullishness is clearly higher for short-leg stocks and markedly lower for long-leg stocks (see Table 3). Traditional media exhibits the opposite pattern. For corporate disclosure, the net bullish count is 0.06 *lower* for short-leg stocks than for quintile 3 stocks and 0.07 higher for long-leg stocks (column 3). Similarly, for other news, the net bullish count is 0.2 lower for short-leg stocks and 0.3 higher for long-leg stocks (column 7). Article fraction shows similar patterns.

Traditional media sentiment thus aligns with the direction of subsequent returns: articles about short-leg stocks are more negative than those about long-leg stocks, consistent with short-leg stocks delivering poorer future performance. Social media sentiment, by comparison, tilts in the *opposite* direction. That coverage volume is higher for short-leg stocks in both media types suggests that some of the additional attention paid to overvalued stocks reflects genuine information production or diffusion.

6. Anomaly Types and Social Transmission Biases

The social finance literature argues that social transmission biases tilt investors toward the stocks in the short-leg of certain anomalies. In particular, Han et al. (2022) argue that investor conversations disproportionately feature stocks with high volatility and high skewness because these stocks are more likely to generate salient high-return experiences. If the StockTwits patterns reflect this mechanism, the short-leg tilt should be strongest for anomalies tied to volatility and lottery-like payoffs.

With this in mind, we decompose the baseline relation between NET and StockTwits activity by anomaly type, using an LLM to assign each anomaly to one of seven categories. Appendix Table A3 reports the mapping between the categories of anomalies from Chen and Zimmermann (2022) and the anomaly types. The categories are: (i) Narrative/Extrapolation/Attention, (ii) Fundamental/Accounting/Valuation, (iii) Investment/Growth/Innovatio
(iv) Risk/Volatility/Lottery, (v) Liquidity/Trading Frictions/Market Microstructure, and (vi) Financing/Payout/Ownership, and (vii) Other.

Table 5 reports the mapping from the Chen and Zimmermann (2022) anomaly categories to these types. To facilitate the decomposition by anomaly type, we use the continuous version of NET rather than NET quintile indicators. Panel A first repeats the baseline specification using this continuous measure. Consistent with the quintile-based results in Tables 2 and 3, the coefficients on NET are negative and significant across all specifications. As a stock moves toward the short leg of anomalies, it receives more StockTwits attention and more bullish sentiment. For example, a stock that belongs to one more long-leg portfolio than short-leg portfolio receives 22 fewer posts and 9 fewer net bullish posts in the following month.

Panel B decomposes this relation by anomaly type: the strongest gradient appears for *Risk/Volatility/Lottery*. Across outcomes, the estimated effects for this type are roughly two to five times as large as the baseline gradient in Panel A. The next strongest gradients come from *Liquidity/Trading Frictions/Market Microstructure*, followed by *Financing/Payout/Ownership*. By contrast, the gradients are small and generally insignificant for the remaining anomaly types. Thus, the social-media tilt we document is strongest for the subset of short-leg stocks whose volatility and lottery-like characteristics make them especially likely to generate attention and discussion, consistent with the theory of Han et al. (2022).

7. A Demand-Based Explanation: Likes and Follows

In this section, we examine several dimensions of social media engagement to understand why StockTwits users disproportionately produce bullish posts about short-leg stocks. One possible explanation is demand-driven: users reward certain types of content with likes and follows, thereby generating clear incentives for content creation. We begin by examining whether users disproportionately reward bullish posts about short-leg stocks, creating incentives to supply such content, and then test whether content creators appear to respond to follower demand.

7.1 Liking behavior and incentives to post

We begin with a univariate analysis. Figure 4 examines the distribution of likes. Panels (a) and (b) show the % of *Likes* and the # of *Likes per post* across bullish, bearish, and unlabeled posts. Both panels tell a consistent story: bullish posts attract substantially more likes than bearish posts, by more than 2 to 1 on a per-post basis. Panels (c) and (d) examine the distribution of likes across NET quintiles. The pattern for likes closely mirrors, and is even stronger than, the pattern for posts: likes are disproportionately concentrated in bullish posts about short-leg stocks, and the bullish-minus-bearish gap is substantially larger for short-leg than for long-leg stocks. This concentration partly reflects the larger number of bullish posts about short-leg stocks. To separate this extensive-margin effect from per-post engagement, Panels (e) and (f) repeat the distribution of likes on a per-post basis. The same pattern remains: bullish posts about short-leg stocks receive the most likes per post among all post types, 30% more than bullish posts about long-leg stocks. Thus, short-leg bullish content is not only more prevalent but also attracts stronger engagement.

Table 6 reports estimates from regressions that relate engagement to stocks' net anomaly exposure in the previous month. All specifications include stock and year-month fixed effects. For each stock-month, we compute the number of likes on posts mentioning the stock, the fraction of likes relative to likes on all posts in that month, and the average number of likes per post. Only stock-month observations with at least one post enter the regressions.

Posts about short-leg stocks receive 1,492 more likes per month than posts about quintile 3 stocks (column 1), or 0.06 pp more of total likes (column 2). On a per-post basis, each post about a short-leg stock receives 0.20 more likes than a post about a quintile 3 stock (column 3), equivalent to 22% of the mean (0.91). By comparison, posts about long-leg stocks receive

841 fewer likes per month than posts about quintile 3 stocks, 0.02 pp less of total likes, and 0.09 fewer likes per post. All estimates are significant at the 1% level, and the relationship across quintiles is monotonic.

Columns 4 through 6 report findings for net bullish likes, that is, the difference in likes between bullish posts and bearish posts. Posts about short-leg stocks receive 925 more net bullish likes per month than posts about quintile 3 stocks (column 4), and each post about a short-leg stock receives 0.20 more net bullish likes per post (column 6). By contrast, posts about long-leg stocks receive 516 fewer net bullish likes per month (column 4) and 0.06 fewer net bullish likes per post (column 6). Thus, not only are short-leg stocks discussed more and with more bullish sentiment, but bullish posts about these stocks are also disproportionately rewarded with more likes.

The per-post engagement patterns are important because they rule out a mechanical explanation: the engagement differential is not simply a consequence of there being more posts about short-leg stocks. Even after scaling by the number of posts, each individual post about a short-leg stock receives substantially more likes than a post about a long-leg stock, and this difference is stronger for bullish posts than for bearish ones. This pattern points to a demand-driven mechanism in which users disproportionately favor bullish content about short-leg stocks, creating incentives to produce such content.

7.2 Providers and recipients of *Likes*

Figures 5 and 6 decompose the engagement patterns by user type, distinguishing between providers and recipients of likes.

Figure 5 plots coefficient estimates from regressions of received likes on NET quintile indicators, estimated separately by the type of user who *wrote* the post. Panel (a) shows that all user types receive more likes on posts about short-leg stocks. The negative slope is steepest among ordinary users, but is also present among professionals and influencers. Panel

(b) examines net bullish likes, showing that all user types receive disproportionately more likes on their bullish posts about short-leg stocks. Panels (c) and (d) repeat the analysis using likes per post and show similar patterns. On a per-post basis, posts about short-leg stocks receive more likes, whereas posts about long-leg stocks receive fewer. This pattern is stronger for bullish posts than for bearish posts, regardless of who wrote the post.

Figure 6 turns to the demand side, plotting coefficients estimated separately by the type of user who *gives a like* to a post. Panel (a) shows that all users, regardless of type, give more likes to posts about short-leg stocks and fewer to those about long-leg stocks, relative to the middle quintile. However, Panel (b) shows that net bullish likes are driven primarily by ordinary users. The per-post results in Panels (c) and (d) show similar patterns.

Taken together, Figures 5 and 6 suggest a demand-driven feedback loop. Ordinary users, who form the majority of the StockTwits audience, disproportionately reward bullish content about short-leg stocks through likes. This reward accrues to all types of posters, including professionals and influencers. If users are motivated by engagement, this creates an incentive to supply bullish content about short-leg stocks.

Underlying this loop is an asymmetry between who supplies bullish short-leg content and who rewards it. Professionals and influencers receive more likes when they post bullishly about short-leg stocks (Figure 5), yet they do not themselves disproportionately like such content: in Figure 6, only ordinary users show a pronounced negative slope in net bullish likes given across NET bins, while the slopes for professionals and influencers are nearly flat. They thus supply bullish short-leg content without sharing the demand for it. This distinguishes our setting from Engelberg et al. (2020), where analysts appear sincerely more optimistic about short-leg stocks in their forecasts and recommendations. By contrast, professionals' and influencers' own like-giving suggests that their bullish posting reflects audience demand rather than sincere optimism.

7.3 Following behavior and incentives to post

Tables 7 and 8 examine how users' posting behavior predicts growth in their follower base. Table 7 regresses the number of new followers user j receives in month t on the share of bullish and bearish posts the user made in the prior month $(t-1)$, controlling for user and year-month fixed effects. The results are clear: bullish posts predict more followers, and bearish posts predict fewer. The coefficient on the bullish share is 0.262 (column 1), implying that a user who increases their share of bullish posts by 10 pp will gain nearly three additional followers in the following month. Conversely, a user who increases their share of bearish posts by 10 pp will lose one follower in the following month. Thus, bullish posting is rewarded with more followers, whereas bearish posting is penalized on StockTwits.

Columns 2 through 4 decompose new followers by follower type. The follower reward for bullish posts comes from all follower types, but it is largest for new follows from ordinary users, and smallest for new follows from professional users. There is an analogous follower penalty for bearish posts, which is strongest for ordinary users and is absent for professionals.

Table 8 extends the analysis by interacting the sentiment of a user's posts with the NET quintile of the stocks they discuss. This specification allows us to assess whether the follower reward for bullish posting varies by the type of stock being discussed. The key comparison is between bullish posts about short-leg stocks and bullish posts about long-leg stocks (*Bullish* $\times$ $NET1$ vs. *Bullish* $\times$ $NET5$). Column 1 shows that a user who posts 10 pp more bullish content about short-leg stocks will gain nearly three new followers in the following month, but they will gain fewer than one new follower when writing bullish posts about long-leg stocks. This gap is economically large and holds across follower types (columns 2 to 4).

