

Quantifying Stock Forum Information: A Precise Analysis of Investor Sentiment, Post Characteristics, and Stock Market Volatility

Pengfan Chen

*Yunnan University of Finance and Economics, Kunming, China
1191477193@qq.com*

Abstract: This study uses data from Eastmoney Stock Bar, focusing on the Shanghai Composite Index (SSE Index) in 2024, to examine the impact of investor sentiment, post volume, and total post views on stock market volatility. Investor sentiment is quantified using the SnowNLP algorithm, and a regression model is constructed with robustness checks performed. The results reveal that investor sentiment is significantly positively correlated with the SSE Index's daily trading volume, with positive sentiment driving higher trading activity. In contrast, the volume of posts does not significantly influence trading volume, while the total number of post views shows a positive association with market volatility. This study provides valuable insights for academic research, investment strategy development, and market regulation.

Keywords: Investor Sentiment, Post Volume, Stock Market Volatility, Shanghai Composite Index, Stock Forum Data

1. Introduction

In financial markets, stock returns and volatility have long been core concerns for investors, scholars, and regulators. The prosperity of the stock market not only affects investors' wealth but also has far-reaching implications for the health of the broader economy. As a crucial benchmark of China's stock market, the Shanghai Composite Index (SSE Index) captures the attention of numerous market participants with its fluctuations.

Traditional research primarily focuses on the influence of macroeconomic indicators such as inflation rates, interest rates, and money supply on stock market volatility. For example, the Efficient Market Hypothesis (EMH) emphasizes the pivotal role of macroeconomic information in stock price formation, positing that stock prices can quickly reflect changes at the macroeconomic level [1]. Building on this framework, extensive studies have explored the relationship between macroeconomic indicators and stock market volatility [2], offering theoretical foundations and empirical evidence at the macroeconomic level. However, with the rapid development of internet technology, information dissemination methods have undergone significant transformations. Investors no longer rely solely on traditional financial media for information. Online platforms such as Eastmoney Stock Bar have emerged as critical venues for investors to exchange ideas and share market insights. The massive amount of data generated on these platforms—including posts published

by investors and the corresponding view counts—contains rich market signals, offering a new lens through which to study stock market volatility.

In this emerging field, some scholars have begun to explore the relationship between online investor behavior and stock market dynamics. For instance, analyses of public sentiment on social media platforms have revealed a certain degree of correlation with market movements [3]. In China, text mining studies have confirmed that investor sentiment—particularly that induced by events like the COVID-19 pandemic—plays a role in stock market volatility beyond economic fundamentals [4]. Nevertheless, in-depth studies focusing specifically on specialized investment forums like Eastmoney Stock Bar remain scarce. Furthermore, a comprehensive understanding of how multiple factors—such as investor sentiment, post volume, and post views—jointly influence stock market volatility is still lacking.

Posts published by investors on stock forums not only reflect individual sentiment but also reveal the level of market participation and activity. Investor sentiment, as a subjective factor, has been recognized as a significant non-economic driver of stock market volatility. Meanwhile, post volume and total views capture different dimensions of market engagement and investor attention. Therefore, a deep exploration of the intrinsic connections between investor sentiment, post activity, and stock market volatility is vital for accurately understanding the stock market's operational mechanisms, formulating effective investment strategies, and implementing sound regulatory policies.

This study aims to bridge the current research gap by using data from Eastmoney Stock Bar, with the 2024 Shanghai Composite Index as the research sample, to investigate the effects of investor sentiment, post volume, and total post views on stock market volatility. It also seeks to uncover the underlying mechanisms from the perspectives of market activity and investor behavior. Through the construction of a regression model and rigorous robustness testing, this research aspires to offer reliable empirical findings. Ultimately, the study enriches the literature on financial market volatility and investor behavior and provides valuable references for both investors and regulatory authorities.

2. Hypothesis development

2.1. Investor sentiment (mood) and stock market volatility

According to behavioral finance theory, investors are not fully rational, and their emotions can significantly influence their investment decisions [5]. Numerous studies have shown that investor sentiment plays a crucial role in the stock market. For instance, research has found that investor sentiment affects the demand and supply for stocks, thereby influencing stock prices [6]. When investor sentiment is positive, investors tend to be more confident about market prospects, are more willing to take on risk, and consequently increase their stock purchases. This heightened demand drives stock prices upward and raises trading volumes, ultimately leading to positive stock market volatility. Conversely, negative sentiment fosters pessimism among investors, leading them to sell stocks, reducing capital inflow into the market, causing stock prices to fall and trading volumes to shrink, and thus triggering negative market volatility. Based on the above analysis, the following hypothesis is proposed:

H1: Investor sentiment is positively correlated with stock market returns; positive sentiment drives upward market volatility.

