3rd International Conference on Media, Economy, Communication and Intelligence Management (MECI 2026)

Article

Construction of Investor Sentiment Factor from Social Media Big Data and Its Spillover Effects on Stock Market Volatility

Quanrong Bao [1,*]

[1] School of Finance, Shandong University of Finance and Economics, Jinan, China
* Correspondence: Quanrong Bao, School of Finance, Shandong University of Finance and Economics, Jinan, China

Abstract: This study investigates the complex association between social-media-based investor sentiment and stock market volatility, with a specific focus on the mediating role of trading activity in contemporary financial markets. As digital platforms increasingly influence financial decision-making, understanding these dynamics is crucial. Using a comprehensive dataset of user posts extracted from the East Money Stock Forum for selected major CSI 300 constituent stocks, we systematically construct robust investor sentiment measures and a composite sentiment index. These metrics are subsequently matched with empirical market turnover and volatility indicators to capture real-time market reactions. The empirical results demonstrate that while the aggregate sentiment direction exhibits only a weak direct association with market volatility, sentiment intensity is significantly and positively related to volatility spikes. Furthermore, comprehensive mediation analysis suggests that the market turnover rate serves as an essential transmission channel linking digital sentiment to observed volatility. To provide a deeper understanding, quantile regression analysis further indicates that this relationship exhibits significant heterogeneity, varying considerably across different market volatility conditions, although the overall economic magnitude of this effect remains relatively limited in scope. Ultimately, these findings suggest that social media sentiment can provide valuable supplementary information for understanding and forecasting market fluctuations, especially when it is considered in conjunction with underlying trading activity and broader behavioral finance paradigms.

Keywords: investor sentiment; social media; sentiment analysis; market volatility; behavioral finance

Received: 23 March 2026
Revised: 14 May 2026
Accepted: 28 May 2026
Published: 03 June 2026

1. Introduction

Stock market volatility is a key indicator of market uncertainty and risk, and it has long been regarded as an important barometer of financial market conditions. Under the traditional finance framework, represented by the Efficient Market Hypothesis, asset prices are assumed to fully reflect available information and investors are assumed to behave rationally [1]. However, persistent market fluctuations, excessive reactions, and other anomalous phenomena are difficult to explain solely within this framework. Behavioral finance provides an alternative perspective by emphasizing that investors are often influenced by psychological biases and emotional factors when making decisions under uncertainty. In this context, investor sentiment has become an important lens through which market dynamics can be understood.

In China's A-share market, where retail investors account for a large proportion of market participants, investor sentiment tends to be highly contagious and can spread rapidly through trading behavior. Prior research has shown that investor sentiment, overtrading, and market volatility are closely related in the A-share market, suggesting that trading activity may play an important role in the transmission of sentiment effects. At the same time, recent studies have further examined the impact of mixed-frequency investor sentiment on stock market volatility, indicating that sentiment-based indicators may provide useful supplementary information for volatility analysis. Nevertheless, existing studies still face several limitations. In particular, the direct association between sentiment and volatility often appears weak or unstable, and the transmission mechanism through which sentiment affects volatility has not been sufficiently examined [1].

To address this issue, this study takes selected major constituent stocks of the CSI 300 Index as the research sample and constructs multidimensional investor sentiment factors based on social media text data from the East Money Stock Forum. By linking sentiment indicators with turnover and volatility measures, this paper examines the relationship between investor sentiment and stock market volatility and further explores the role of trading activity in this process. The study aims to provide additional empirical evidence on how social-media-based investor sentiment is associated with market fluctuations in China's A-share market [2].

2. Literature Review and Research Hypotheses

2.1. Investor Sentiment and Stock Market Fluctuation

Investor sentiment has become an important concept in behavioral finance for explaining market fluctuations that cannot be fully interpreted by traditional rational pricing theories. Behavioral finance argues that investors are not always fully rational and that their decisions are often influenced by psychological biases, subjective beliefs, and emotional reactions. As a result, asset prices may deviate from fundamentals, and sentiment-driven trading may amplify market instability [1].

With the rapid development of internet platforms and text analysis techniques, investor sentiment can now be measured more directly from online discussions and social media content. This has expanded the scope of empirical research on the relationship between sentiment and financial markets. Existing studies generally suggest that investor sentiment is associated with stock market performance, but the direction and strength of this relationship are not always consistent. In particular, sentiment may affect the market not only through price expectations, but also through changes in trading activity, liquidity, and risk perception [1].