The asymmetry is much weaker on the bearish side. A user who posts 10 percentage points more bearish content about short-leg stocks is predicted to lose 1.6 followers, while the same increase for long-leg stocks predicts a similar-sized loss of 1.9 followers. In short, bearish posting is penalized regardless of the stock's net anomaly exposure. The patterns are similar across follower types: the effect is weakest among new follows from professionals and strongest among new follows from ordinary users.

Once again, our results point to a demand-based explanation for why bullish short-leg content proliferates on StockTwits and, more broadly, on investor social media. The follower channel makes this incentive especially durable, and it cuts both ways. Unlike a like, which is tied to a single post, a follower attaches to the user: gaining one expands the audience for all of a creator's future posts, while losing one shrinks it. Because bullish short-leg content attracts followers while bearish content drives them away, creators have a strong, cumulative reason both to produce bullish posts and to avoid bearish ones.

7.4 Audience Capture

The preceding results show that users disproportionately reward bullish content about short-leg stocks and that posters who supply such content gain followers. Together, these findings point to a broader audience-capture channel ("consumer-driven slant" in the economics literature), in which posters adapt their content to follower demand. If this channel is operative, influencers should respond not only to follower demand for short-leg stocks, but also to the stocks their followers are interested in more generally. We test this prediction directly by examining whether followers' stock interests in month $t-1$ predict an influencer's posting behavior in month t.

Table 9 presents the first set of results. The sample consists of influencer-stock-month observations in which stock i was discussed by followers of influencer j in month $t-1$, but *not* by the influencer. The outcome variable in column 1 equals 100 if the influencer mentions stock i in month t; in columns 2 and 3, it equals 100 if the influencer's mention is bullish or bearish, respectively. All regressions include influencer fixed effects and stock $\times$ year-month fixed effects. This specification absorbs time-invariant influencer characteristics as well as any time-varying stock-level shocks common to all influencers in a given month, isolating within-stock-month cross-influencer variation in follower discussion.

The results show that an influencer is significantly more likely to pick up coverage of a stock when that stock has been actively discussed by the influencer's followers in the preceding month, even though the influencer did not discuss it. Specifically, a stock that ranks among the top three most discussed by the influencer's followers in month $t-1$ ("Top 3 follower posts") is 0.60 pp more likely to be mentioned by the influencer in month t, a large effect relative to the baseline mean of 0.93 pp.

Influencers appear to respond to the sentiment of follower discussion, not just to the volume. When follower discussion is more bullish, the influencer is 0.07 pp more likely to mention the stock with bullish sentiment (column 2). This effect increases by 0.63 pp when the stock also ranks among the top three in follower discussion. Similarly, bearish follower discussion predicts bearish influencer mentions, although both the main effect and the interaction with "Top 3 follower posts" are smaller in magnitude.

In Appendix Table A4, we restrict the sample to stock discussions by followers whom the influencer also follows back by the end of the prior month. The results remain qualitatively similar to those in Table 9, and several estimates are larger in magnitude. Mutual following makes follower content more likely to appear in the influencer's feed, providing a natural channel through which the influencer can track follower discussion.

Table 10 provides complementary evidence using a different approach. Rather than examining which stocks followers discuss, it examines whether influencers respond to the engagement their own posts receive. This table focuses on influencer-stock-month observations for which the influencer *did* mention the stock in month $t-1$ and asks whether the number of likes the influencer's posts about that stock received in month $t-1$ predicts the influencer's behavior in month t.

It does. An influencer is more likely to post about a stock in month t when the influencer's posts about that stock received more likes in the prior month, with the relationship strongest for likes on bullish posts (column 1). The sentiment of an influencer's posts also relates to followers' liking activity in the prior month. An influencer writes more bullish posts about a stock when its bullish posts received more likes the preceding month (column 2). The same pattern holds for bearish posts (column 3). These results suggest that influencers actively monitor the like-response to their content and adjust their future posting accordingly. All regressions control for whether the influencer posted about the stock in the preceding month, so the estimated relationship is not simply driven by persistence in the influencer's own posting behavior.

Together, the evidence in this section points to a coherent mechanism. Investor social media users have a demand for bullish content about short-leg (overpriced) stocks, which they express by disproportionately liking such content and following the users who supply it. In response to these incentives, influencers track follower discussion and like responses, adjusting their future content to cater to audience demand. The result is a self-reinforcing cycle in which social media attention and positive sentiment are systematically directed toward stocks with low expected returns, potentially amplifying the mispricing of those stocks.

8. Conclusion

We study how the largest investor social media platform, StockTwits, disseminates information about anomaly stocks. Combining the 209 anomalies from Chen and Zimmermann (2022) with StockTwits data from 2010 to 2024, we find that StockTwits coverage and sentiment are systematically tilted toward the stocks the asset pricing literature suggests investors should avoid. In an average month, stocks in the bottom NET quintile – anomaly short-leg stocks – attract over five times as many posts as those in the top quintile – long-leg stocks

— and receive 472 more net bullish posts. As a result, a StockTwits user who encounters a sentiment-labeled post at random is nearly five times more likely to see a bullish post about a short-leg stock than about a long-leg stock.

Our evidence points to a demand-driven, audience-capture mechanism. Bullish posts about short-leg stocks receive more likes than those about long-leg stocks, their authors gain over three times as many new followers as those who post bullishly about long-leg stocks. Professionals and influencers receive this engagement premium for short-leg posts but, unlike other users, do not disproportionately give likes to such posts, suggesting that they are supplying what their audience demands rather than acting on private beliefs. Overall, the incentive to supply bullish content about overpriced stocks provides a micro-foundation for why StockTwits investors end up on the wrong side of anomaly portfolios and may amplify the mispricing of those stocks.

Asset pricing anomalies offer an unusually clean setting to detect whether this incentive leads to good or bad advice because expected returns provide an objective benchmark for the quality of stock recommendations. However, the "supply follows demand" mechanism demonstrated here is not unique to finance, and should operate wherever engagement incentives shape what social media users produce.

References

Allcott, Hunt, and Matthew Gentzkow, 2017, Social media and fake news in the 2016 election, *Journal of economic perspectives* 31, 211–236.

Ammann, Manuel, and Nic Schaub, 2021, Do individual investors trade on investment-related internet postings?, *Management Science* 67, 5679–5702.

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

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic, 2015, Exposure to ideologically diverse news and opinion on facebook, *Science* 348, 1130–1132.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Barber, Brad M, and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

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

Barberis, Nicholas, and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98, 2066–2100.

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

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 27, 207–264.

Chen, Andrew Y, and Tom Zimmermann, 2023, Publication bias in asset pricing research, in *Oxford Research Encyclopedia of Economics and Finance*.

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

Chen, Hailiang, Byoung-Hyoun Hwang, and Zhuozhen Peng, 2025, Why do investors like short-leg securities? evidence from a textual analysis of buy recommendations, *Review of Financial Studies* 38, 3729–3767.

Chu, Yongqiang, David Hirshleifer, and Liang Ma, 2020, The causal effect of limits to arbitrage on asset pricing anomalies, *Journal of Finance* 75, 2631–2672.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini, 2021, The echo chamber effect on social media, *Proceedings of the National Academy of Sciences* 118, e2023301118.

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

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

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

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

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, et al., 2016, The spreading of misinformation online, *Proceedings of the National Academy of Sciences* 113, 554–559.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113–2141.

Dim, Chukwuma, 2025, Social media analysts' skill: Insights from text-implied beliefs, *Journal of Financial and Quantitative Analysis* .

Engelberg, Joseph, R David McLean, and Jeffrey Pontiff, 2018, Anomalies and news, *The Journal of Finance* 73, 1971–2001.

Engelberg, Joseph, R David McLean, and Jeffrey Pontiff, 2020, Analysts and anomalies, *Journal of Accounting and Economics* 69, 101249.

Farrell, Michael, T Clifton Green, Russell Jame, and Stanimir Markov, 2022, The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145, 616–641.

Gentzkow, Matthew, and Jesse M Shapiro, 2010, What drives media slant? evidence from us daily newspapers, *Econometrica* 78, 35–71.

Han, Bing, David Hirshleifer, and Johan Walden, 2022, Social transmission bias and investor behavior, *Journal of Financial and Quantitative Analysis* 57, 390–412.

Hirshleifer, David, 2020, Presidential address: Social transmission bias in economics and finance, *Journal of Finance* 75, 1779–1831.

Jiang, Lei, Jinyu Liu, Lin Peng, and Baolian Wang, 2022, Investor attention and asset pricing anomalies, *Review of Finance* 26, 563–593.

Kogan, Shimon, Tobias J Moskowitz, and Marina Niessner, 2023, Social Media and Financial News Manipulation, *Review of Finance* 27, 1229–1268.

Kuchler, Theresa, and Johannes Stroebel, 2021, Social finance, *Annual Review of Financial Economics* 13, 37–55.

Kumar, Alok, 2009, Who gambles in the stock market?, *Journal of Finance* 64, 1889–1933.

La Porta, Rafael, 1996, Expectations and the cross-section of stock returns, *Journal of Finance* 51, 1715–1742.

List, John A, Lina M Ramirez, Julia Seither, Jaime Unda, and Beatriz H Vallejo, 2024, Critical thinking and misinformation vulnerability: Experimental evidence from colombia, *PNAS nexus* 3, pgae361.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

McLean, R David, Jeffrey Pontiff, and Christopher Reilly, 2025, Taking sides on return predictability, *Journal of Financial Economics* 173, 104158.