2.2. Post volume (issue) and stock market volatility

Post volume represents the level of investor activity within the market. A greater number of posts indicates higher investor participation and more frequent information flow in the market [7]. On one hand, an abundance of posts can lead to increased information disclosure, including updates on corporate fundamentals, industry developments, and market rumors. This influx of information can

influence investors' assessments of stock value, thereby altering their investment decisions, increasing market activity, and intensifying volatility [8]. On the other hand, a rise in post volume may trigger imitation behavior among investors, leading to herding effects. When a large number of investors post bullish or bearish opinions, others may follow their lead, further amplifying market volatility [9]. Therefore, the following hypothesis is proposed:

H2: Post volume is positively correlated with stock market returns; the greater the number of posts, the more pronounced the market volatility.

2.3. Total post views (skim) and stock market volatility

Total post views reflect the degree of market attention to specific information. A high number of views suggests that the information has garnered widespread attention, which can rapidly disseminate among investors and trigger collective reactions [10]. If highly viewed posts convey positive news, such as favorable corporate announcements, they may attract more investors to buy stocks, driving the market upward. Conversely, if the posts convey negative information, such as pessimistic industry forecasts, they may prompt investors to sell stocks, leading to a market downturn. Furthermore, changes in the number of post views can also reflect shifts in market attention cycles. Sustained increases or decreases in attention to certain types of information can concentrate trading behavior, thereby intensifying market volatility [11]. Consequently, the following hypothesis is proposed:

H3: Total post views are positively correlated with stock market returns; the higher the total post views, the greater the stock market returns.

3. Research design

3.1. Data source and sample selection

This study uses the Shanghai Composite Index (SSE Index) in 2024 as the research sample. Data were collected by web scraping from Eastmoney Stock Bar and processed according to the following principles: (1) Posts were selected exclusively from the SSE Index forum to eliminate interference from discussions related to other stocks; (2) Data from non-trading days (weekends and public holidays) were excluded; (3) Samples with missing or abnormal data were removed. After data cleaning, the final dataset contains trading information for 226 trading days, comprising a total of 1,130 data points.

3.2. Variable definitions

3.2.1. Dependent variable — daily trading volume of the SSE index (trade)

Trading volume represents the scale of stock transactions in the market and directly reflects the inflow and outflow of capital in the stock market. In this study, the daily trading volume of the SSE Index, obtained from Eastmoney Stock Bar, is used to indicate the overall market condition. A higher trading volume reflects a healthier trading environment and a more prosperous stock market.

3.2.2. Independent variables

(1) Investor Sentiment (mood): Posts from the SSE Index Stock Bar were scraped and analyzed to assess the sentiment of their authors. The SnowNLP algorithm was applied to classify posts into negative, positive, and neutral categories, ultimately producing an aggregated sentiment score. (2) Post Volume (issue): The number of posts represents the level of market participation by investors and to some extent reflects market activity. (3) Total Post Views (skim): The total number of views

measures the degree of investor attention toward the market and serves as a key indicator of market attention and information dissemination breadth.

3.2.3. Control variables

To isolate the effect of the independent variables on trading volume, this study controls for factors that may influence the SSE Index trading volume, including the inflation rate, interest rates, money supply (M2), financial sector index performance, market volatility, and the performance of other major indices.

3.3. Model specification

To test the proposed hypotheses, the following model is constructed:

$$\text{trade}_i = \beta_0 + \beta_1 * \text{mood}_i + \beta_2 * \text{issue}_i + \beta_3 * \text{skim}_i + \varepsilon$$

This model is based on the classical theory of multiple linear regression [12], which has been widely applied in financial research to analyze the relationship between multiple independent variables and a dependent variable. In the context of this study, it effectively integrates several key variables to quantify their association with stock market volatility.

The rules for interpreting the regression results are as follows: If coefficient β_1 is significant, it indicates that investor sentiment influences stock market returns. If coefficient β_2 is significant, it indicates that post volume affects stock market returns. If coefficient β_3 is significant, it indicates that total post views impact stock market returns. If a coefficient β is positive, it implies a positive impact; if negative, it implies a negative impact.

