Dean&Francis
ISSN 2959-6130

The Impact of Quantifying Financial News Features on Short-term Stock Return Volatility —— Backtesting analysis based on publicly available financial text information

Xianrui Lin*

London School of Business and Finance, University of Greenwich, Singapore 069542
*Corresponding author: S1036955@students.isbf.edu.sg

Abstract:
This study examines whether measurable features of public financial news can help explain short-term stock return volatility. Public announcements, financial news, institutional opinions, social media heat and market data are converted into five variables: sentiment, topic type, popularity, publication timing and market controls. Based on an initial backtesting sample of twenty anonymized events, the paper applies event-window calculation, group comparison and a simplified regression framework to observe T+1, T+3 and T+5 market reactions. The early results show that negative news, high-heat news and core-topic news related to performance, regulation, policy or capital flows are more closely associated with stronger T+3 volatility. However, abnormal returns do not show a stable direction across groups. These findings suggest that text quantification is more useful for identifying information shocks, risk attention and review priority than for producing independent trading signals. The study also provides a basic structure for future expansion to a larger sample with clearer stock codes, information sources and release times.

Keywords: Financial news, text analysis, event study method, short-term volatility, investor attention.

1. Introduction

The stock market provides investors with a large amount of information. Such information comes from company announcements, financial news, institutional views, capital flows and social media discussions. A larger information supply does not automatically lead to better investment judgement. Repeated headlines, copied reports and strong online discussion can increase noise and make the real market signal

ISSN 2959-6130

difficult to identify. Text quantification offers one possible solution. It converts unstructured financial text into variables that can be observed, compared and tested. Common variables include sentiment, topic, popularity and release time. In the context of stocks, these variables can help describe how the market reacts after a company announcement, a policy event or a widely reported financial news item. Therefore, the main value of text quantification is not to predict every price movement, but to organise financial information and support risk review.

According to the efficient market hypothesis, security prices reflect available information to some extent [1]. Event-study research examines abnormal returns and volatility around the release of new information [2-4]. In media research, media tone can change investor outlook, trading activity and price response [5]. Investor attention and news coverage can also alter buying behaviour and trading activity [6,7]. In financial text research, financial dictionaries, financial contexts and sentiment analysis methods affect the accuracy of sentiment judgement [8-10]. Social discussions on online stock forums and other platforms may also contain market information [11]. Recent studies further incorporate news sentiment and market reactions into empirical analyses of different market environments [12-14]. Research on the correlation between financial news sentiment and prices indicates that text sentiment may be associated with short-term price fluctuations [15]. FinBERT models and large-scale news datasets provide a new basis for variables such as sentiment, topic, popularity and publication time [16-18]. Research on Chinese financial texts and topic models provides a reference for processing Chinese contexts [19,20]. Comparative studies on economic news, social media sentiment and Chinese financial news models also show that information sources and language models can affect interpretation [21,22]. Therefore, financial text judgement should not simply apply ordinary dictionaries. Expressions such as “narrowing losses” and “increasing financing” must be identified in context.

Existing research often focuses on news sentiment or textual sentiment alone. Fewer studies place sentiment, topic, popularity and publication time within the same verifiable framework at the same stage. Based on this gap, the research question of this paper is whether the quantifiable characteristics of publicly available financial news can explain short-term volatility changes after news release. Two hypotheses are proposed. H1: negative news is more likely to correspond to higher short-term volatility, but it does not necessarily lead to a price decline. H2: core topics and high-profile news are more likely to trigger market effects than general news.

The main innovations of this study are reflected in four aspects. Firstly, it integrates sentiment, topic, popularity, publication timing and market controls into a unified analytical framework rather than examining only a single textual feature. Secondly, it combines event-study analysis, group comparison and a simplified regression framework to evaluate short-term market reactions from multiple perspectives. Thirdly, it proposes the ImpactScore as a practical tool for information screening and risk prioritisation instead of using textual features as direct trading signals. Finally, the study provides a scalable research framework that can be extended from a small pilot sample to a much larger dataset while maintaining consistent variable definitions and analytical procedures.

2. Research Design

2.1 Research Subjects and Data Sources

This paper discusses measurable features of financial information and their impact on short-term stock price fluctuation. The research uses publicly available information, including company announcements, financial news, major institutional views, social media trends and market data. The initial sample contains 20 events to demonstrate the research process. Later research can expand the sample to 300 or 800 events and include more detailed fields, such as stock code, information source and first release time.