In the context of China's A-share market, investor sentiment is particularly important because retail investors account for a large share of market participants, making sentiment more likely to spread rapidly and influence short-term market dynamics [3]. Prior studies have shown that major events and information shocks can significantly affect investor sentiment, which in turn may alter market behavior and increase uncertainty. In addition, empirical evidence based on dynamic econometric models indicates that investor sentiment may have time-varying and nonlinear effects on market outcomes, suggesting that its impact cannot be fully captured by simple static relationships.

Although the existing literature has provided valuable evidence on the relationship between investor sentiment and market fluctuation, several limitations remain. First, the measurement of investor sentiment still varies substantially across studies, which may partly explain the inconsistency in empirical findings. Second, many studies focus mainly on the direct association between sentiment and market performance, while less attention is given to the possible transmission mechanisms through which sentiment affects volatility. Therefore, further research is still needed to clarify how sentiment is related to stock market fluctuation and through which channels this relationship may operate [4].

2.2. The Hypotheses of Research

Existing studies suggest that investor sentiment extracted from online platforms contains useful information for understanding stock market behavior, but its direct predictive effect is often weak and unstable across different samples and market settings. In addition, prior research has shown that the relationship between investor sentiment and market outcomes may vary across different market environments and may exhibit nonlinear characteristics. Regime-dependent modeling studies further indicate that sentiment-related effects may differ across market states, although the explanatory power of sentiment for regime transitions is not always strong [5]. Meanwhile, recent evidence suggests that investor sentiment may also influence market volatility indirectly through trading activity or liquidity-related channels.

Based on the above literature and the characteristics of social-media-based sentiment data, this study proposes the following hypotheses [1].

H1: In the current research setting, the direct association between aggregate investor sentiment and market volatility is weak, whereas sentiment intensity is positively associated with volatility.

This hypothesis is based on the idea that sentiment direction alone may not provide stable explanatory power for volatility, while the extremity or intensity of sentiment may better capture market panic, disagreement, or emotional fluctuation, and therefore show a stronger relationship with volatility [6].

H2: Investor sentiment affects market volatility indirectly through trading activity, and turnover rate plays an important mediating role in the sentiment--volatility relationship [7].

This hypothesis reflects the view that sentiment influences investors' willingness to trade and market participation, which may increase short-term fluctuations in trading behavior and further amplify market volatility [8].

H3: The impact of investor sentiment on volatility is heterogeneous across different volatility conditions [9].

This hypothesis is motivated by prior evidence that the effect of sentiment may not be constant across market environments. In relatively calm periods, sentiment may have limited influence, whereas in high-volatility periods, sentiment shocks may be more easily amplified and generate stronger market reactions [3].

H4: Market volatility exhibits regime-switching characteristics, but investor sentiment has limited ability to explain or drive market regime transitions [10].

This hypothesis is proposed because financial markets often display different volatility states, while sentiment variables may be more effective in explaining short-term fluctuations within a state than in identifying changes between states [11].

3. Methodology

3.1. Data Sources and Sample Construction

This study utilizes a panel of representative sample stocks to examine the relationship between investor sentiment and stock market fluctuations. The sample was constructed based on the availability, continuity, and comparability of market data over the study period. To ensure the reliability of the empirical analysis, stocks with incomplete observations, prolonged trading suspensions, or other abnormal trading conditions were excluded from the final sample [12]. Consequently, the selected firms provide a relatively stable and representative basis for analyzing sentiment-driven market dynamics.

The data used in this study were collected from authoritative financial databases and publicly available market records, including stock prices, trading volume, and other relevant market indicators. These variables were employed to calculate the core measures required for subsequent empirical testing. To reflect the actual market influence of each stock more accurately, this study adopts free-float market capitalization rather than total market capitalization as the weighting standard [12]. Compared with total market capitalization, free-float market capitalization better captures the portion of shares that are actively tradable in the market and therefore provides a more realistic representation of each stock's effective contribution to market movements.

Table 1 reports the free-float market capitalization weights of individual sample stocks included in this study. The weights indicate the relative importance of each stock in the sample portfolio and serve as the basis for constructing weighted market-level measures. By applying this weighting approach, the empirical design reduces the potential bias caused by treating all firms equally regardless of their market size, while also improving the representativeness of the aggregated indicators. In this way, larger and more liquid firms exert a proportionally greater influence on the sample-based market measures, which is consistent with actual market conditions [13].

