Proceedings of CONF-CDS 2026 Symposium: Machine Learning and Neural Network Applications in Engineering
DOI: 10.54254/2755-2721/2026.AST37266

The Role of Financial News Sentiment Bias in Stock Market Information Asymmetry

Yuetong Yang

School of Management, University College London, London, UK
wellington589125@gmail.com

Abstract. The growth of digital financial media has led to the faster and larger dissemination of market information while varying levels of emotionality in the news reporting from different sources can create conflicting public signals and add to information asymmetry. While there is some existing literature on the connection between aggregate news sentiment and stock returns, volatility, and trading volume, the phenomenon of sentiment bias in cross-media news reporting has not been adequately studied yet. In order to fill the gap, financial news about S&P 100 firms published during 2021–2025 is gathered from GDELT and matched against the daily market data from Yahoo Finance. Headlines are analyzed using FinBERT to find the level of their sentiment while Sentence-BERT and HDBSCAN determine whether the news item refers to the same event. Event-level median sentiment becomes a reference point while deviations from the consensus sentiment measure the cross-media bias in public sentiment. The impact of sentiment biases on the market is assessed using the Corwin-Schultz bid-ask spread, Amihud illiquidity, and two-way fixed effects regressions. It was found that greater cross-media divergence in news sentiments causes the widening of next day bid-ask spread and increased illiquidity; moreover, negative deviations have a stronger impact than positive. These correlations were consistent after controlling for return, volatility, trading volume, and news attention as well as exclusion of bias outliers.

Keywords: financial news sentiment, information asymmetry, FinBERT, event clustering, stock market liquidity

1. Introduction

Financial media has emerged as a key medium for investors in receiving and processing information regarding specific firms since digital media channels have made the dissemination of earnings releases, mergers, regulatory announcements, and corporate strategy announcements increasingly fast-paced. Linguistic framing of media tone may influence market attention and trading activity because it affects how investors assess the same event [1]. News reporting has an effect on visibility of firms and also on the incorporation of information in prices [2]. Prior studies have primarily considered aggregate tone and return, volume, price/volatility, or investor attention in connection with each other, but cross-media disagreement in terms of tone has not been directly studied before. Different outlets may cover the same news event using different levels of emotion, resulting in

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AST37266

disagreement even when facts presented are alike. This study considers cross-media bias in tone as a predictor of next-day information asymmetry. This involves the use of FinBERT sentiment classifier, semantic event clusters, and market microstructure data to distinguish between consensus of sentiment at the event level and deviation at the outlet level to see whether more deviation results in larger trading frictions and lower liquidity.

2. Literature review

2.1. Financial news sentiment signals

Financial news serves two functions, that of being a conveyor of information and an interpreter. Negative language influences the workings of the market as the investors react to the tone contained in financial reports [3]. Financial news affects the valuation of firms due to the selection and conveyance of information from the firm in question. Sentiment analysis usually classifies the articles in terms of whether they are positive, negative, or neutral. The classification process is followed by the aggregation of these sentiments based on the firm or day of trade. Aggregation of these sentiments gives the direction of sentiment, but not the level of disagreement that exists among reporters writing about the same occurrence. Event-based comparison helps to hold constant the firm event and compare emotions across sources [4].

2.2. Information asymmetry transmission

Information asymmetry results from a difference in information access or interpretation of value-related signals among market players. Media can affect this process through influencing the speed and extent of information dissemination, thus affecting price discovery and trading. Literature on market microstructure indicates that increased information asymmetries lead to wider bid-ask spreads, lower liquidity, and greater responsiveness of prices to trading [5]. Sentiment divergence between media outlets increases the effect of information asymmetry as different interpretations of the same news create different investor expectations regarding the value of the company. When the emotional frame becomes diversified, public information is still available but its interpretation is less uniform for different audiences [6].