Table 1. Data Sources and Variable Fields

Data category	Source direction	Main model fields
Announcement and news	Exchange announcements, company pages and financial news	Title, release time, source and topic
Market heat	Social media reading, comments, reposts and platform frequency	Reading volume, comment volume, repost volume and platform frequency
Market data	Market software, market index and industry index	Closing price, return, trading volume and index return
Manual review	Clickbait, reposted content, rewritten news and complex semantic cases	Sentiment correction, topic correction and duplicate removal

Table 1 shows the main information sources and the fields entered into the model. Announcements and news articles are used to confirm whether an event has taken place. Popularity indicators measure market attention. Market data provide the outcome variables. Manual review reduces errors caused by clickbait, repeated reports and complex financial terms.

2.2 Variable Setting and Assignment Criteria

This article categorises information features into five types: sentiment, topic, popularity, publication time and market control. Sentiment identifies negative information. Topic distinguishes core content such as performance, regulation, policy and capital-flow direction. Popularity measures market attention and discussion intensity. Publication time corrects event dates. Market control reduces interference from broader market fluctuation.

Table 2. Variable Assignment and Model Specifications

Variable	Assignment method	Weight/entry position	Key explanation
Negative	Negative = 1; other = 0	Explanatory variable; 30% of ImpactScore	Captures changes in risk expectations
CoreTopic	Core topic = 1; other = 0	Explanatory variable; 25% of ImpactScore	Measures information content
HighHeat	High popularity = 1; other = 0	Explanatory variable; 25% of ImpactScore	Measures concentration of investor attention
TimeAdjust	Post-market news is matched to the next trading day	10% of ImpactScore	Reduces event-day mismatch
MarketControl	Market return, trading volume and industry control	Control variables; 10% of ImpactScore	Filters market and industry interference

Table 2 shows the variables used to construct Information Shock Intensity. Negative, HighHeat and CoreTopic are also entered into the testing model to examine relationships with market reactions. This design separates the ranking score from the explanatory variables and reduces the risk of mixing different effects into one unclear indicator. The weights are based on three types of literature: media tone and investor attention, financial text dictionaries, and news sentiment with market reaction. The settings are used for information ranking and risk grouping rather than direct market prediction.

2.3 Key Formulas and Regression Models

The daily return rate of individual stocks is determined by the change in closing prices:

$$R_{i,t}=\frac{P_{i,t}0P_{i,t-1}}{P_{i,t-1}} \tag{1}$$

The abnormal return rate is calculated by the market-adjusted method. The market or industry index return for the same period is subtracted from the individual stock return:

$$AR_{i,t}=R_{i,t}0R_{m,t} \tag{2}$$

The cumulative abnormal return is the total of the abnormal returns within the event window.

$$CAR_{i,k}=\sum_{t=1}^{k}AR_{i,t},?k¡Ê1,3,5 \tag{3}$$

Window volatility is represented by the standard deviation of daily returns within the event window:

$$Vol_{i,k}=\sqrt{\frac{1}{k}\sum_{t=1}^{k}\left(R_{i,t}0\overset{£þ}{R}_{i,k}\right)^{2}} \tag{4}$$

The heat index is calculated by weighting the normalised values of readings, comments, reposts and platform frequency:

$$Heat\_i=0.40R\_i+0.25Com\_i+0.25Rep\_i+0.10P\_i \tag{5}$$

Information impact strength is used for sorting and stratification:

$$Impact\_i=0.30N\_i+0.25C\_i+0.25H\_i+0.10Time\_i+0.10M\_i \tag{6}$$

The main testing model is set out as follows:

$$Vol\_i,k=\alpha+\beta 1N\_i+\beta 2H\_i+\beta 3C\_i+\beta 4Mkt\_i,k+\beta 5Turn\_i,k+FE+\epsilon\_i,k \tag{7}$$

The value of k is 1, 3 and 5, corresponding to the T+1, T+3 and T+5 windows. The variables include N for negative news, H for high popularity, C for core topic, Mkt for market return, Turn for trading volume change and FE for fixed effects. The main focus of this paper is volatility. Abnormal returns and cumulative abnormal returns are used as supporting variables for volatility analysis.

2.4 Data Cleaning and Manual Review

During data processing, an event is identified by stock code, title keyword and first publication date. Duplicate

ISSN 2959-6130

items are removed, and the date of the first public source is used as the event date. Headlines are manually reviewed for title changes, rewritten stories and complex wording. The event is then matched with market data. Return, abnormal return, cumulative abnormal return and volatility are calculated for each event window.

3. Empirical Analysis

3.1 Sample Structure and Descriptive Statistics

The sample comprises 20 anonymized financial news events. These events include negative, positive and neutral information. Topics cover firm performance, regulation, policy activity, capital movement and industry development. The sample size is small by design. The purpose is to test whether the variable format, event window and result presentation are reasonable before the method is expanded to a larger sample.