Table 1. Free-Float Market Capitalization Weights of Individual Sample Stocks

constituent stocks	weight	constituent stocks	weight
SSE: 600519	0.1720	XSHE: 000333	0.0767
SSE: 601318	0.1454	SSE: 601166	0.0711
XSHE: 300308	0.1351	SSE: 600900	0.0664
SSE: 601899	0.1132	XSHE: 300059	0.0585
SSE: 600036	0.1039	XSHE: 002475	0.0576

Overall, the sample construction procedure and weighting method enhance the robustness of the dataset and provide a solid foundation for the subsequent analysis of how investor sentiment affects stock market fluctuations.

3.2. Construction of Investor Sentiment Factors

3.2.1. Sentiment Analysis

Investor sentiment is quantified using posts collected from the East Money Guba stock forum. Textual sentiment is identified with the SnowNLP model, which assigns a sentiment score to each post on a scale from 0 to 1, with higher values indicating more positive sentiment [1]. Based on these post-level sentiment scores, three daily firm-level investor sentiment factors are constructed: the Proportion of Positive Sentiment Information (PSI), the Proportion of Negative Sentiment Information (NSI), and the Sentiment Volatility Index (SVI).

To classify the sentiment polarity of forum posts, a post is defined as positive when its sentiment score is greater than or equal to 0.6, negative when its sentiment score is less than or equal to 0.4, and neutral when its score falls between 0.4 and 0.6. In the baseline regression analysis, neutral posts are excluded to focus more directly on clearly polarized investor sentiment [14]. On this basis, PSI and NSI are calculated as the daily proportions of positive and negative posts for each firm, respectively, while SVI is measured as the standard deviation of daily sentiment scores, capturing the degree of divergence or fluctuation in investor sentiment.

3.2.2. Market Composite Sentiment

To capture overall market sentiment more comprehensively, this study applies Principal Component Analysis (PCA) to three market-level sentiment indicators: market-level PSI, market-level NSI, and market-level SVI [15]. PCA is utilized to extract the common variation embedded in these sentiment-related variables and to reduce potential multicollinearity among them.

The first principal component, which explains 58.2% of the total variance, is retained as the proxy for aggregate market sentiment [11]. To improve interpretability and facilitate subsequent empirical analysis, this component is further standardized and defined as the composite investor sentiment index (hereafter denoted as composite sentiment_std).

3.2.3. Sentiment Absolute Value

To further capture the intensity of market sentiment regardless of its direction, this study defines the Sentiment Absolute Value as the absolute value of the standardized composite investor sentiment index, as shown in Equation (1).

Sentiment Absolute Value = ∣Comprehensive Sentiment_std∣ (1)

This indicator measures the extremity of market sentiment. A larger value of Sentiment Absolute Value indicates that market sentiment deviates further from the neutral state, regardless of whether such sentiment is optimistic or pessimistic. In other words, this measure focuses on the strength rather than the direction of sentiment [15]. By transforming the composite sentiment index into its absolute value, this study examines whether stronger sentiment polarization is associated with greater stock market fluctuations.

3.3. Volatility and Turnover Measurement

3.3.1. Volatility

Volatility is measured using the rolling 20-day annualized standard deviation of daily logarithmic returns [16]. Specifically, daily log returns are first calculated for the market index, and then a 20-trading-day rolling window is applied to obtain the standard deviation of returns. The resulting measure is annualized to effectively capture the recent level of market volatility.

3.3.2. Turnover Rate

Turnover Rate is used to capture market trading activity and liquidity conditions. It is measured as the ratio of the 5-minute transaction amount of the CSI 300 Index to the total tradable market capitalization of CSI 300 constituent stocks on the same day, as shown in Equation (2). Because tradable market capitalization data are available at the daily frequency, they are down-filled to the 5-minute frequency to match the intraday transaction data.

$$\text{Turnover Rate} = \frac{\text{5-minute transaction amount of the CSI 300 Index}}{\underset{\text{(Daily data, down-filled to 5-minute frequency)}}{\text{Total tradable market capitalization of CSI 300 constituent stocks on the day}}} \quad (2)$$

A higher Turnover Rate indicates more active trading and stronger market participation, while a lower Turnover Rate reflects relatively weaker trading activity. This variable is included to assess whether fluctuations in investor sentiment are associated not only with changes in market volatility but also with variations in trading intensity [15].