4. Empirical results

4.1. Descriptive statistics

After data collection and model construction, descriptive statistics were conducted on the 168 valid sample observations to gain an initial understanding of the basic characteristics of each variable. The results are presented in Table 1. Descriptive statistics provide insights into central tendencies, dispersion, and data distribution, laying a foundation for subsequent correlation and regression analyses.

The daily trading volume of the SSE Index (trade) has a maximum of 1.51e+12 and a minimum of 2.07e+11, with a mean of 4.48e+11. This indicates that, overall, the SSE Index maintained a relatively high level of trading activity during the sample period. Although the range is wide, the mean reflects a moderately active trading environment on average. The total post views (skim) range from 500,547 to 3,801,918, with a mean of 1,245,085. This suggests that posts received substantial attention on average each trading day, indicating a certain level of investor interest in Stock Bar information. The investor sentiment (mood) variable ranges from -2,442 to 772, with a mean of -98.67857 and a standard deviation of 390.304. The negative mean indicates that investor sentiment was slightly pessimistic overall, and the relatively large standard deviation reflects substantial sentiment fluctuation. The post volume (issue) ranges from 1,281 to 15,357, with a mean of 3,936.03, indicating that a significant number of posts were published daily, reflecting a moderate level of investor engagement and discussion activity.

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. dev.	Min	Max
trade	168	4.48e+11	2.40e+11	2.07e+11	1.51e+12
skim	168	1245085	603426.4	500547	3801918
mood	168	-98.67857	390.304	-2442	772
issue	168	3936.03	2202.496	1281	15357

4.2. Correlation analysis

Following the descriptive analysis, correlation analysis was conducted to examine the linear relationships between investor sentiment (mood), post volume (issue), total post views (skim), and the daily trading volume of the SSE Index (trade). The results are shown in Table 2.

The correlation coefficient between trade and skim is 0.6649, indicating a strong and significant positive correlation—higher post views are associated with greater daily trading volume. The correlation between trade and mood is 0.2041, suggesting a weaker positive correlation, implying that while sentiment does influence trading volume, it is not as impactful as total post views. The correlation between trade and issue is 0.4569, indicating a moderate positive relationship, though this may be affected by other confounding factors.

Table 2: Correlation matrix

Variable	trade	skim	mood	issue
trade	1.0000			
skim	0.6649	1.0000		
mood	0.2041	-0.1494	1.0000	
issue	0.4569	0.7841	-0.4066	1.0000

4.3. Regression analysis results

The regression results based on the proposed model are summarized in Table 3 below:

Table 3: Regression results

VARIABLES	(1) trade	(2) trade
mood	2.063e+08*** (37431603.397)	1.672e+06 (9.097e+06)
issue	9233031.106 (10567648.816)	240,717*** (31,259)
skim	258,069.899*** (35,640.495)	2.403e+08*** (3.322e+07)
Constant	1.104e+11*** (2.952e+10)	1.528e+11*** (2.478e+10)
Observations	168	226
R-squared	0.538	0.471

4.3.1. Investor sentiment (mood)

The coefficient is positive and highly significant at the 1% level, confirming a strong positive relationship with trading volume. Therefore, Hypothesis 1 is supported. All else equal, a one-unit increase in investor sentiment leads to an estimated increase of $2.063e+08$ in daily trading volume. From a behavioral finance perspective, heightened investor sentiment tends to fuel optimism and trigger more aggressive investment behavior, which resonates within online forums and stimulates buying activity. Market participants may use investor sentiment on Stock Bar as a market trend indicator, while regulators should monitor extreme sentiment fluctuations to prevent market instability.

4.3.2. Post volume (issue)

The coefficient (9,233,031.11) is not statistically significant, suggesting that post volume does not have a clear or direct impact on trading volume. While post volume may reflect investor engagement, it may not translate into actual trading behavior due to potential multicollinearity or other confounding factors. Thus, Hypothesis 2 is not supported. Investors should not rely solely on post volume to gauge market trends and activity but should consider multiple dimensions.

4.3.3. Total post views (skim)

The coefficient is 258,069.90 and is highly significant at the 1% level, indicating a substantial and positive effect on trading volume. Therefore, Hypothesis 3 is supported. High view counts reflect strong market attention and may trigger a “herding effect,” attracting more potential investors and increasing both market participation and trading activity. Financial institutions and information providers can monitor post views as a proxy for market sentiment, and traders may adjust their strategies accordingly.