Mullainathan, Sendhil, and Andrei Shleifer, 2005, The market for news, *American economic review* 95, 1031–1053.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

Thaler, Michael, 2024, The fake news effect: Experimentally identifying motivated reasoning using trust in news, *American Economic Journal: Microeconomics* 16, 1–38.

Vosoughi, Soroush, Deb Roy, and Sinan Aral, 2018, The spread of true and false news online, *Science* 359, 1146–1151.

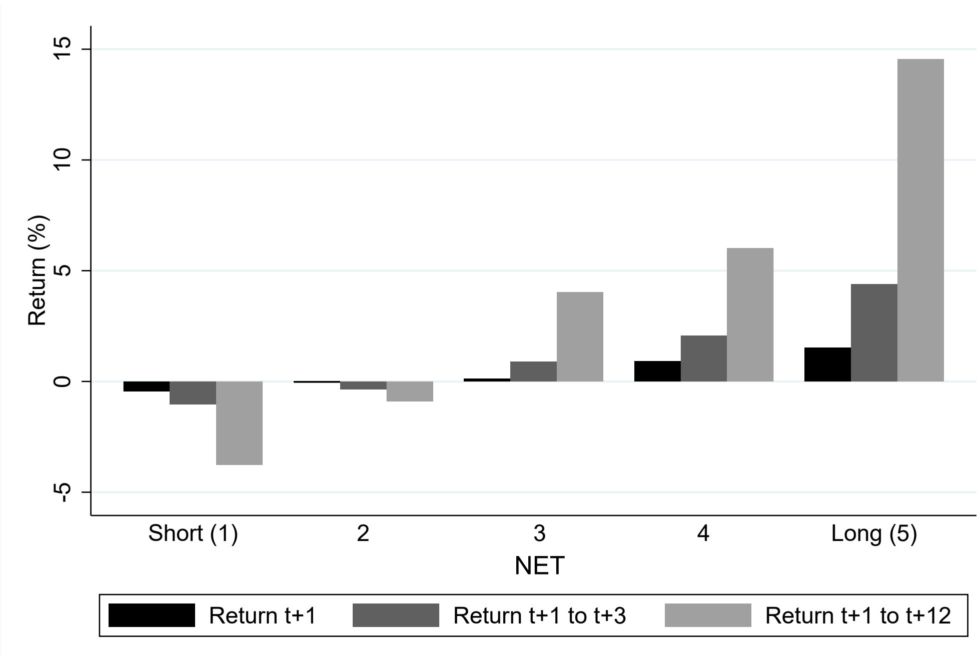
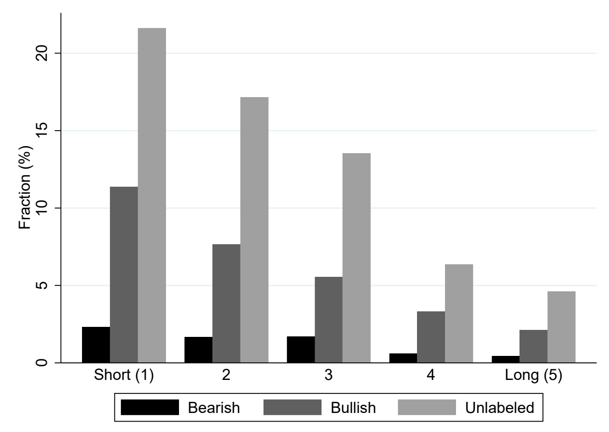
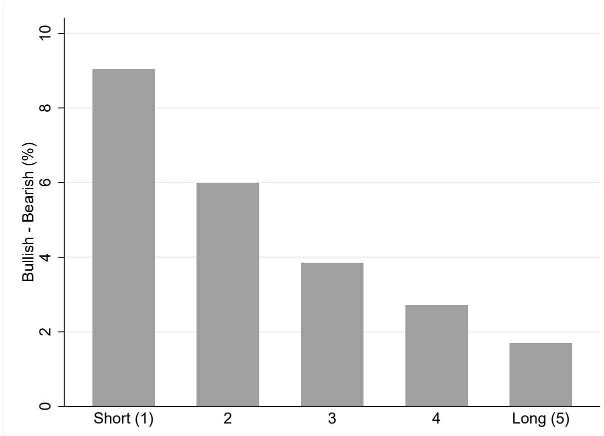


Figure 1: Returns by NET quintile portfolio

This figure plots the average returns of stocks sorted by their exposure to the long versus short legs of return anomalies in the preceding month. Each month, for each stock, we count the number of anomalies (out of 209) for which the stock falls into the top quintile of the corresponding firm characteristic (the "long leg") and the number of anomalies for which it falls into the bottom quintile (the "short leg"). We define NET as the difference between these two counts. Stocks are then ranked by NET each month and sorted into quintile portfolios. For each portfolio, we compute equal-weighted average returns over the subsequent month (t), quarter ($t+1$ to $t+3$), and year ($t+1$ to $t+12$). The sample is from August 2010 to December 2024. Appendix Figure A1 reports these averages by year and by year-month.



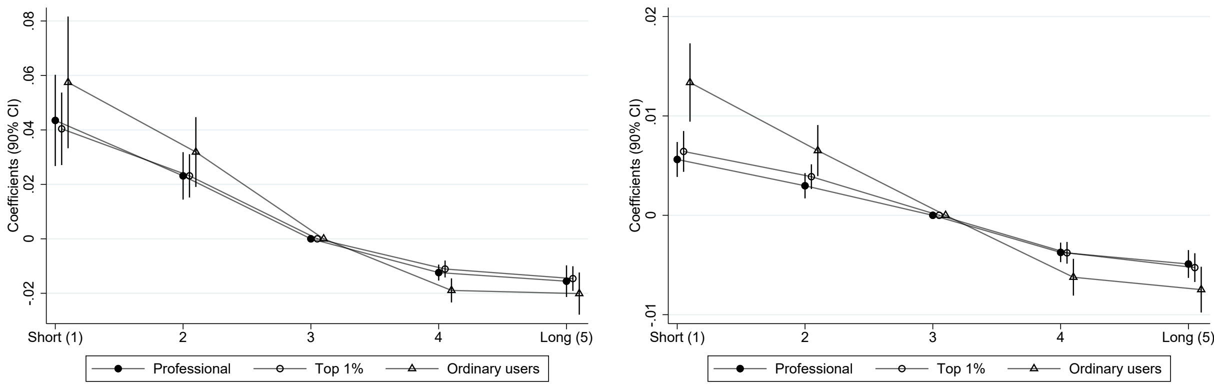


(a) % of posts

(b) % of posts: Bullish minus Bearish

Figure 2: Distribution of *Posts* by their sentiment and NET quintile

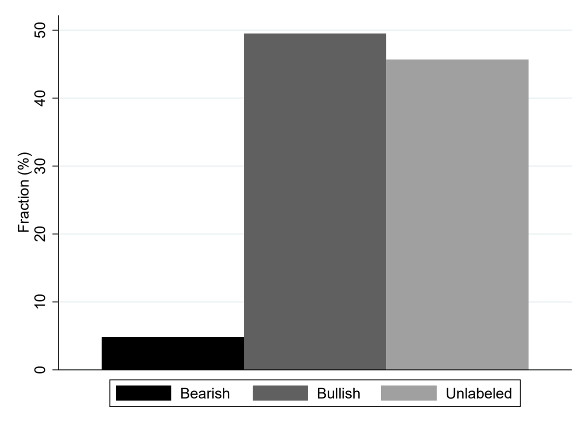
This figure plots the distribution of posts across post sentiment and the corresponding stock's net anomaly quintile (horizontal axis) in the previous month. Each month, we allocate all posts across fifteen bins based on whether they are (i) self-labeled as bullish, (ii) self-labeled as bearish, or (iii) unlabeled, and whether they pertain to stocks in NET quintile portfolios one (short-leg stocks) through five (long-leg stocks) in the previous month. Panel (a) reports the average distribution of posts across these bins. Panel (b) reports, by NET quintile portfolio, the average difference between the fraction of posts self-labeled as bullish and the fraction self-labeled as bearish. The sample period spans August 2010 to December 2024.



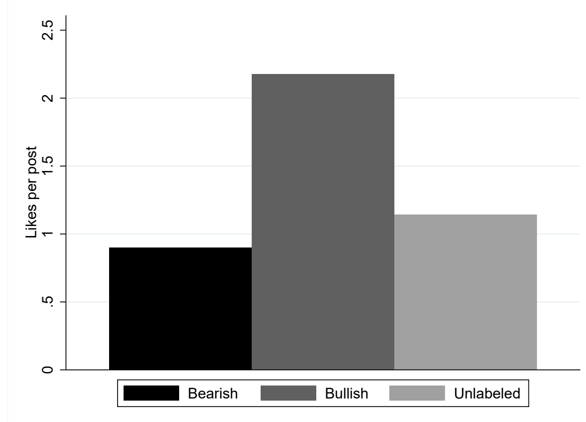
(a) % of posts **(b)** % of posts: Bullish minus Bearish

Figure 3: Which user groups *Post* about stocks in each NET quintile?

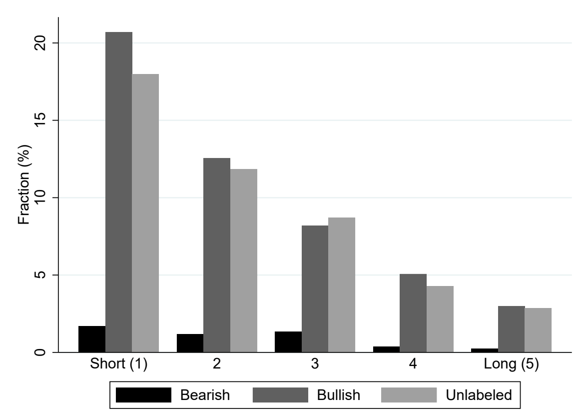
This figure plots coefficient estimates from regressions relating fractions of posts (and the difference in fraction of bullish vs. bearish posts in plot (b)) to stocks' net anomaly quintile bin in the previous month (equation 1). For each stock in each month, we compute the fraction of posts referring to that stock relative to all posts in that month and the difference in the fractions for bullish and bearish posts. We then regress these stock-month outcomes on indicator variables for whether the stock belongs to NET quintile portfolios one (short-leg stocks), two, four, or five (long-leg stocks) in the previous month; NET quintile three serves as the reference group. We estimate separate regressions for posts written by professional users (solid circles), influencers with top one percent follower counts (hollow circles), and ordinary users (hollow triangles). All specifications include stock fixed effects and year–month fixed effects. 90% confidence intervals; standard errors are double clustered at the year-month and stock levels. The sample is from September 2012 to December 2024.