3.4. Empirical Model

This study adopts a comprehensive empirical framework to examine the relationship between investor sentiment and stock market behavior. First, Pearson correlation analysis is conducted to provide preliminary evidence on the associations among the main variables [17]. Second, ordinary least squares (OLS) regression is employed to estimate the direct effect of investor sentiment on market volatility. To further investigate whether this effect varies across different volatility conditions, quantile regression is applied, allowing the analysis to capture heterogeneous effects at different points of the volatility distribution.

In addition, a vector autoregression (VAR) model is constructed to explore the dynamic interactions between investor sentiment and market volatility over time. To test whether turnover acts as a mediating variable in the relationship between sentiment and market fluctuations, this study applies stepwise regression analysis together with relevant mediation testing procedures. At the individual stock level, panel data models, including both fixed-effects and random-effects specifications, are used to examine how firm-specific investor sentiment influences stock returns while controlling for unobservable heterogeneity across firms [17].

Furthermore, a two-regime Markov switching model is employed to identify potential regime changes in market volatility and to determine whether the effect of investor sentiment differs across volatility states. Finally, a series of robustness checks is conducted, including alternative variable specifications, subsample analyses, and model adjustments, in order to ensure the stability and reliability of the empirical findings.

4. Results

4.1. Empirical Analysis at the Market Level

4.1.1. Descriptive Statistics of Sentiment Factors

Figure 1 presents the comprehensive relationships among sentiment factors and market volatility. As shown in the upper-left panel of Figure 1, the composite sentiment fluctuates around zero, while market volatility exhibits clear clustering characteristics. This pattern suggests that investor sentiment varies continuously over time, whereas volatility tends to persist once the market enters a relatively turbulent period [7].

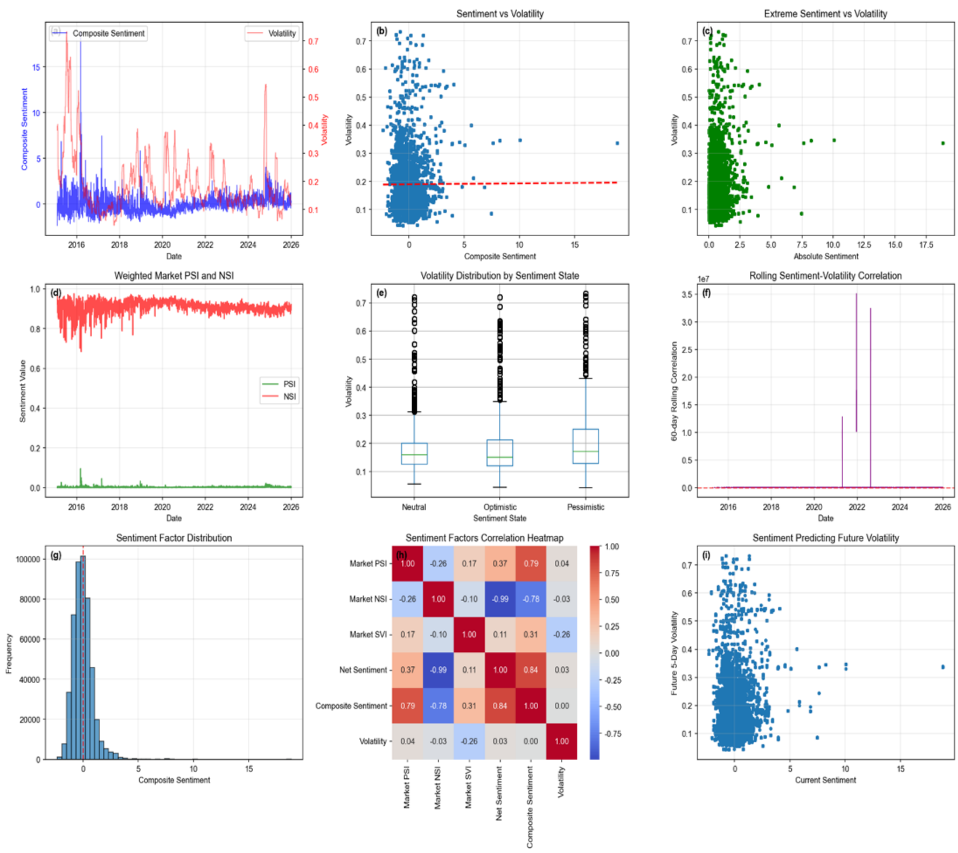

Figure 1. Comprehensive Analysis Chart of Sentiment Factor and Volatility

As further illustrated in Figure 1, the weighted market-level PSI remains close to zero over the sample period, whereas the weighted market-level NSI stays around 0.9, indicating that sentiment expressed in stock forums is predominantly negative. This result suggests that investors in online stock discussions tend to exhibit stronger pessimistic sentiment than optimistic sentiment, reflecting a generally cautious or risk-sensitive market atmosphere.

