Proceedings of CONF-SEML 2026 Symposium: Multimodal Data Acquisition: Applications in Physiological and Behavioral Research
DOI: 10.54254/2755-2721/2026.AD33020

An Empirical Analysis of Firm-Level Online Investor Sentiment and Short-Term Stock Price Changes

Gengyuan Zhang

University of Florida, Gainesville, Florida, USA
gengyuanngng@gmail.com

Abstract. Online financial forums have become a popular channel through which investors express their opinions to the firm-level information. This raises a natural question: does user-generated sentiment predict short-term stock price movements? This study examines the relationship between firm-level online investor sentiment and next-day stock price changes using user comment data from a Chinese financial forum. We construct a daily sentiment measure using the proportion of negative comments and relate it to next-day stock price change. Besides, two regression models are employed for estimate if the stock price change can be predicted by daily sentiment measure. Our linear regression results show that, although the estimated sentiment effects are consistent in direction of price change, the model exhibits low explanatory power and the effects are not statistically significant. The logistic regression model achieves moderate classification accuracy in determining the direction of price change, but the ability is still limited. Overall, the results suggest that sentiment derived from online investor comments contains limited informative signals for next-day stock prices. The evidence is consistent with the interpretation that such comments primarily reflect immediate reactions or noise rather than forward-looking signals.

Keywords: Online Investor Sentiment, Stock Price Changes, Textual Analysis, Short-Term Predictability, Chinese Equity Market

1. Introduction

Online financial discussion platforms have become an increasingly important channel. Through those platforms, individual investors can express their opinions, share interpretations of firm-level events, and react to market changes. With the growth of user-generated content in financial markets, a natural question is whether sentiment expressed in online investor comments contains information that is useful for predicting short-term stock price changes. While some literature documents links between textual sentiment and asset prices, the empirical relevance of firm-level online comments for short-term price change remains under explored.

Early studies on individual investor sentiment and asset prices primarily focus on media and professionally curated content. For example, Tetlock [1]shows that negative tone in news articles is associated with short-term downward pressure on stock prices and subsequent reversal. It suggests that media sentiment may capture temporary sentiment-driven mispricing rather than fundamental information. Related work finds that sentiment embedded in news and analyst reports can affect

DOI: 10.54254/2755-2721/2026.AD33020

trading volume and volatility, particularly over short terms. These findings motivate the broader hypothesis that sentiment signals may proxy for shifts in investor attention or mood that influence prices in the short term.

More recent research extends sentiment analysis to online platforms and social media [2-4] where individual investors actively generate comments. Studies using Twitter data, stock message boards, and online forums report mixed evidence on predictability [5]. Some papers document statistically significant associations between aggregated online sentiment and market returns or volatility. However, others find that such effects are unstable, weak, or highly sensitive to sample period and model specification. In particular, user-generated comments often reflect heterogeneous beliefs, emotional reactions, or immediate responses to price changes. Hence, it raises concerns about endogeneity and noise. Whether online investor sentiment provides financially meaningful predictive power beyond contemporaneous correlation remains an open empirical question.

Motivated by this gap, this study examines the relationship between firm-level online investor sentiment and next-day stock price changes using comment data from a major Chinese financial forum. Focusing on a single listed firm allows for a direct alignment between sentiment measures and firm-level price outcomes. A daily sentiment index is constructed based on the proportion of negative comments and is aligned with one-day-ahead price outcomes. Besides, linear regression models and logistic classification models are employed, with performance evaluated using out-of-sample testing. The results indicate that firm-level online investor sentiment exhibits limited predictive power for next-day stock price changes. While the estimated sentiment effects are generally consistent in direction of stock price change, the linear models show low explanatory power and lack statistical significance. Besides, the classification models provide limited discrimination between upward and downward price changes.

The remainder of the paper is organized as follows. Section 2 describes the stock price data and online comment data and explains how the two datasets are aligned. Section 3 details the construction of the firm-level investor sentiment index based on online comments. Section 4 outlines the empirical methods used to examine the relationship between sentiment and short-term stock price changes. Section 5 presents the empirical results. Section 6 provide the final discussion.

2. Data

2.1. Stock price data

The stock price data used in this study are obtained for Xiaomi Group (1810.HK), a publicly listed firm in the Chinese equity market. The sample period spans from August 1st , 2025 to December 31st , 2025, covering all official trading days within this interval. Daily trading data include opening and closing prices, which are used to construct next-day price outcomes. Trading days are defined according to the official exchange calendar, and non-trading days are excluded from the analysis.