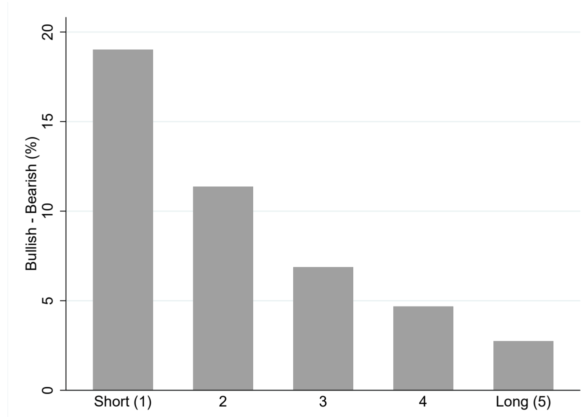
(a) % of likes



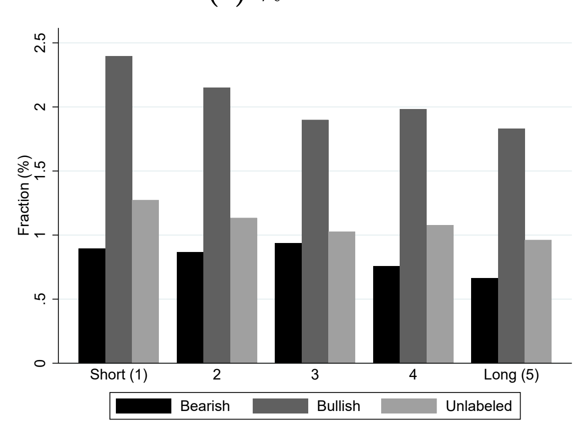
(b) # of likes per post



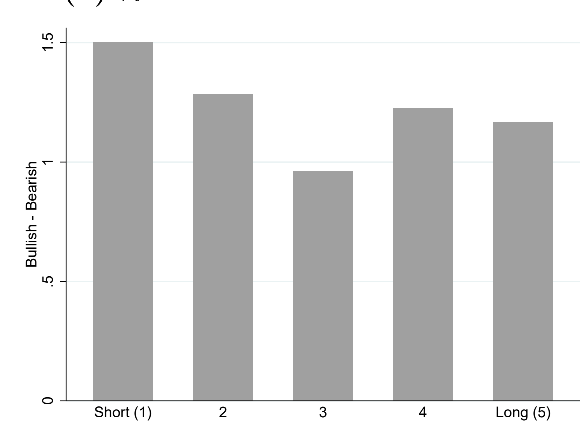
(c) % of likes



(d) % of likes: Bullish minus Bearish



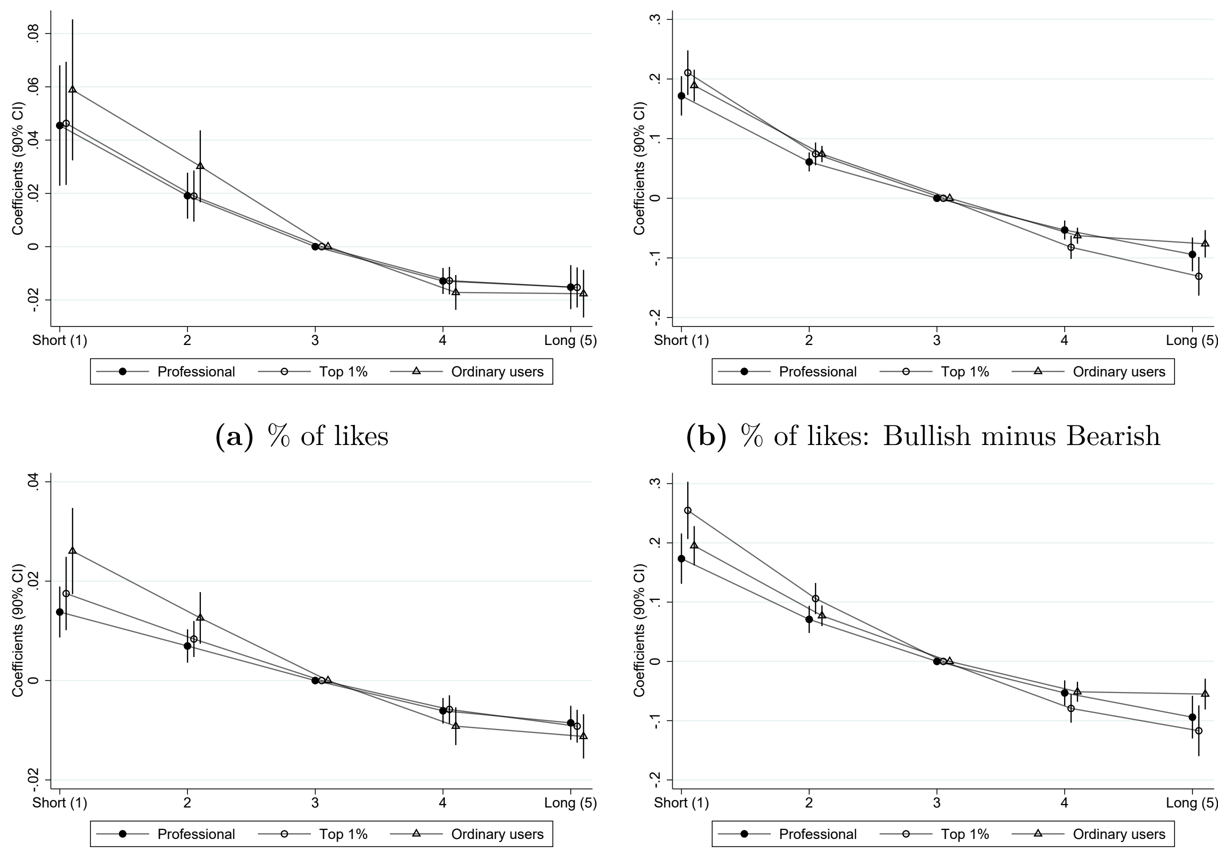
(e) # of likes per post



(f) # of likes per post: Bullish minus Bearish

Figure 4: The distribution of *Likes*

This figure plots the distribution of likes by post sentiment (panels (a) and (b)), and also by the stock's NET anomaly quintile in the previous month (panels (c) to (f)). Each month, we allocate all posts across fifteen bins based on whether they are (i) self-labeled as bullish, (ii) self-labeled as bearish, or (iii) unlabeled, and whether they pertain to stocks in NET quintile portfolios one (short-leg stocks) through five (long-leg stocks) in the previous month. Panels (a) and (b) report the fraction of likes and the average number of likes per post by post sentiment, respectively. Panel (c) reports the average distribution of likes across these bins. Panel (d) reports, by NET quintile portfolio, the average difference between the fraction of likes self-labeled as bullish and the fraction self-labeled as bearish. Panel (e) reports the average number of likes per post across these bins. Panel (f) reports, by NET quintile portfolio, the average difference between the number of likes per post self-labeled as bullish and the number of likes per post self-labeled as bearish. The sample period spans September 2012 to December 2024.



(a) % of likes

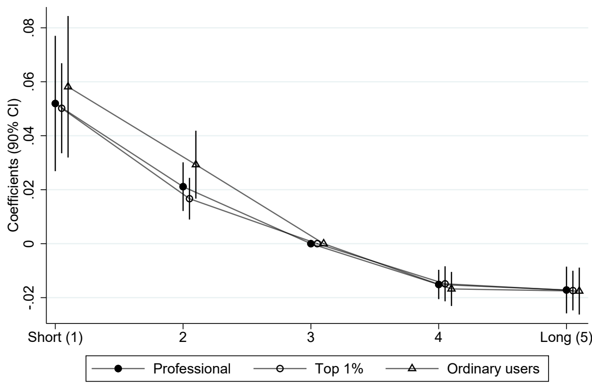
(b) % of likes: Bullish minus Bearish

(c) # of likes per post

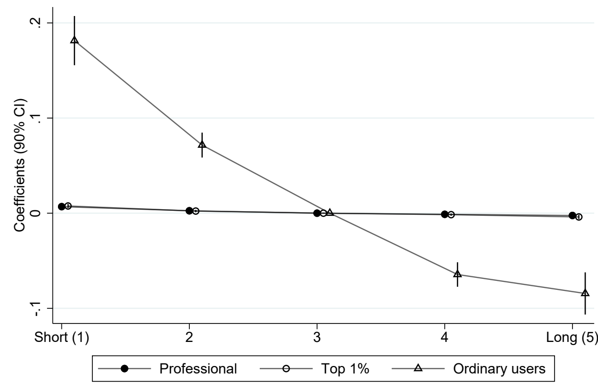
(d) # of likes per post: Bullish minus Bearish

Figure 5: Which user groups' posts *receive Likes* in each NET quintile?