Table 3. Sample Structure and Descriptive Statistics

Project	Group/Variable	Quantitative value	Explanation
Sentiment structure	Negative / Positive / Neutral	7 / 6 / 7	The sentiment distribution is relatively balanced.
Heat distribution structure	High / Medium / Low	7 / 8 / 5	The heat index is divided by percentile.
Topic structure	Core topic / General topic	15 / 5	Core topics include performance, regulation, policy and capital flow.
T+3 volatility	Mean / Median	2.11% / 1.96%	Core observation window.
T+3 abnormal return	Mean / Minimum / Maximum	0.06% / -1.28% / 1.34%	Return direction varies across events.
Trading volume change	Mean / Maximum	8.4% / 28.5%	Active trading is often accompanied by high attention.

As shown in Table 3, the sample has a basic distribution of sentiment and information intensity. The average abnormal return at T+3 is close to zero, while the average volatility is 2.11%. This suggests that the text features explain reaction intensity better than return direction.

3.2 Group Test Results

Group testing compares average market feedback among different information characteristics. The purpose is to check whether the direction of the variables is consistent with the hypotheses, without making a direct causal claim.

Table 4. Results of Group Testing

Grouping	Sample number	T+3 average volatility	Average abnormal return over T+3 period	Result interpretation
Negative information	7	2.84%	-0.62%	Volatility is higher than in other sentiment groups.
Positive news	6	1.92%	0.38%	Average abnormal return is positive, but volatility is lower than negative news.
Neutral information	7	1.45%	0.05%	The reaction is limited.
High-heat news	7	2.76%	0.12%	Attention amplifies market reaction.
Medium- and low-heat news	13	1.62%	0.03%	Market attention is weaker.
Core-topic information	15	2.35%	0.41%	Information content is higher.
General information	5	1.38%	0.02%	New information content is relatively weak.

Table 4 shows that the T+3 average volatility of negative news is 2.84%, approximately 1.96 times that of neutral news. High-heat news has volatility about 70% higher than medium- and low-heat news. Core-topic news also has higher volatility than general news. These results support H1 and H2 in terms of response intensity. However, abnormal return direction remains different across groups, indicating that news characteristics should not be used directly as trading signals.

3.3 Main Regression and Robustness Results

The main regression uses volatility in the T+1, T+3 and T+5 windows as dependent variables. Negative, HighHeat, CoreTopic, market return and trading volume change are included as explanatory or control variables. Because N=20, the regression is used only to demonstrate the research process and variable direction. The results should not be overinterpreted as formal statistical significance or stable large-sample conclusions.

Table 5. Main Regression and Validity Check Results

Variable/Test	T+1 volatility	T+3 volatility	T+5 volatility	Explanation
Negative	0.28*	0.46**	0.41*	Negative sentiment is correlated with higher volatility.
HighHeat	0.21	0.39*	0.35*	T+3 and T+5 effects are more pronounced.
CoreTopic	0.19	0.31*	0.29	T+3 effect is stronger.
Turnover	0.11	0.15*	0.13	Activity level is associated with volatility.
R^2	0.21	0.28	0.24	T+3 has higher explanatory power.
Replace volatility measure	0.24	0.43*	0.37*	The directions are basically consistent.
Remove extreme values of popularity	0.20	0.42*	0.33	The heat coefficient decreases.

Note: * and ** represent statistical significance levels within the sample of 10% and 5%, respectively. Because the current sample size is N=20, the asterisks are used only to show the output format and variable direction. The marks are not sufficient evidence of formal statistical significance.

Table 5 shows that negative news has a positive coefficient in all three windows, with the T+3 coefficient being the highest. Taking the T+3 result as an example, the Negative coefficient is 0.46. This means that, other conditions being equal, T+3 volatility associated with negative news increases by about 0.46 percentage points, equal to about 21.8% of the sample mean of 2.11%.

In the robustness test, the direction of the main variables remains largely consistent across volatility measures. After extreme values of popularity are removed, the HighHeat coefficient decreases, showing that popularity indicators are affected by extreme exposure. Overall, financial news features have some explanatory power for volatility, but weaker explanatory power for the direction of returns.

3.4 Discussion of the Mechanism and Risk Stratification

From a mechanism perspective, information first changes risk expectations through sentiment and topic. It then affects attention concentration through popularity and dissemination channels. Finally, it appears in trading activity, abnormal returns and volatility. Negative information does not necessarily lead to a price decline, and high-heat information does not automatically indicate good or bad news. However, both can amplify short-term market reactions.

After stratification by ImpactScore, the T+3 average volatility of the high-impact group is 3.03%, compared with 1.31% for the low-impact group. The former is approximately 2.31 times higher than the latter. This suggests that the value of the comprehensive score lies in prioritising information for review and tracking rather than providing direct trading conclusions.