The primary outcome variables are defined at the daily frequency. For continuous outcomes, next-day price changes are computed using prices from consecutive trading days. For directional outcomes, a binary indicator is constructed to capture whether the stock price increases or decreases on the following trading day. All price variables are aligned to ensure that sentiment measured on day t is matched with price outcomes on day $(t+1)$. The resulting stock price dataset provides the basis for evaluating short-term price changes associated with firm-level investor sentiment.

DOI: 10.54254/2755-2721/2026.AD33020

2.2. Online comment data

Online investor comment data are collected from Eastmoney Guba ("股吧"), a major Chinese financial discussion forum that hosts firm-level message boards for listed companies. The platform allows individual investors to post comments and opinions related to specific firms, generating a large volume of user-generated content at a daily frequency.

The dataset consists of all available comments associated with Xiaomi Group (1810.HK) during the sample period from August 1, 2025 to December 31, 2025. Each comment is time-stamped, which allows comments to be aggregated at the daily level and aligned with trading days.

To ensure consistency with the stock price data, comments are aggregated to the daily frequency based on their posting dates. Trading days with no comments are retained in the sample and treated as zero-comment days. The resulting dataset forms a daily panel that aligns firm-level online investor sentiment with subsequent stock price changes.

2.3. Data alignment

The stock price and online comment datasets are merged at the daily level using trading dates. For each trading day t, investor sentiment is constructed using comments posted on that day, while price outcomes are measured on the subsequent trading day $t+1$. This alignment ensures that sentiment measures are constructed prior to the realization of the corresponding price outcomes.

The final sample consists of trading days for which both stock price data and corresponding online comment data are available. This merged dataset is used for for the empirical analysis in following sections.

3. Investor sentiment index construction

Each comment is treated as an individual textual observation and is classified into one of three sentiment categories: positive, negative, or neutral, based on its overall emotional tone. To obtain consistent sentiment labels at scale, this study uses GPT-4o as a text classifier to assign a sentiment label to each comment. To enhance reproducibility, the classification procedure is implemented with a fixed prompt and deterministic decoding settings. In particular, the model is run with a low temperature and a fixed top-p value (temperature = 0.0, top-p = 1.0), and all other generation settings are held constant across runs. This configuration minimizes stochastic variation in model outputs and helps ensure that identical or highly similar inputs receive stable sentiment labels.

Daily investor sentiment is constructed by aggregating comment-level sentiment at the daily frequency. For each trading day t, the number of comments classified as negative is divided by the total number of comments posted on that day, yielding the daily negative comment ratio.

$$\mathit{S}\text{entiment}_{\mathrm{t}} = \frac{(\text{Number of negative comments})_{\mathrm{t}}}{(\text{Total number of comments})_{\mathrm{t}}} \tag{1}$$

This ratio captures the relative intensity of negative sentiment while controlling for variation in daily comment volume. Higher values indicate a greater proportion of nagative investor opinions, while lower values reflect neutral or optimistic sentiment.The resulting sentiment time series serves as the primary representative for firm-level investor sentiment and is used in subsequent analyses to examine its association with next-day stock price changes.

DOI: 10.54254/2755-2721/2026.AD33020

4. Emprical method

This section outlines the empirical method used to examine the relationship between firm-level online investor sentiment and short-term stock price changes.

4.1. Outcome variables

The primary outcomes of interest are defined at the daily frequency. For continuous outcomes, stock price change are measured using next-day price changes constructed from consecutive trading days.

$$\Delta \mathrm{P}_{t+1} = \mathrm{P}_{\mathrm{close},t+1} - P_{\mathrm{open},\ t+1} \tag{2}$$

For directional outcomes, a binary indicator is defined to capture whether the stock price increases or decreases on the following trading day.

$$\mathrm{Y}_{t+1} = \begin{cases} 1, & \mathrm{P}_{\mathrm{close},\ t+1} > \mathrm{P}_{\mathrm{open},\ t+1} \\ 0, & otherwise \end{cases} \tag{3}$$

All outcome variables are aligned such that sentiment constructed from comments posted on trading day t is matched with price outcomes observed on trading day $t+1$.

4.2. Regression model

To examine the association between investor sentiment and next-day price changes, the following linear regression model is estimated:

$$\Delta \mathrm{P}_{t+1} = \alpha + \beta \cdot \mathrm{Sentiment}_t + \varepsilon_{t+1} \tag{4}$$

where $\Delta \mathrm{P}_{t+1}$ denotes the next-day stock price change, and $\mathrm{Sentiment}_t$ represents the firm-level investor sentiment measure constructed on day t . The coefficient β captures the average relationship between current investor sentiment and subsequent price changes.