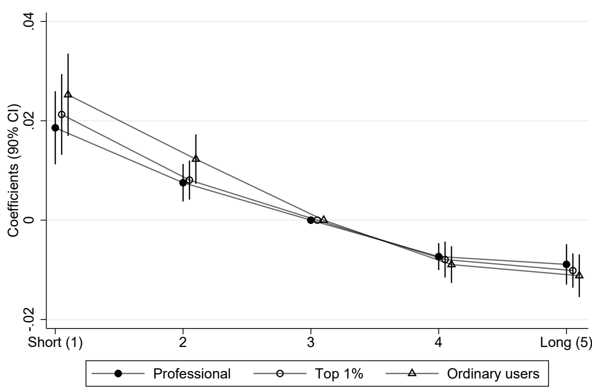
This figure plots coefficient estimates from regressions relating fractions of likes and likes per post to stocks' net anomaly quintile bin in the previous month (equation 1). For each stock in each month, we compute (i) the fraction of likes on posts about that stock relative to all likes on all posts in that month, and (ii) the average number of likes per post. We also compute these measures separately for bullish and bearish posts and take their difference. The sample includes only stock-month observations with non-zero posts. We then regress these stock-month outcomes on indicator variables for whether the stock belongs to NET quintile portfolios one (short-leg stocks), two, four, or five (long-leg stocks) in the previous month. NET quintile three serves as the reference group. We estimate separate regressions for posts written by professional users (solid circles), influencers with top one percent follower counts (hollow circles), and ordinary users (hollow triangles). Panels (a) and (c) report results for the fraction of likes and the number of likes per post. Panels (b) and (d) report results for the difference between bullish and bearish posts. All specifications include firm fixed effects and year–month fixed effects. 90% confidence intervals; standard errors are double clustered at the year-month and stock levels. The sample period spans September 2012 to December 2024.



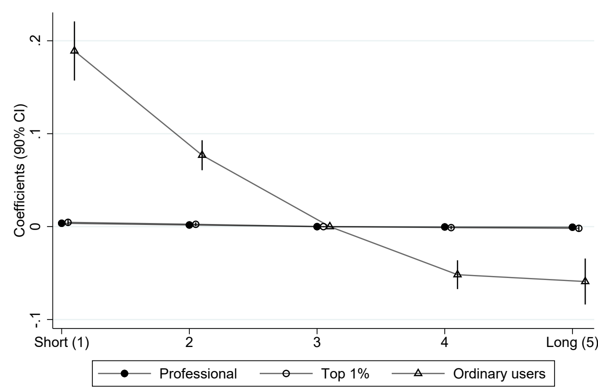
(a) % of likes



(b) % of likes: Bullish minus Bearish



(c) # of likes per post



(d) # of likes per post: Bullish minus Bearish

Figure 6: Which user groups *give Likes* to posts in each NET quintile?

This figure plots coefficient estimates from regressions relating fractions of likes and likes per post to stocks' net anomaly quintile in the previous month. For each stock in each month, we compute (i) the fraction of likes on posts about that stock relative to all likes on all posts in that month, and (ii) the average number of likes per post. We also compute these measures separately for bullish and bearish posts and take their difference. The sample includes only stock-month observations with non-zero posts. We then regress these stock-month outcomes on indicator variables for whether the stock belongs to NET quintile portfolios one (short-leg stocks), two, four, or five (long-leg stocks) in the previous month. NET quintile three serves as the reference group. We estimate separate regressions for likes by professional users (solid circles), influencers with top one percent follower counts (hollow circles), and ordinary users (hollow triangles). Panels (a) and (c) report results for the fraction of likes and the number of likes per post. Panels (b) and (d) report results for the difference between bullish and bearish posts. All specifications include firm fixed effects and year–month fixed effects. 90% confidence intervals; standard errors are double clustered at the year-month and stock levels. The sample period spans September 2012 to December 2024. Appendix Figure A4 re-plots panels (b) and (d) for influencers and professional users only so that the patterns for these user groups can be shown more clearly.

Table 1: Summary statistics at the stock-month level

	All posts		Bullish posts		Bearish posts		Bullish minus Bearish posts	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: All stocks								
# of posts	917	7,359	383	4,064	75	874	308	3,730
% of posts	0.089	0.499	0.027	0.160	0.006	0.052	0.021	0.139
# of likes	2,082	21,133	1,354	15,960	109	1,724	1,244	15,609
% of likes	0.084	0.559	0.041	0.273	0.004	0.047	0.037	0.262
# of likes per post	0.909	1.039	1.297	1.352	0.510	0.595	0.825	1.218
N stock-month (posts)					210,337			
N stock-month (likes)					176,125			
Panel B: Short-leg stocks								
# of posts	1,573	11,655	696	7,122	121	785	575	6,659
% of posts	0.142	0.666	0.046	0.238	0.009	0.050	0.036	0.211
# of likes	4,022	36,648	2,581	27,728	166	1,461	2,416	27,223
% of likes	0.153	0.825	0.079	0.411	0.006	0.043	0.072	0.394
# of likes per post	1.178	1.229	1.616	1.581	0.571	0.594	1.090	1.418
N stock-month (posts)					44,170			
N stock-month (likes)					38,853			
Panel C: Long-leg stocks								
# of posts	327	2,006	126	900	23	193	103	763
% of posts	0.035	0.198	0.010	0.070	0.002	0.017	0.008	0.057
# of likes	564	4,227	361	2,905	24	244	337	2,778
% of likes	0.028	0.239	0.014	0.119	0.001	0.012	0.013	0.112
# of likes per post	0.663	0.852	1.012	1.139	0.435	0.580	0.605	1.034
N stock-month (posts)					39,800			
N stock-month (likes)					32,123			
Panel D: Sample periods								
Posts					August 2010–December 2024			
Likes					September 2012–December 2024			

This table reports summary statistics at the stock-month level for the StockTwits sample. Panel A covers all stock-month observations, while Panels B and C cover stock-month observations for stocks in the bottom NET quintile in the prior month ("short-leg stocks") and in the top NET quintile ("long-leg stocks"). For each outcome, the table reports the mean and standard deviation for all posts, bullish posts, bearish posts, and the difference between bullish and bearish posts per stock-month. Post counts and like counts are set to zero for stock-months with no posts and no likes, respectively. Fraction of posts (likes) measures a stock's share of total posts (likes) across all stocks in a given month and are defined for months with at least one post and at least one like, respectively. Likes per post is defined for stock-months with at least one post. The sample spans August 2010 to December 2024 for posts and sentiment, and September 2012 to December 2024 for likes.

Table 2: Post volume by NET quintile

	Outcome: All posts			
	#Posts	%Posts	#Posts	%Posts
	(1)	(2)	(3)	(4)
NET $1_{i,t-1}$	623.402**	0.052**	520.636***	0.054***
	(293.622)	(0.025)	(111.756)	(0.013)
NET $2_{i,t-1}$	179.904	0.019	294.093***	0.030***
	(164.651)	(0.017)	(67.818)	(0.007)
NET $4_{i,t-1}$	-414.120**	-0.041**	-200.803***	-0.018***
	(202.525)	(0.020)	(50.614)	(0.002)
NET $5_{i,t-1}$	-621.446***	-0.054**	-312.491***	-0.019***
	(222.104)	(0.021)	(56.370)	(0.004)
Firm FE	N	N	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	916.591	0.082	916.591	0.082
Outcome SD	7,358.827	0.478	7,358.827	0.478
Observations	210,337	210,337	210,337	210,337
R^2	0.013	0.008	0.215	0.381

This table reports coefficient estimates from regressions relating posting activity to stocks' net anomaly quintile bin in the previous month (1). For each stock in each month, we compute the number of posts the stock appears in and its share of all posts in that month. NET quintile portfolio three serves as the reference group. All specifications include year-month fixed effects, and columns 3–4 additionally include firm fixed effects. Standard errors are double clustered at the year month and firm levels. The sample period spans August 2010 to December 2024.

Table 3: Post sentiment by NET quintile

	Outcome: Bullish minus Bearish posts			
	#Posts (1)	%Posts (2)	#Posts (3)	%Posts (4)
NET $1_{i,t-1}$	322.264***	0.019***	196.804***	0.012***
	(95.874)	(0.003)	(57.492)	(0.002)
NET $2_{i,t-1}$	140.396***	0.008***	118.996***	0.006***
	(42.331)	(0.002)	(37.743)	(0.001)
NET $4_{i,t-1}$	-60.589***	-0.004***	-97.981***	-0.006***
	(18.433)	(0.001)	(25.104)	(0.001)
NET $5_{i,t-1}$	-149.267***	-0.009***	-137.995***	-0.007***
	(25.511)	(0.001)	(27.169)	(0.001)
Firm FE	N	N	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	308.045	0.019	308.045	0.019
Outcome SD	3,730.374	0.133	3,730.374	0.133
Observations	210,337	210,337	210,337	210,337
R^2	0.010	0.009	0.120	0.154

This table reports coefficient estimates from regressions relating net post sentiment to stocks' net anomaly quintile bin in the previous month (1). For each stock in each month, we compute the number of posts the stock appears in and its share of all posts in that month. We do this separately for bullish and bearish posts and take the difference to obtain the outcome variable, *Bullish minus Bearish*. NET quintile portfolio three serves as the reference group. All specifications include year-month fixed effects, and columns (3)–(4) additionally include firm fixed effects. Standard errors are double clustered at the year month and firm levels. The sample period spans August 2010 to December 2024.