2.3. Computational measurement gap

The development of computational financial text analysis has evolved from lexical methods of counting to the context-aware language model. The Loughran-McDonald dictionary has contributed to financial classification by identifying the presence of domain-specific meaning carried by words in corporate documents [7]. FinBERT enhances contextual approaches through sentiment probability generated through financial language instead of counting isolated words [8]. Sentence-BERT captures the semantically similar headlines in one vector space regardless of the differences in their wording. These instruments enable event-based measurement methodology that implies grouping of articles according to events followed by comparison within each event group [9]. Event clustering minimizes the risk of contamination by irrelevant news, whereas the distance between sentiment scores of each article and the event median measures the abnormal emotional framing by media outlets.

DOI: 10.54254/2755-2721/2026.AST37266

3. Experimental methods

3.1. Sample construction

The sample period includes S&P 100 firms for January 2021 to December 2025 using firm-trading day as the basic unit of analysis. Financial news from English language news sources is obtained from GDELT database in terms of headline, timestamp, source domain, URL, and company-specific entities. Companies are identified based on their name, abbreviations, and ticker symbol, while the false positives and duplicates or syndicated news are discarded. Daily market data from Yahoo Finance are obtained as opening price, highest price, lowest price, closing price, adjusted closing price, and number of trades that are used to calculate returns, volatility, dollar volume, Corwin-Schultz spreads, and Amihud illiquidity measures. News are converted to U.S. Eastern Time stamps and matched to trading days where after-hours events are assigned to the next day. Events from at least three unique domains are kept in the dataset.

3.2. Sentiment bias identification

Sentiment bias is identified in two connected stages: financial sentiment estimation and same-event comparison. Each cleaned headline is first processed with FinBERT, which returns positive, neutral, and negative class probabilities [10]. A continuous sentiment score is computed as shown in Equation (1), allowing direction and intensity to be represented on the same scale.

$$s_i = p_i^+ - p_i^- \tag{1}$$

Where s_i i is the sentiment score for headline i, while p_i^+ and p_i^- denote the FinBERT probabilities of positive and negative sentiment.

Headlines are encoded with Sentence-BERT to generate semantic vectors and clustered with HDBSCAN within a 24-hour firm-specific window. HDBSCAN groups semantically similar reports while treating weakly connected headlines as noise, allowing events to be identified without predefining the number of clusters. For each retained event cluster j, the median sentiment score across independent outlets represents the event-level consensus. Cross-media sentiment bias is then calculated as shown in Equation (2).

$$B_{ij} = |s_{ij} - \text{median}(s_j)| \tag{2}$$

Where B_{ij} is the absolute sentiment bias of report i within event j, s_{ij} is its FinBERT score, and $\text{median}(s_j)$ is the cluster-level sentiment consensus. Larger values indicate stronger emotional deviation from the prevailing interpretation of the same event. Positive and negative deviations are retained separately for asymmetric-effect tests. Cluster quality is validated using a manually reviewed event-pair sample. Report-level deviations are then averaged across qualifying event clusters for each firm-day, producing a daily absolute-bias measure that can be directly matched with next-day market outcomes.

3.3. Asymmetry effect modeling

Market information asymmetry is examined through two observable microstructure outcomes and a panel regression framework. The primary outcome is the Corwin-Schultz high-low spread estimator,

DOI: 10.54254/2755-2721/2026.AST37266

calculated from adjacent daily high and low prices to approximate bid-ask spreads when intraday quote data are unavailable [11]. Amihud illiquidity is used as an alternative outcome because it captures how strongly prices move relative to traded dollar volume. The daily measure is calculated as shown in Equation (3).

$$ILLIQ_{it} = \frac{|R_{it}|}{DVOL_{it}} \tag{3}$$

Where $ILLIQ_{it}$ is illiquidity for firm i on day t, R_{it} is the daily stock return, and R_{it} is dollar trading volume. Larger values indicate that a smaller amount of trading is associated with a larger price movement.