4.3. Directional prediction model

In addition to continuous price changes, a logistic classification model is employed to assess whether investor sentiment is informative about the direction of next-day price direction changes (either increase or decrease). Here, the dependent variable is an indicator for positive next-day price changes, and the explanatory variable is the sentiment measure constructed on day t .

$$\log\left(\frac{\Pr(\mathrm{Y}_{t+1}=1\mathrm{Sentiment}_t)}{1-\Pr(\mathrm{Y}_{t+1}=1\mathrm{Sentiment}_t)}\right) = \alpha + \beta \cdot \mathrm{Sentiment}_t \tag{5}$$

This specification allows for a comparison between explanatory power in a regression setting and predictive performance in a directional forecasting context.

4.4. Model evaluation

Model performance is evaluated using an out-of-sample test. The sample is divided into training and testing subsets (70% versus 30%) based on time ordering, with models estimated on the training

DOI: 10.54254/2755-2721/2026.AD33020

sample and evaluated on the testing sample. For the regression models, explanatory power is assessed using standard goodness-of-fit measures. For the classification models, predictive performance is evaluated using classification accuracy and related metrics.

5. Empirical results

5.1. Linear regression: sentiment and next-day price change

We first examine whether daily negative investor sentiment is associated with next-day stock price changes using the linear regression model specified in Section 4.2. The dependent variable is defined as the difference between the closing price and the opening price on day $t+1$, and the explanatory variable is the daily negative comment ratio constructed on day t.

The regression model is estimated using the first 70% of the sample as training data. The estimated coefficient on the daily negative comment ratio is positive ($\beta = 0.204$), indicating that higher levels of negative sentiment are associated with higher next-day price changes. However, this coefficient is not statistically significant, with a p-value of 0.519.

The overall explanatory power of the model is extremely limited, with an R^2 value of 0.001. This means that daily negative sentiment alone explains almost none of the variation in next-day stock price changes during the sample period. Taken together, these results provide no evidence of a statistically significant linear relationship between firm-level online investor sentiment and next-day price changes.

5.2. Logistic regression: sentiment and price direction

We next investigate whether daily negative sentiment is informative about the direction of next-day stock price changes using a logistic regression framework, as described in Section 4.3. The dependent variable is a binary indicator equal to one if the stock price increases on day $t+1$, and zero otherwise. Similarly, the model is trained on the first 70% of the sample and evaluated on the remaining 30%.

The out-of-sample classification accuracy is 56.3%, which exceeds the random guessing benchmark of 50%. However, examination of the confusion matrix reveals that the model predicts only a single class. As a result, the observed accuracy reflects weak discriminative performance rather than meaningful directional prediction.

These findings indicate that although daily negative sentiment may contain limited information about price direction, it is insufficient on its own to reliably predict next-day stock price changes.

5.3. Summary of results

Overall, the results from both the linear regression and logistic classification analyses point to weak predictive ability of firm-level online investor sentiment for short-term stock price changes. While the estimated sentiment effects exhibit an interpretable sign in the linear regression, they are not statistically significant and explain little variation in prices. Similarly, the classification results show limited improvement over naive benchmarks and poor class separation.

Together, these findings suggest that sentiment extracted from online investor comments primarily reflects contemporaneous reactions or noise rather than forward-looking information relevant for short-term stock price prediction.

DOI: 10.54254/2755-2721/2026.AD33020

6. Conclusion

This study examines whether firm-level online investor sentiment extracted from user-generated comments contains predictive information for short-term stock price changes. Using comment data from a major Chinese financial forum and focusing on a single listed firm, we construct a daily sentiment measure based on the proportion of negative comments and evaluate its association with next-day price changes and price direction.

The empirical results provide limited evidence of short-term predictability. Linear regression analyses show that daily negative sentiment is not significantly related to next-day price changes and explains little variation in returns. Logistic classification models achieve modest classification accuracy but exhibit weak discriminative power, indicating that sentiment alone is insufficient for reliable directional prediction.

Taken together, these findings indicate that online investor comments primarily reflect contemporaneous reactions or noise rather than forward-looking information relevant for short-term price forecasting at the firm level. While user-generated sentiment captures aspects of investor mood, its incremental value for predicting next-day price changes appears limited in this setting.

This study contributes to the literature by providing firm-level evidence on the limits of short-term predictability using online investor sentiment. Future research may explore whether richer sentiment representations, longer forecasting time windows, or multi-firm settings can yield stronger predictive insights.

References

[1] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139–1168. https: //doi.org/10.1111/j.1540-6261.2007.01232.x

[2] Zhao, Chuanjun, Lu Kang, Xiaoxiong Xi, Shuiping Du, and Jie Li. "Investor sentiment and stock market volatility: Exploring the relationship using sentiment analysis of stock bar comments." Finance Research Open (2025): 100016.