The principal component analysis shows that the first principal component explains 58.2% of the total variance, supporting its use as a representative measure of overall market sentiment. The standardized composite sentiment index has a mean of 0 and a standard deviation of 1, while the Sentiment Absolute Value has a mean of 0.674 and a standard deviation of 0.739. As shown in Figure 1, the distribution of the Sentiment Absolute Value is right-skewed, indicating that market sentiment remains relatively moderate in most periods, while episodes of extreme sentiment occur less frequently. Overall, these descriptive results provide preliminary evidence that investor sentiment exhibits substantial time variation and may be closely associated with fluctuations in market volatility.

4.1.2. Correlation Analysis and OLS Regression

Table 2 reports the Pearson correlation coefficients among the main market-level variables. The results show that Composite Sentiment_std is strongly positively correlated with Market_PSI and Market_Net Sentiment, and strongly negatively correlated with Market_NSI, indicating that the composite sentiment index effectively captures the directional variation in market sentiment [1]. By contrast, Abs_Sentiment exhibits a relatively moderate correlation with Composite Sentiment_std but shows a stronger association with sentiment dispersion and extremity, reflecting the intensity rather than the direction of sentiment.

Table 2. Pearson Correlation Coefficient Matrix

	Composite Sentiment_std	Abs_Sentiment	Market_PSI	Market_NSI	Market_Net Sentiment	Vol_20d	Ret
Composite Sentiment_std	1.0000	0.5313	0.7945	-0.7782	0.8412	0.0031	0.0165
Abs_Sentiment	0.5313	1.0000	0.6848	-0.2370	0.3071	0.1576	-0.0081
Market_PSI	0.7945	0.6848	1.0000	-0.2633	0.3686	0.0416	-0.0081
Market_NSI	-0.7782	-0.2370	-0.2633	1.0000	-0.9938	-0.0263	-0.0272
Market_Net Sentiment	0.8412	0.3071	0.3686	-0.9938	1.0000	0.0302	0.0253
Vol_20d	0.0031	0.1576	0.0416	-0.0263	0.0302	1.0000	-0.0411
Ret	0.0165	-0.0081	-0.0081	-0.0272	0.0253	-0.0411	1.0000

These results preliminarily support Hypothesis 1, indicating that emotional direction has no predictive power, while emotional intensity shows a weak but robust positive correlation with volatility [16].

More importantly, as shown in Table 2, the correlation between Composite Sentiment_std and Vol_20d is close to zero (0.0031), suggesting that sentiment direction has little linear association with market volatility. In contrast, Abs_Sentiment is positively correlated with Vol_20d (0.1576), although the magnitude remains modest. This finding provides preliminary support for Hypothesis 1, indicating that emotional direction has limited predictive value for volatility, whereas emotional intensity is more closely related to market fluctuations [5]. In addition, the correlations between sentiment variables and market returns are generally weak, suggesting that sentiment may affect market conditions primarily through volatility rather than through contemporaneous return changes.

The OLS regression results reported in Table 3 further support this conclusion. In Model 1, Composite Sentiment_std has a statistically significant coefficient, but its explanatory power is negligible, with an R2 of 0.0000, indicating that sentiment direction alone contributes almost nothing to explaining market volatility. In Model 2, Abs_Sentiment has a significantly positive coefficient (0.0221, p = 0.000) and explains approximately 2.48% of the variation in volatility, suggesting that stronger sentiment extremity is associated with higher market volatility [17]. In Model 3, when both Composite Sentiment_std and Abs_Sentiment are included simultaneously, the coefficient on Abs_Sentiment remains significantly positive (0.0305, p = 0.000), whereas Composite Sentiment_std becomes significantly negative (-0.0117, p = 0.000). This result suggests that after controlling for sentiment intensity, the directional component of sentiment may exhibit a different marginal effect; however, the dominant explanatory contribution still comes from sentiment intensity.