The effect of current news sentiment bias on next-day market conditions is then estimated using the two-way fixed-effects model shown in Equation (4).

$$Y_{i,t+1} = \alpha + \beta B_{it} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{4}$$

Where $Y_{i,t+1}$ denotes either the next-day Corwin-Schultz spread or Amihud illiquidity, B_{it} is firm-day sentiment bias, and X_{it} ontains lagged return, realized volatility, dollar trading volume, and news attention. Firm fixed effects μ_i absorb persistent company characteristics, while date fixed effects λ_t absorb market-wide shocks. Standard errors are two-way clustered by firm and date. Separate models test positive and negative sentiment deviations to compare asymmetric effects. Amihud values are rescaled for stability, while continuous controls are winsorized at the 1st and 99th percentiles. All sentiment-bias observations precede the corresponding next-day market outcomes.

4. Results

4.1. Model validation results

Model validation proves that sentiment classification and event clustering are reliable. For a sample of 2,400 hand-labeled financial headlines, FinBERT has Accuracy and Macro-F1 of 0.874 and 0.861, respectively, which outperforms the Loughran-McDonald method at 0.703 and 0.681. Event clustering was assessed based on 1,200 pairs of headlines rated by humans. Sentence-BERT along with HDBSCAN obtained Precision, Recall, and F1 score of 0.892, 0.846, and 0.868. As can be seen from Table 1, both steps have well-balanced metrics. Positive, neutral, and negative F1 scores are 0.872, 0.842, and 0.869, respectively, thus proving stability of classification for different sentiments. This makes the approach of event-level sentiment agreement and article-level divergence relevant for further study of effects on the market.

Table 1. Validation performance of sentiment classification and event clustering

Method	Task	Accuracy	Macro-F1	Precision	Recall	F1
FinBERT	Sentiment classification	0.874	0.861	0.866	0.858	0.861
Loughran-McDonald	Sentiment classification	0.703	0.681	0.690	0.673	0.681
Sentence-BERT + HDBSCAN	Event clustering	—	—	0.892	0.846	0.868

DOI: 10.54254/2755-2721/2026.AST37266

4.2. Market effect results

The market effect results indicate that increased cross-media sentiment bias is linked to larger next-day transaction costs. In the exemplary panel data of 31,842 observations, a 0.10 rise in absolute sentiment bias increases the Corwin-Schultz spread by 0.61 basis points ($p < 0.001$) and has a positive association with Amihud illiquidity ($\beta = 0.028$, $p = 0.002$). Negative bias affects spread positively by 0.83 basis points, whereas positive bias does so by only 0.39 basis points. Figure 1 indicates how the mean spreads rise from 10.8 to 17.3 basis points across sentiment-bias deciles. The association is consistent even after adjusting for the effects of returns, volatility, volume, and news attention. If one excludes the top 1% of bias observations, then the association becomes 0.55 basis points.