[3] Woodward, Larry, and Jack Qu. "USING SOCIAL MEDIA COMMENTS AS A LEADING INDICATOR FOR STOCK PRICE MOVEMENT." International Journal of Accounting, Economics, and Finance Perspectives 5, no. 1 (2025): 106-120.

[4] Betzer, André, and Jan Philipp Harries. "How Online Comments affect Stock Trading-The Case of Gamestop." Forthcoming, Financial Markets and Portfolio Management (2021).

[5] Goel, R., & Mittal, A. (2011). Stock market prediction using Twitter sentiment analysis (CS229 Project). Stanford University. Retrieved from https: //cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf

Appendix

Section A: Exploratory data analysis

This appendix presents exploratory figures that summarize the time-series behavior of stock prices, trading activity, and online investor sentiment for Xiaomi Group (1810.HK) over the sample period from August 1st , 2025 to December 31st , 2025. The figures are provided for descriptive purposes only and are not used to motivate or support the empirical identification strategy in the main analysis.

DOI: 10.54254/2755-2721/2026.AD33020

1. Iinvestor sentiment and stock price

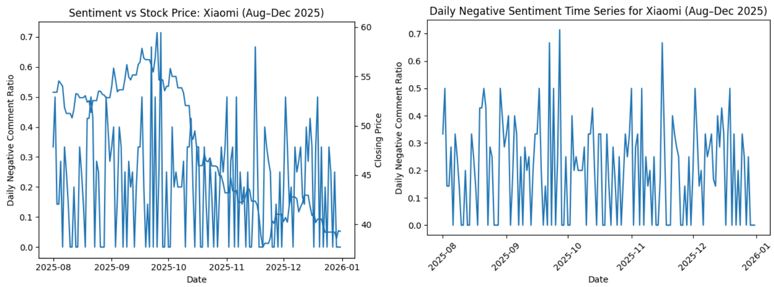

Figure A1. Investor sentiment and stock price (daily frequency)

(Figure notes: Panel (a) plots the daily negative comment ratio together with the daily closing price of Xiaomi Group (1810.HK) over the sample period from August 1, 2025 to December 31, 2025. Panel (b) shows the time series of the daily negative comment ratio. Sentiment is measured as the proportion of negative comments posted on each trading day. These figures are presented for descriptive purposes only.)

Figure A1 presents the daily evolution of online investor sentiment and stock prices for Xiaomi Group. Panel (a) plots the daily negative comment ratio together with the stock closing price, while Panel (b) shows the time series of the daily negative comment ratio alone. Investor sentiment is measured as the proportion of negative comments posted on each trading day, consistent with the construction described in Section 3.

The negative sentiment ratio exhibits noticeable variation across trading days. In several periods, the share of negative comments rises to relatively high levels, indicating temporary increases in pessimistic sentiment among individual investors. Overall, the figure provides a descriptive summary of the temporal patterns in investor sentiment and stock prices. It serves to illustrate the variability of the sentiment measure but does not imply any causal relationship between sentiment and subsequent price changes.

The sentiment series exhibits substantial day-to-day variation over the sample period. These figures provide a visual summary of the temporal patterns in investor sentiment and stock prices but do not impose or suggest any causal relationship between the two variables.

2. Stock price and trading volume

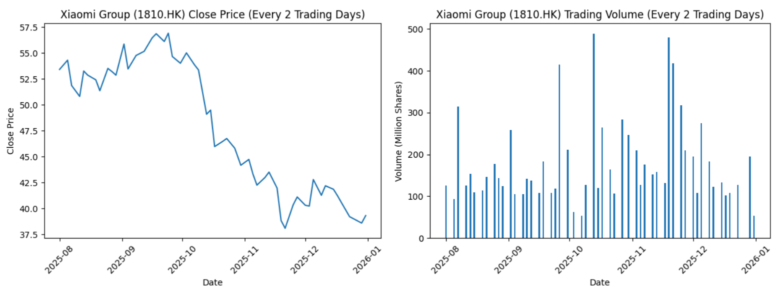

Figure A2. Stock price and trading volume

(Figure notes: Panel (a) plots the closing price of Xiaomi Group (1810.HK), and Panel (b) plots the corresponding trading volume over the sample period. To improve visual clarity, the series are downsampled to one observation every two trading days. This downsampling is applied solely for visualization purposes and does not affect any variables used in the empirical analysis.)

DOI: 10.54254/2755-2721/2026.AD33020

Figure A2 displays the time series of the stock closing price and trading volume for Xiaomi Group. To improve visual clarity, both series are downsampled to one observation every two trading days. This downsampling is applied solely for visualization purposes and does not affect any variables used in the empirical analysis.

The figures illustrate short-term fluctuations in stock prices and trading activity over the sample period. All empirical results reported in the main text are based on daily-frequency data without any temporal smoothing.