Table 4: Traditional news coverage and sentiment by NET quintile

	Outcomes: Corporate disclosure				Other news			
	#Articles (1)	%Articles (2)	#Bullish minus Bearish (3)	%Bullish minus Bearish (4)	#Articles (5)	%Articles (6)	#Bullish minus Bearish (7)	%Bullish minus Bearish (8)
NET $1_{i,t-1}$	0.462***	0.007***	-0.062*	-0.004***	1.919**	0.010***	-0.213***	-0.002**
	(0.093)	(0.001)	(0.038)	(0.001)	(0.846)	(0.004)	(0.072)	(0.001)
NET $2_{i,t-1}$	0.251***	0.004***	-0.041	-0.003**	1.610**	0.007**	0.002	-0.000
	(0.071)	(0.001)	(0.032)	(0.001)	(0.775)	(0.004)	(0.065)	(0.001)
NET $4_{i,t-1}$	-0.169**	-0.004***	-0.063*	-0.002	0.093	-0.002	0.163***	0.003***
	(0.070)	(0.001)	(0.033)	(0.001)	(0.720)	(0.004)	(0.062)	(0.001)
NET $5_{i,t-1}$	-0.002	-0.000	0.068	0.002	-1.177*	-0.007	0.309***	0.006***
	(0.123)	(0.002)	(0.049)	(0.002)	(0.681)	(0.005)	(0.092)	(0.001)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	5.651	0.082	0.773	0.029	14.971	0.082	1.324	0.019
Outcome SD	12.153	0.184	4.078	0.101	121.380	0.571	8.228	0.073
Observations	210,337	210,337	210,337	86,832	210,337	210,337	210,337	82,581
R^2	0.525	0.534	0.327	0.306	0.551	0.640	0.205	0.195

This table reports coefficient estimates from pooled regressions relating news coverage and sentiment to stocks' net anomaly exposure in the previous month. For each stock in each month, we compute the number of *Dow Jones Newswires* referring to that stock and the fraction of *Dow Jones Newswires* relative to all *Dow Jones Newswires* in that month. In addition, we compute these measures separately for bullish and bearish articles and take their difference. We define an article as bullish (bearish) if its RavenPack Event Sentiment Score (ESS) is above zero (below zero). Columns (1)–(4) are focused on articles relating to corporate disclosures, while columns (5)–(8) are based on other firm news. We estimate pooled regressions of these outcomes on indicator variables for whether the stock belongs to NET quintile portfolios one, two, four, or five in the previous month, with NET quintile portfolio three serving as the reference group. All specifications include year-month and firm fixed effects. Standard errors are double clustered at the year–month and firm levels. The sample period spans August 2010 to December 2024 and includes only news articles with relevance scores above seventy-five.

Table 5: Post volume and sentiment by anomaly type

	Outcome			
	All posts		Bullish minus Bearish posts	
	#Posts	%Posts	#Posts	%Posts
	(1)	(2)	(3)	(4)
Panel A: Baseline				
$\text{NET}_{i,t-1}$	-21.658***	-0.002***	-8.860***	-0.001***
	(4.094)	(0.000)	(2.165)	(0.000)
Firm FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	916.591	0.082	308.045	0.019
Outcome SD	7,358.827	0.478	3,730.374	0.133
Observations	210,337	210,337	210,337	210,337
R^2	0.215	0.381	0.120	0.154
Panel B: Decomposition by anomaly type				
Narrative/Extrapolation/Attention	-2.661	0.000	1.240	-0.000
	(5.626)	(0.000)	(3.037)	(0.000)
Fundamental/Accounting/Valuation	-6.798	-0.001	-2.952	-0.000
	(9.580)	(0.001)	(4.057)	(0.000)
Investment/Growth/Innovation	6.221	0.000	4.809**	0.000***
	(5.064)	(0.001)	(2.274)	(0.000)
Risk/Volatility/Lottery	-115.636***	-0.007***	-53.068***	-0.002***
	(26.712)	(0.001)	(15.552)	(0.000)
Liquidity/Trading Frict./Mkt. Microstructure	-91.507***	-0.011***	-39.161***	-0.003***
	(12.146)	(0.001)	(5.682)	(0.000)
Financing/Payout/Ownership	-30.412**	-0.002***	-17.222**	-0.001***
	(14.452)	(0.001)	(8.101)	(0.000)
Other	12.644	0.002	6.321	0.000
	(15.866)	(0.002)	(7.957)	(0.000)
Firm FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	916.591	0.082	308.045	0.019
Outcome SD	7,358.827	0.478	3,730.374	0.133
Observations	210,337	210,337	210,337	210,337
R^2	0.218	0.387	0.123	0.161

This table reports coefficient estimates from regressions relating posting activity and net bullishness to stocks' net anomaly exposure in the previous month. Columns (1)–(2) examine overall posting activity, measured using the number of posts and the fraction of posts relative to all posts in a given month. Columns (3)–(4) examine net bullishness, defined as the difference between bullish and bearish posting activity. Panel A reports baseline specifications using the continuous measure of net anomaly exposure. Panel B decomposes net anomaly exposure by anomaly type. Appendix Table A3 reports the mapping between individual anomalies and anomaly-type categories. All specifications include year–month and firm fixed effects. Standard errors are double clustered at the year-month and firm levels. The sample period spans August 2010 to December 2024.

Table 6: Engagement by sentiment and NET quintile: *Liking* behavior

	Outcome					
	All posts			Bullish minus Bearish posts		
	#Likes (1)	%Likes (2)	#Likes per post (3)	#Likes (4)	%Likes (5)	#Likes per post (6)
NET $1_{i,t-1}$	1492.601***	0.058***	0.196***	925.711***	0.025***	0.198***
	(369.897)	(0.015)	(0.017)	(254.037)	(0.005)	(0.020)
NET $2_{i,t-1}$	807.805***	0.028***	0.076***	540.390***	0.012***	0.081***
	(216.289)	(0.007)	(0.008)	(162.510)	(0.003)	(0.010)
NET $4_{i,t-1}$	-578.100***	-0.017***	-0.068***	-380.243***	-0.009***	-0.054***
	(145.665)	(0.004)	(0.008)	(103.916)	(0.002)	(0.010)
NET $5_{i,t-1}$	-841.340***	-0.018***	-0.091***	-516.386***	-0.011***	-0.061***
	(159.085)	(0.005)	(0.014)	(108.449)	(0.003)	(0.016)
Firm FE	Y	Y	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y	Y	Y
Outcome Mean	2,081.756	0.084	0.909	1,156.826	0.037	0.825
Outcome SD	21133.438	0.559	1.039	15052.999	0.262	1.218
Observations	189,464	176,125	176,125	189,464	176,125	176,125
R^2	0.158	0.248	0.585	0.129	0.187	0.434

This table reports coefficient estimates from regressions relating engagement on posts to stocks' net anomaly exposure in the previous month. For each stock in each month, we compute the number of likes on posts referring to that stock, the fraction of likes relative to all likes on all posts in that month, and the average number of likes per post. In addition, we compute these measures separately for bullish and bearish posts and take their difference. Like count is set to zero for stock-months with no likes; fraction of likes is defined only for stock-months with at least one like; likes per post is defined only for stock-months with at least one post. We estimate pooled regressions of these variables on indicator variables for whether the stock belongs to NET quintile portfolios one (short-leg stocks), two, four, or five (long-leg stocks) in the previous month. NET quintile portfolio three serves as the reference group. All specifications include year–month and firm fixed effects. Standard errors are double clustered at the year-month and firm levels. The sample period spans September 2012 to December 2024.

Table 7: Engagement by sentiment: *Following* behavior

	Outcome: New follows from			
	All	Influencers	Professionals	Ordinary users
	(1)	(2)	(3)	(4)
Bullish share$_{j,t-1}$	0.262***	0.018***	0.011***	0.232***
	(0.021)	(0.004)	(0.001)	(0.017)
Bearish share$_{j,t-1}$	-0.112***	-0.023***	0.004*	-0.095***
	(0.024)	(0.004)	(0.002)	(0.021)
User FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	2.553	0.270	0.150	2.142
Outcome SD	5.164	0.744	0.440	4.557
Observations	4,946,366	4,946,366	4,946,366	4,946,366
R^2	0.613	0.388	0.357	0.615

This table reports coefficient estimates from regressions relating the number of new follows a user receives in a given month to the share of bullish and bearish posts written by that user in the previous month. For each user j in each month, we compute the share of the user's posts in the prior month that are labeled bullish or bearish; the share of unlabeled posts is omitted due to multicollinearity. The sample includes only users who made at least one post in the prior month. The outcome variable is the number of new follows received in the month, with follower counts winsorized at 0.5% for both tails. Columns (2)–(4) report results using new follows from influencers, professionals, and ordinary users as the outcome variable. All specifications include year-month and user fixed effects. Standard errors are double clustered at the year–month and user levels. The sample period spans August 2010 to December 2024.

Table 8: Engagement by sentiment and NET quintile: *Following* behavior

	Outcome: New follows from			
	All users	Influencers	Professionals	Ordinary users
	(1)	(2)	(3)	(4)
Bullish × NET $1_{j,t-1}$ share	0.287***	0.022***	0.010***	0.249***
	(0.027)	(0.006)	(0.002)	(0.023)
Bullish × NET 2-$4_{j,t-1}$ share	0.255***	0.014***	0.009***	0.230***
	(0.024)	(0.005)	(0.001)	(0.019)
Bullish × NET $5_{j,t-1}$ share	0.080***	0.009	0.005*	0.064***
	(0.029)	(0.006)	(0.003)	(0.024)
Unlabeled × NET $1_{j,t-1}$ share	-0.004	-0.002	-0.004***	-0.002
	(0.013)	(0.003)	(0.001)	(0.011)
Unlabeled × NET $5_{j,t-1}$ share	-0.071***	-0.002	-0.002	-0.068***
	(0.021)	(0.003)	(0.002)	(0.018)
Bearish × NET $1_{j,t-1}$ share	-0.156***	-0.029***	-0.002	-0.130***
	(0.033)	(0.005)	(0.003)	(0.028)
Bearish × NET 2-$4_{j,t-1}$ share	-0.090***	-0.020***	0.005*	-0.075***
	(0.027)	(0.005)	(0.003)	(0.023)
Bearish × NET $5_{j,t-1}$ share	-0.194***	-0.026***	-0.002	-0.177***
	(0.049)	(0.008)	(0.004)	(0.043)
User FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	2.553	0.270	0.150	2.142
Outcome SD	5.164	0.744	0.440	4.557
Observations	4,946,366	4,946,366	4,946,366	4,946,366
R^2	0.613	0.388	0.357	0.615

This table reports coefficient estimates from regressions relating the number of new follows a user receives in a given month to the share of bullish, bearish, and unlabeled posts written by that user in the previous month, separately for stocks in NET bin one (short-leg stocks), bins two through four, and bin five (long-leg stocks). For each user j in each month, we compute the share of the user's posts in the prior month that are bullish, bearish, or unlabeled within each NET bin category. The share of unlabeled posts for stocks in NET bins two through four is omitted due to multicollinearity. The sample includes only users who made at least one post in the prior month. The outcome variable is the number of new follows received in the month, with follower counts winsorized at 0.5% on both tails. Columns (2)–(4) report results using new follows from influencers, professionals, and ordinary users. All specifications include year–month and user fixed effects. Standard errors are double clustered at the user and year–month levels. The sample period spans August 2010 to December 2024.