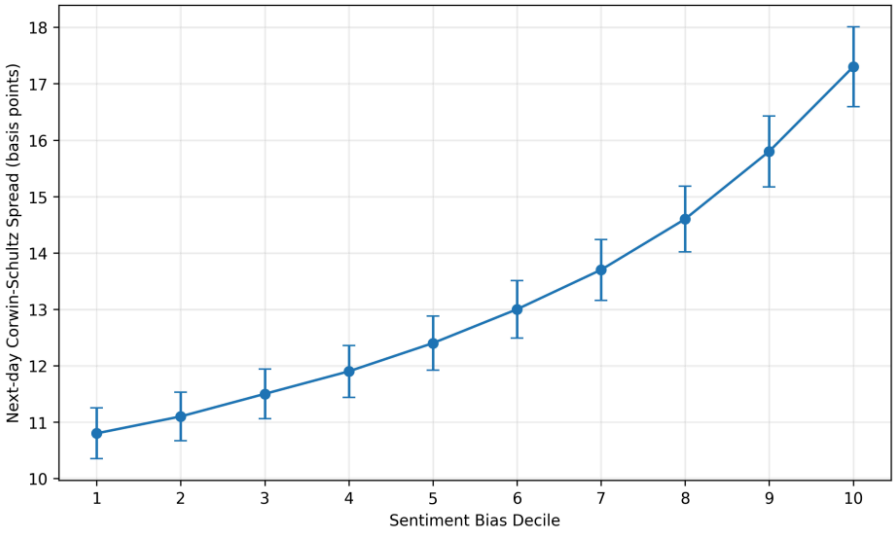

Figure 1. Next-day bid-ask spread across sentiment bias deciles

5. Discussion

The results imply that sentiment bias dispersion adds to public signal noise and causes information asymmetry. In case the sentiment bias of different media outlets is not the same, the investors' interpretation signals become inconsistent, which will lead to more valuation discrepancy and higher transaction costs. The evidence of the next day increase in spread between sentiment-bias deciles proves this theory, while the more substantial impact of negative deviation implies that pessimistic sentiment bias causes more trading friction than positive one. The positive correlation with Amihud illiquidity proves that sentiment bias influences the illiquidity in both spreads and price impact terms. It means that the results are relevant for communication system of financial market since media aggregators and recommendation algorithms can provide the investors with the repeated exposure to highly divergent or highly biased news. Thus, the sentiment bias must be considered as the additional information measure in addition to total news amount and sentiment bias average.

6. Conclusion

The analysis develops a computational methodology to investigate the connection between the bias in sentiment conveyed through financial news and the information asymmetries in the stock market. Financial news for the companies in S&P 100 obtained through GDELT from 2021 to 2025 is

DOI: 10.54254/2755-2721/2026.AST37266

combined with trading data for the companies on Yahoo Finance in order to maintain the order of events related to the time precedence between exposure and market reactions. FinBERT is employed to assess the sentiment conveyed in the headlines, whereas Sentence-BERT and HDBSCAN detect which articles discuss the same corporate events. Median sentiment at the event level constitutes the benchmark against which the bias can be assessed. The market impact is analyzed using Corwin-Schultz spread, Amihud illiquidity, and two-way fixed-effects regression accounting for returns, volatility, trading volume, and news exposure.

References

[1] Liu, J., Wu, K., & Zhou, M. (2023). News tone, investor sentiment, and liquidity premium. *International Review of Economics & Finance, 84*, 167–181.
[2] Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research, 40*(2), 806–841.
[3] Chamberlain, B., Liu, Z., & Smales, L. A. (2023). Short interest and the stock market relation with news sentiment from traditional and social media sources. *Australian Economic Papers, 62*(2), 321–334.
[4] Gan, B., Alexeev, V., & Yeung, D. (2024). Tweets versus broadsheets: Sentiment impact on stock markets around the world. *Journal of Financial Research, 47*(3), 601–633.
[5] Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters, 62*, 105227.
[6] Huang, C., Huang, H.-Y., & Ho, K.-C. (2024). Media coverage and stock liquidity: Evidence from China. *International Review of Economics & Finance, 89*, 665–682.
[7] Xia, J. (2024). Stealing the show: The negative effects of media coverage on peers' stock liquidity. *Finance Research Letters, 59*, 104691.
[8] Carlini, F., et al. (2024). Do stress and overstatement in the news affect the stock market? Evidence from COVID-19 news in The Wall Street Journal. *International Review of Financial Analysis, 93*, 103178.
[9] Koch, A., Huynh, T. L. D., & Wang, M. (2024). News sentiment and international equity markets during BREXIT period: A textual and connectedness analysis. *International Journal of Finance & Economics, 29*(1), 5–34.
[10] Groß-Klußmann, A. (2024). Learning deep news sentiment representations for macro-finance. *Digital Finance, 6*(3), 341–377.
[11] Todd, A., Bowden, J., & Moshfeghi, Y. (2024). Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions. *Intelligent Systems in Accounting, Finance and Management, 31*(1), e1549.