4.4. Robustness check

To verify the robustness of the regression results, this study adopts a sample expansion method. While the previous regression used data from April to October 2024 (168 observations), the updated model incorporates data from January to October 2024, increasing the sample size to 226 observations. The new regression results are shown in Table 3 (Column 2).

The updated results remain consistent: The coefficient for post volume remains insignificant. The coefficient for total post views is 240,717, significant at the 1% level, confirming its positive effect on trading volume. The coefficient for investor sentiment is $2.403e+08$, also significant at the 1% level, indicating a strong positive correlation with trading volume. These findings confirm the robustness of the study's conclusions and the explanatory power of the regression model.

5. Research conclusions and implications

This study employs data from Eastmoney Stock Bar and constructs a regression model to investigate the effects of investor sentiment, post volume, and total post views on stock market volatility, offering a unique perspective for analyzing fluctuations in the stock market. The results reveal a strong positive correlation between investor sentiment and the daily trading volume of the SSE Index, indicating that sentiment fluctuations serve as a critical driver of trading activity. A well-constructed quantitative sentiment index system can effectively capture market sentiment trends and provide a solid foundation for investor decision-making. However, extreme fluctuations in sentiment can significantly impact market stability. The study also finds that post volume does not have a significant effect on daily trading volume, possibly because post activity does not directly translate into trading

behavior, or due to multicollinearity with other variables. In contrast, total post views exert a significant positive influence on trading volume. High view counts can trigger herding behavior, thereby increasing market activity and participation. Moreover, different types of investors respond differently to post view metrics—long-term investors tend to focus on fundamental analysis, while short-term speculators are more sensitive to market trends and sentiment signals.

Based on these findings, the following recommendations are proposed: For investors: It is essential to make full use of two key indicators—investor sentiment and total post views. By incorporating a quantitative sentiment index and closely monitoring changes in post views in real time, investors can develop more rational and effective trading strategies while avoiding blind herd-following. At the same time, investors should pay close attention to expert analysis and insights published by financial institutions to enhance their understanding and interpretation of market information, leading to more informed investment decisions. For regulatory authorities: Close monitoring of extreme fluctuations in investor sentiment is necessary, and effective intervention mechanisms should be established. In periods of excessive market optimism, which may lead to asset bubbles, regulators should strengthen oversight of information disclosure, ensuring that listed companies publish accurate and timely information, thereby maintaining orderly and stable market operations. For financial institutions and information providers: Greater attention should be paid to the market value embedded in total post views. Institutions should actively guide market sentiment, optimize information dissemination strategies, and offer personalized services and recommendations tailored to different investor profiles and behavioral preferences.

References

- [1] Fama E F. *Efficient capital markets: a review of theory and empirical work*[J]. *Journal of Finance*, 1970, 25(2): 383 - 417.
- [2] Chen, N., Roll, R. and Ross, S.A. (1986) *Economic Forces and the Stock Market*. *Journal of Business*, 59, 383-403.
- [3] Bollen, J., Mao, H., & Zheng, X. (2011). *Twitter mood predicts the stock market*. *Journal of Computational Science*, 2 (1), 1 - 8.
- [4] Cao, X. F., Yang, S. H., & Fan, W. B. (2023). *The impact of pandemic-related emotions on the stock market based on text mining*. *Financial Theory and Teaching*, 5(5), 44–52.
- [5] Shefrin H. *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing* [M]. Harvard Business School Press, 2000.
- [6] Baker M, Wurgler J. *Investor sentiment and the cross-section of stock returns* [J]. *Journal of Finance*, 2006, 61 (4): 1645-1680.
- [7] Dye R A. *Proprietary information, trading, and efficiency* [J]. *Journal of Finance*, 1986, 41 (1): 103-120.
- [8] Hong H, Stein J C. *A unified theory of underreaction, momentum trading, and overreaction in asset markets* [J]. *Journal of Finance*, 1999, 54 (6): 2143-2184.
- [9] Banerjee A V. *A simple model of herd behavior* [J]. *The Quarterly Journal of Economics*, 1992, 107 (3): 797-817.
- [10] Shiller R J. *Irrational Exuberance* [M]. Princeton University Press, 2000.
- [11] Kahneman D, Tversky A. *Prospect theory: An analysis of decision under risk* [J]. *Econometrica*, 1979, 47 (2): 263-291.
- [12] Gujarati, D. N., & Porter, D. C. (2008). *Basic Econometrics* (5th ed.). McGraw Hill.