Table 9: Predicting Influencer stock mentions with follower interest in stocks in the prior month

	Outcome: Influencer mention$_{j,i,t}$ (%)		
	Any mention	More bullish mention	More bearish mention
	(1)	(2)	(3)
Top 3 follower posts$_{j,i,t-1}$	0.6025***	0.1916***	0.0547***
	(0.1150)	(0.0198)	(0.0114)
Follower bullish$_{j,i,t-1}$	0.2128***	0.0676***	0.0047***
	(0.0082)	(0.0033)	(0.0007)
Follower bearish$_{j,i,t-1}$	0.1172***	0.0146***	0.0202***
	(0.0122)	(0.0029)	(0.0015)
Top 3 follower posts × Follower bullish$_{j,i,t-1}$	1.6541***	0.6341***	0.1068***
	(0.1521)	(0.0558)	(0.0117)
Top 3 follower posts × Follower bearish$_{j,i,t-1}$	1.5362***	0.2445***	0.2601***
	(0.1613)	(0.0195)	(0.0193)
Influencer FE	Y	Y	Y
Firm×Ym FE	Y	Y	Y
Outcome Mean	0.928	0.262	0.061
Outcome SD	9.587	5.115	2.473
Observations	234,745,071	234,745,071	234,745,071
R^2	0.045	0.018	0.010

This table tests whether an influencer mentions in month t a stock previously discussed by their followers in month $t-1$. The sample includes influencer-stock-month observations in which stock i is mentioned by followers of influencer j, but not by the influencer, in month $t-1$. In column (1), the outcome variable equals 100 if the influencer mentions the stock in month t. In columns (2) and (3), the outcome variable equals 100 if the influencer mentions the stock with bullish or bearish sentiment, respectively. "Top 3 follower posts" equals one if stock i ranks among the three stocks most discussed by influencer j's followers in month $t-1$. "Follower bullish" equals one if follower discussion of stock i in month $t-1$ is more bullish than bearish; "Follower bearish" is defined analogously. The omitted category is the case in which follower discussion is equally bullish and bearish. All specifications include influencer and firm-by-year-month fixed effects. Standard errors are double clustered by firm and year-month. The sample period spans August 2010 to December 2024.

Table 10: Predicting Influencer stock mentions with follower likes in the prior month

	Outcome: Influencer mention$_{j,i,t}$ (%)		
	Any mention	More bullish mention	More bearish mention
	(1)	(2)	(3)
Likes to bullish posts$_{j,i,t-1}$ (%)	0.3022***	0.4116***	-0.0269***
	(0.0093)	(0.0113)	(0.0019)
Likes to bearish posts$_{j,i,t-1}$ (%)	0.0569***	-0.0294***	0.2147***
	(0.0146)	(0.0070)	(0.0359)
Likes to unlabeled posts$_{j,i,t-1}$ (%)	0.1444***	0.0338***	-0.0116***
	(0.0068)	(0.0059)	(0.0024)
Influencer post$_{j,i,t-1}$	Y	Y	Y
Influencer FE	Y	Y	Y
Firm FE	N	N	N
Ym FE	N	N	N
Firm×Ym FE	Y	Y	Y
Outcome Mean	39.707	12.653	2.896
Outcome SD	48.929	33.245	16.770
Observations	5,877,325	5,877,325	5,877,325
R^2	0.255	0.271	0.175

This table tests whether an influencer mentions in month t a stock they mentioned in month $t-1$. The sample includes influencer-stock-month observations in which stock i is mentioned by influencer j in month $t-1$. In column (1), the outcome variable equals 100 if the influencer mentions the stock again in month t. In columns (2) and (3), the outcome variable equals 100 if the influencer mentions the stock with bullish or bearish sentiment, respectively. "Likes to bullish posts (%)" is the share of influencer j's total likes in month $t-1$ that go to bullish posts about stock i, measured in percentage points. "Likes to bearish posts (%)" and "Likes to unlabeled posts (%)" are defined analogously. Shares are coded as zero if the influencer received zero likes. All specifications control for the influencer's fraction of bullish, bearish, and neutral posts about stock i in month $t-1$, as well as whether the influencer received zero likes for posts about stock i in month $t-1$. All specifications include influencer and firm-by-year-month fixed effects. Standard errors are double clustered by firm and year-month. The sample period spans September 2012 to December 2024.

Online Appendix

Giving Users What They Want: Social Media and Anomalies

Joseph Engelberg, Byoung-Hyoun Hwang, Runjing Lu, and William Mullins

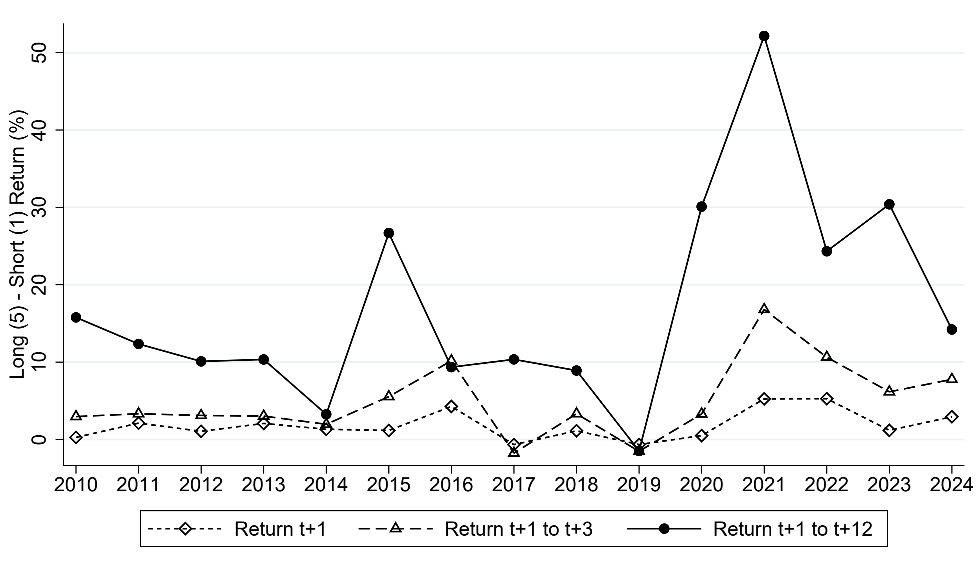
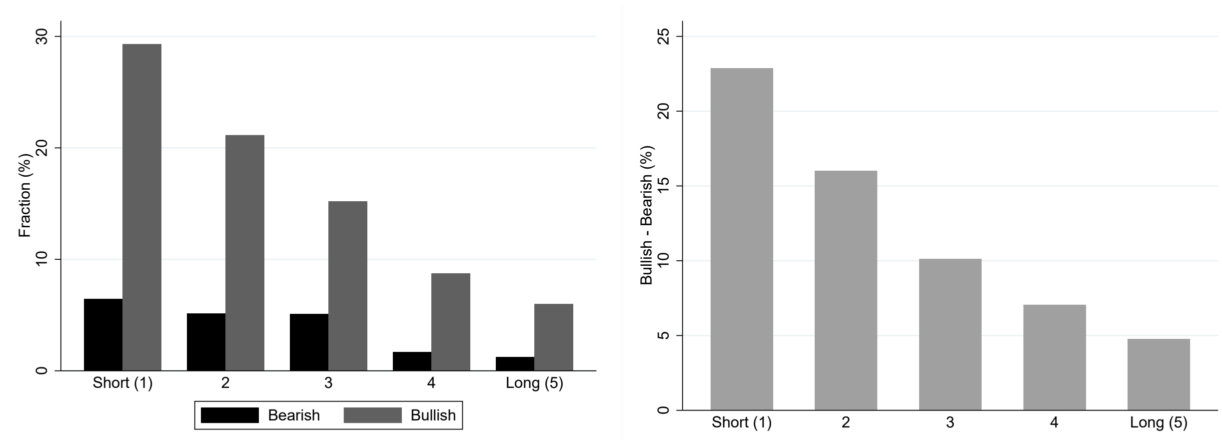


Figure A1: Returns by NET quintile *over time*

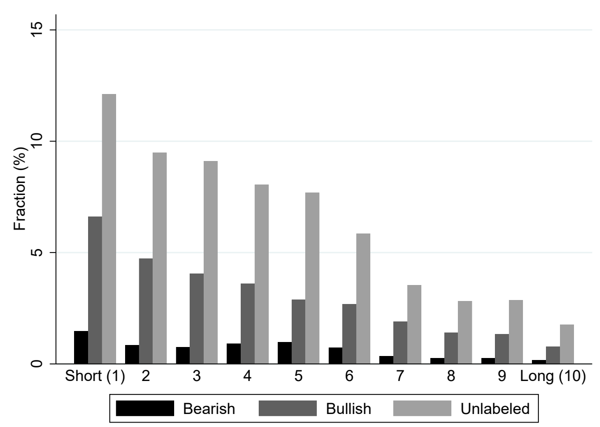
This figure plots the difference in average subsequent returns between stocks in the top and the bottom *NET* quintile by year. See note to Figure 1 for portfolio construction.