Table 3. Market Volatility Regression Results

Model	Var.	Coef.	P-val.	R^2
1	Composite Sentiment_std	0.0003	0.034	0.0000
2	Abs_Sentiment	0.0221	0.000	0.0248
3	Composite Sentiment_std	-0.0117	0.000	0.0248
	Abs_Sentiment	0.0305	0.000	

Overall, the evidence from Table 2 and Table 3 supports Hypothesis 1: the absolute value of sentiment shows a weak but robust positive relationship with market volatility, while aggregate sentiment direction by itself has very limited explanatory power.

4.1.3. Quantile Regression Model

As shown in Figure 2, the effect of aggregate sentiment on market volatility exhibits clear nonlinear characteristics across different quantiles. Specifically, the estimated coefficients are significantly negative at lower quantiles and significantly positive at higher quantiles, indicating that the relationship between investor sentiment and volatility varies across different volatility regimes.

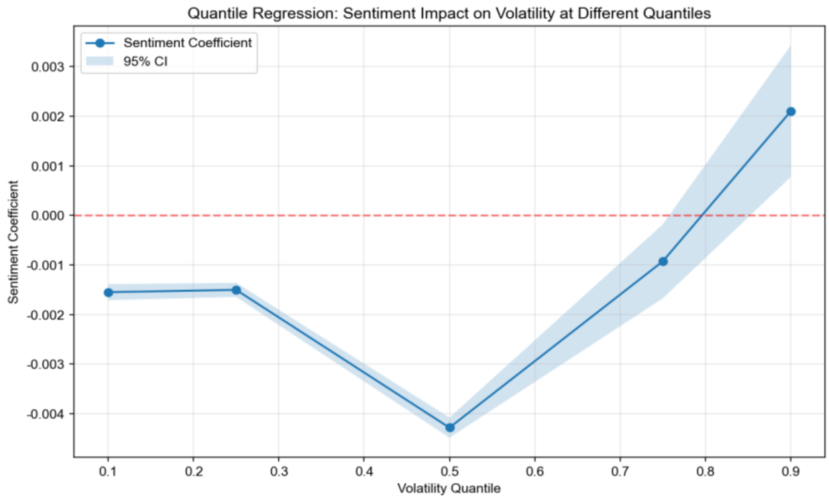

Figure 2. Quantile Regression: The Impact of Sentiment on Volatility Across Different Quantiles

However, although the coefficients are statistically significant at several quantiles, their absolute values are all below 0.005, suggesting that the economic significance of aggregate sentiment remains relatively weak. In other words, while aggregate sentiment shows heterogeneous effects across the volatility distribution, its overall marginal impact on market volatility is limited in magnitude.

Overall, the results presented in Figure 2 suggest that the influence of investor sentiment on volatility is not constant but depends on the prevailing market condition. This finding provides further evidence that the sentiment-volatility relationship is asymmetric across different levels of market volatility [5].

4.1.4. VAR Model and Impulse Response Function

This study takes Composite Sentiment_std as the independent variable, Volatility_20d as the dependent variable, and Turnover Rate as the mediating variable to conduct the mediation effect test [17]. The results revealed that sentiment exerted a weak positive influence on volatility through a total effect coefficient of c=0.0031 (p=0.0335). Path coefficients demonstrated significant positive correlations between sentiment and turnover rate, as well as between turnover rate and volatility. After controlling for turnover rate, the direct effect of sentiment on volatility became insignificant, while the indirect effect remained statistically significant (Bootstrap 95% CI: [0.0036, 0.0041], Sobel test Z=29.06).

Figure 3 presents the impulse response functions of sentiment, volatility, and returns over a 10-period horizon. As shown in Figure 3, investor sentiment exhibits extremely strong persistence in response to its own shocks. The initial response is close to 1 in the first period and remains above 0.96 even in the tenth period. This pattern is consistent with the VAR autoregressive coefficient of 0.996, indicating near-unit-root behavior and a high degree of persistence in sentiment dynamics.