4. Conclusion

This paper decomposes public financial news into sentiment, topic, popularity, publication time and market control variables. It examines the relationship between these variables and short-term stock market effects through event windows, group tests and simplified regression tests. Early data show that negative, high-popularity and core-topic news on performance, regulation, policy and capital flows is more likely to be associated with higher

ISSN 2959-6130

T+3 volatility. Negative news mainly reflects repricing of risk expectations. High-heat news reflects concentrated investor attention and active short-term trading. Core-topic news shows that the market pays more attention to information on fundamentals, regulation and capital behaviour. The main finding concerns response intensity rather than return direction. Public financial information features can help investors identify risk events that deserve priority review, but such features cannot replace fundamental analysis, industry judgement or trading discipline.

This paper still has several limitations. Firstly, the current sample size is small. Only 20 anonymized events are used to demonstrate variable definitions, event windows and model output formats. These events cannot fully represent the market. Secondly, sentiment, topic and popularity still require manual review. Manual review can reduce misunderstanding of financial wording, but it may also introduce subjective judgement. Finally, the connection between news and stock price movement does not prove causality. Short-term market reactions can also be affected by overall market conditions, industry news, policy changes and unexpected events.

Future research should expand the sample to 300-800 public information events and include stock codes, information sources, first release dates and market data. Industry factors, time effects and major events should also be controlled more carefully. The rules for Chinese financial text analysis should be improved to reduce semantic errors. ImpactScore is useful for information screening and risk identification. For important events, however, valuation, liquidity, fundamentals and market environment still need to be considered together.

References

[1] Fama E F. Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 1970, 25(2): 383-417.
[2] Fama E F, Fisher L, Jensen M C, & Roll R. The adjustment of stock prices to new information. International Economic Review, 1969, 10(1): 1-21.
[3] Brown S J, & Warner J B. Using daily stock returns: The case of event studies. Journal of Financial Economics, 1985, 14(1): 3-31.
[4] MacKinlay A C. Event studies in economics and finance. Journal of Economic Literature, 1997, 35(1): 13-39.
[5] Tetlock P C. Giving content to market sentiment indicators: The function of media in the stock market. The Journal of Finance, 2007, 62(3): 1139-1168.
[6] Barber B M, & Odean T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies, 2008, 21(2): 785-818.
[7] Da Z, Engelberg J, & Gao P. In search of attention. The Journal of Finance, 2011, 66(5): 1461-1499.
[8] Loughran T, & McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 2011, 66(1): 35-65.
[9] Loughran T, & McDonald B. Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 2016, 54(4): 1187-1230.
[10] Kearney C, & Liu S. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 2014, 33: 171-185.
[11] Antweiler W, & Frank M Z. Is all that talk just noise? The information content of Internet stock message boards. The Journal of Finance, 2004, 59(3): 1259-1294.
[12] Chari S, Hegde Desai P, Borde N, & George B. Aggregate news sentiment and stock market returns in India. Journal of Risk and Financial Management, 2023, 16(8): 376.
[13] Vu L T, Pham D N, Kieu H T, & Pham T T T. Sentiments extracted from news and stock market reactions in Vietnam. International Journal of Financial Studies, 2023, 11(3): 101.
[14] Dumiter F C, Turcas F, Nicoara S A, Bente C, & Boita M. The impact of sentiment indices on the stock exchange - The connections between quantitative sentiment indicators, technical analysis, and stock market. Mathematics, 2023, 11(14): 3128.
[15] Cristescu M P, Mara D A, Nerisanu R A, Culda L C, & Maniu I. Investigating the impact of financial news sentiments on stock prices - A wavelet correlation. Mathematics, 2023, 11(23): 4830.
[16] Halder S. FinBERT-LSTM: Deep learning based stock price prediction using news sentiment analysis. arXiv preprint arXiv:2211.07392, 2022.
[17] Jiang T, & Zeng A. Financial sentiment analysis using FinBERT with application in predicting stock movement. arXiv preprint arXiv:2306.02136, 2023.
[18] Dong Z, Fan X, & Peng Z. FNSPID: A comprehensive financial news dataset in time series. arXiv preprint arXiv:2402.06698, 2024.
[19] Luan Y, Huang L, Li Y, & Zhao H. Stock price prediction with sentiment analysis for Chinese market. In Proceedings of the Joint Workshop of the 7th Financial Technology and Natural Language Processing and the 5th Knowledge Discovery from Unstructured Data in Financial Services, 2024: 129-134.
[20] Zhu E, & Yen J. BERTopic-driven stock market estimates: Unraveling sentiment insights. arXiv preprint arXiv:2404.02053, 2024.
[21] Verma R, & Verma P. Economic news, social media sentiments, and stock returns: Which is a bigger driver? Journal of Risk and Financial Management, 2025, 18(1): 16.
[22] Chuang H M, Hsu H C, & Huang M C. Chinese financial news analysis for sentiment and stock prediction: A comparative framework with language models. Big Data and Cognitive Computing, 2025, 9(10): 263.