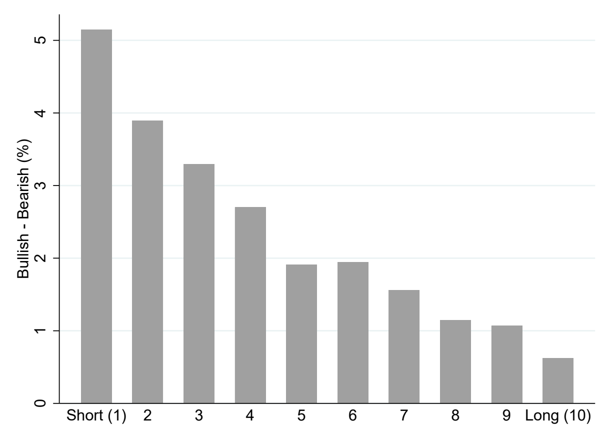
(a) % of Posts **(b)** % of Posts: Bullish minus Bearish

Figure A2: Distribution of *Posts* by their sentiment and NET quintile among posts *with self-labeled sentiment*

This figure repeats Figure 2 Panels (a) and (b) while focusing only on posts with a self-labeled sentiment. Everything else follows the original figure.



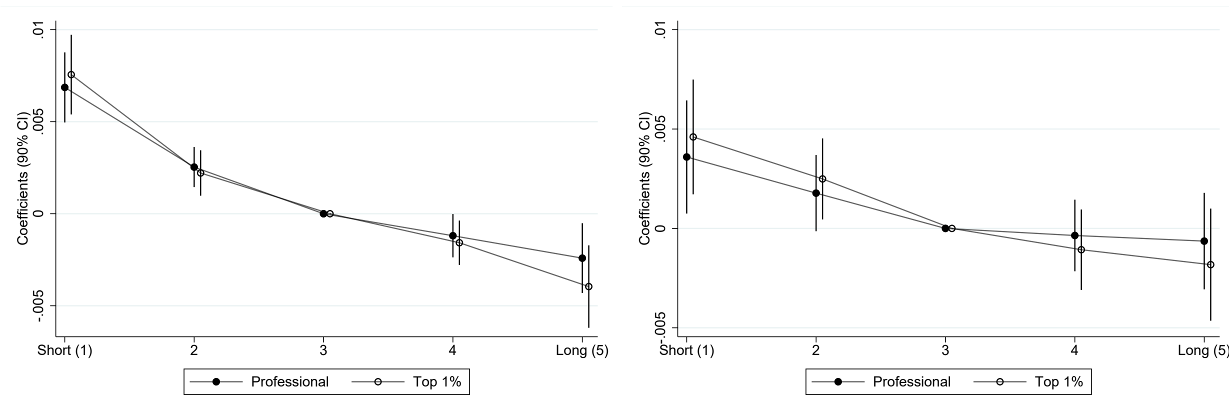
(a) % of Posts



(b) % of Posts: Bullish minus Bearish

Figure A3: Distribution of *Posts* by post sentiment and NET decile bins

This figure repeats Figure 2 using NET *decile* bins. Everything else follows Figure 2.



(a) % of likes: Bullish minus Bearish **(b)** # of likes per post: Bullish minus Bearish

Figure A4: Which user groups *give Likes* to posts in each NET quintile? Professionals and Influencers only

This figure re-plots Figure 6 panels (b) and (d) excluding the line for ordinary users. The lower range of the Y-axis allows the patterns for professionals and influencers to be seen more clearly.

Table A1: Post volume and sentiment by NET quintile
Excluding stock-months with analyst coverage

	All posts		Bullish minus bearish posts	
	#Posts	%Posts	#Posts	%Posts
	(1)	(2)	(3)	(4)
NET $1_{i,t-1}$	486.216***	0.053***	171.462***	0.012***
	(95.304)	(0.010)	(44.919)	(0.002)
NET $2_{i,t-1}$	367.826***	0.032***	149.642***	0.007***
	(93.051)	(0.007)	(49.932)	(0.002)
NET $4_{i,t-1}$	-161.630***	-0.017***	-81.725***	-0.005***
	(54.370)	(0.002)	(20.199)	(0.001)
NET $5_{i,t-1}$	-287.641***	-0.019***	-126.221***	-0.007***
	(55.241)	(0.004)	(24.222)	(0.001)
Firm FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	911.437	0.082	292.476	0.018
Outcome SD	6,634.452	0.473	2,775.083	0.114
Observations	122,713	122,713	122,713	122,713
R^2	0.305	0.458	0.144	0.168

This table reports robustness tests corresponding to columns (3)–(4) of Table 2 and columns (3)–(4) of Table 3. The sample excludes stock-month observations with analyst coverage during months $t-5$ through t. All specifications include year–month and firm fixed effects. Standard errors are double clustered by firm and year-month.

Table A2: Post volume and sentiment by NET quintile
Controlling for firm news

	All posts		Bullish minus bearish posts	
	#Posts	%Posts	#Posts	%Posts
	(1)	(2)	(3)	(4)
NET $1_{i,t-1}$	514.156***	0.054***	194.707***	0.012***
	(110.297)	(0.013)	(56.550)	(0.002)
NET $2_{i,t-1}$	290.316***	0.029***	117.762***	0.006***
	(67.625)	(0.007)	(37.617)	(0.001)
NET $4_{i,t-1}$	-194.435***	-0.017***	-95.779***	-0.006***
	(49.973)	(0.002)	(24.622)	(0.001)
NET $5_{i,t-1}$	-308.218***	-0.019***	-136.628***	-0.007***
	(55.030)	(0.004)	(26.339)	(0.001)
Firm news controls	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Ym FE	Y	Y	Y	Y
Outcome Mean	916.591	0.082	308.045	0.019
Outcome SD	7,358.827	0.478	3,730.374	0.133
Observations	210,337	210,337	210,337	210,337
R^2	0.216	0.382	0.121	0.154

This table reports robustness tests corresponding to columns (3)–(4) of Table 2 and columns (3)–(4) of Table 3. All specifications additionally control for firm news, defined as indicators for whether the firm has any 8-K filings or earnings announcements during months $t-1$ through $t+1$. All specifications include year–month and firm fixed effects. Standard errors are double clustered by firm and year-month.

Table A3: Anomaly classification by type

Type	Original economic classification	Anomaly count
Panel A: Narrative/Extrapolation/Attention		
	Earnings event	2
	Earnings forecast	8
	Info proxy	1
	Lead lag	9
	Long-term reversal	6
	Momentum	11
	Recommendation	3
	Short-term reversal	1
	Total	*41*
Panel B: Fundamental/Accounting/Valuation		
	Accruals	5
	Composite accounting	5
	Profitability	8
	Profitability alt	3
	Valuation	17
	Total	*38*
Panel C: Investment/Growth/Innovation		
	Earnings growth	4
	Investment	8
	Investment alt	10
	Investment growth	3
	R&D	5
	Sales growth	6
	Total	*36*
Panel D: Risk/Volatility/Lottery		
	Cash flow risk	1
	Default risk	1
	Leverage	4
	Option risk	4
	Risk	6
	Size	1
	Volatility	6
	Total	*23*
Panel E: Liquidity/Trading Frictions/Market Microstructure		
	Informed trading	3
	Liquidity	9
	Short-sale constraints	5
	Volume	6
	Total	*23*
Panel F: Financing/Payout/Ownership		
	Asset composition	5
	External financing	12
	Ownership	2
	Payout indicator	4
	Total	*23*
Panel G: Other		
	Other	25
	Total	*25*

This table reports the mapping between the original economic classifications of the 209 anomalies from Chen and Zimmermann (2022) and the broader anomaly types used in Table 5. The broader anomaly types were constructed using an LLM-assisted classification procedure guided by the social-transmission-bias framework in Han et al. (2022). Specifically, we provided the model with the anomaly names, original economic classifications, and the framework in Han et al. (2022), and asked it to group the anomaly classifications into economically meaningful categories. We then manually reviewed the classifications for consistency and interpretability. The resulting classification consists of seven broad types: Narrative/Extrapolation/Attention, Fundamental/Accounting/Valuation, Investment/Growth/Innovation, Risk/Volatility/Lottery, Liquidity/Trading Frictions/Market Microstructure, Financing/Payout/Ownership, and Other. For each type, the table reports the original economic classifications assigned to that type and the corresponding anomaly counts. Anomalies without clear economic content are grouped into the Other category.

Table A4: Predicting Influencer stock mentions with follower discussion in the prior month
Focusing on followers followed back by Influencers

	Outcome: Influencer mention$_{j,i,t}$ (%)		
	Any mention (1)	More bullish mention (2)	More bearish mention (3)
Top 3 follower posts$_{j,i,t-1}$	1.0177***	0.2024***	0.0276**
	(0.0958)	(0.0298)	(0.0139)
Follower bullish$_{j,i,t-1}$	0.4505***	0.1606***	0.0112***
	(0.0176)	(0.0067)	(0.0021)
Follower bearish$_{j,i,t-1}$	0.2555***	0.0365***	0.0536***
	(0.0195)	(0.0067)	(0.0046)
Top 3 follower posts × Follower bullish$_{j,i,t-1}$	1.0018***	0.5200***	0.0626***
	(0.1298)	(0.0419)	(0.0142)
Top 3 follower posts × Follower bearish$_{j,i,t-1}$	1.1086***	0.1757**	0.3168***
	(0.2044)	(0.0689)	(0.0366)
Influencer FE	Y	Y	Y
Firm×Ym FE	Y	Y	Y
Outcome Mean	1.430	0.368	0.092
Outcome SD	11.874	6.057	3.027
Observations	46,543,266	46,543,266	46,543,266
R^2	0.059	0.026	0.013

This table repeats Table 9, restricting the sample to followers who are followed back by the influencer by the end of month $t-1$. All other variable definitions and specifications follow the original table. Standard errors are double clustered by firm and year-month.