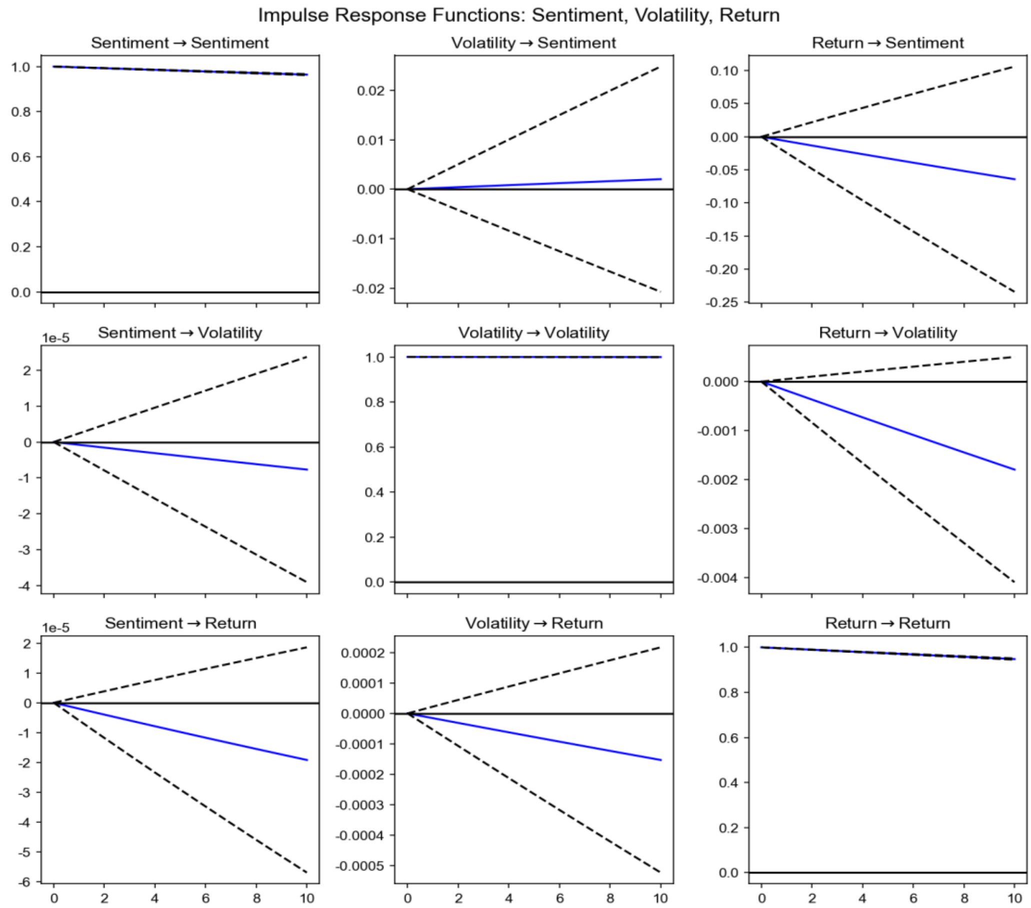

Figure 3. Impulse Response Function of Sentiment, Volatility and Returns (10-Period Lag)

By contrast, the response of market volatility to sentiment shocks fluctuates around zero, and the confidence interval consistently includes zero, suggesting that sentiment does not exert a statistically significant dynamic effect on volatility. Similarly, the response of returns to sentiment shocks is weak and statistically insignificant throughout the forecast horizon. These results are in line with the findings from the Granger causality test and the previous correlation analysis, both of which indicate that the direct predictive power of sentiment for volatility and returns is limited.

Further, as shown in Figure 3, sentiment remains close to zero and does not display significant deviation during the initial phase of volatility shocks, again suggesting the absence of a strong predictive relationship. Overall, the VAR results indicate that sentiment is largely driven by its own historical values and follows a highly persistent dynamic process, whereas volatility exhibits more typical clustering and mean-reverting characteristics. Therefore, the direct linear effect of sentiment on volatility appears to be extremely weak, and conventional linear models may have limited ability to capture the underlying relationship [15].

4.1.5. Analysis of Periods with Extreme Sentiment

To further examine the relationship between extreme sentiment and market volatility, Composite Sentiment_std is sorted in ascending order [10]. The bottom 20% of observations are classified as pessimistic periods, while the top 20% are classified as optimistic periods. Based on this grouping, the difference in average market volatility between pessimistic and optimistic periods is -1.77 percentage points, and the result is statistically significant according to the t-test ($p < 0.01$).

This finding indicates that market volatility is significantly higher during pessimistic periods than during optimistic periods. The result is consistent with the interpretation of the absolute sentiment measure, suggesting that extreme negative sentiment is more likely to be associated with elevated market fluctuations. More broadly, it provides additional evidence that the effect of investor sentiment on market behavior is more pronounced when sentiment becomes extreme, particularly on the pessimistic side.

4.1.6. Two-Regime Markov-Switching Regression Model

To further explore whether the relationship between investor sentiment and market volatility differs across market states, this study estimates a two-regime Markov-switching regression model. However, the estimation results indicate that the model does not yield reliable or interpretable parameters [15].

As shown in Figure 4, the constant term in Regime 0 takes a sign opposite to that of the overall sentiment coefficient, while its standard error is extremely small, suggesting the presence of severe multicollinearity. In Regime 1, the estimated coefficients have very large absolute values, and their standard errors are reported as NaN, further indicating instability in parameter estimation. In addition, the transition probabilities reveal extremely high persistence in State 0 ($p = 0.9999$), implying that the model is unable to effectively distinguish between alternative volatility regimes. The high condition number of the covariance matrix also suggests that the model is severely singular and fails to converge properly [3].

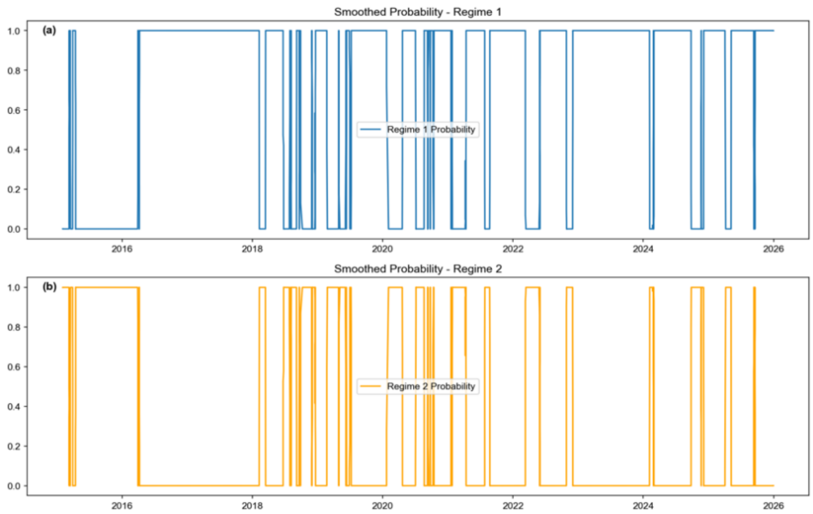

Figure 4. Markov State Probability

Several factors may explain this estimation failure. First, the high-frequency volatility series exhibits strong autocorrelation, which makes it difficult for a simple two-regime linear structure to identify distinct market states. Second, investor sentiment has only limited explanatory power for volatility in the baseline analysis and therefore does not provide sufficient discriminatory information to support stable state classification. Third, given the very large sample size of approximately 480,000 observations, the numerical optimization procedure is more susceptible to local singularities and convergence problems.

Overall, the results reported in Figure 4 suggest that the two-regime Markov-switching specification is not suitable for the current dataset. Therefore, its estimates should not be interpreted substantively, and the findings are better treated as evidence of model instability rather than as support for regime-dependent sentiment effects [1, 11].

4.2. Daily Panel Analysis at the Individual Stock Level

To maintain consistency with the market-level analysis, this study further conducts fixed-effects and random-effects regressions using the daily panel data of the 10 sample stocks [2, 12]. The estimation results are reported in Table 4.

Table 4. Panel Regression Results (Individual Stock Daily)

Var.	FE	RE
const	-4.22e-05 (0.984)	8.05e-05 (0.950)
PSI	-0.0190 (0.172)	-0.0182 (0.102)
NSI	-0.0012 (0.535)	-0.0013 (0.297)
SVI	0.0086 (0.003)	0.0079 (0.012)
R^2	0.0004	0.0004
Observations	25,854	25,854

As shown in Table 4, SVI exhibits a significantly positive association with individual stock returns in both model specifications. Specifically, the coefficient of SVI is 0.0086 in the fixed-effects model ($p = 0.003$) and 0.0079 in the random-effects model ($p = 0.012$), indicating that greater sentiment divergence is associated with higher daily stock returns. By contrast, neither PSI nor NSI has a statistically significant effect on returns in either specification, suggesting that the directional components of sentiment do not provide strong explanatory power at the individual stock level.

However, the overall explanatory power of the panel models remains extremely limited. As reported in Table 4, the R2 value is only 0.0004 in both the fixed-effects and random-effects regressions, indicating that the sentiment variables explain less than 0.1% of the variation in daily stock returns. This result suggests that although sentiment divergence has a statistically significant positive relationship with individual stock returns, its economic significance is weak and its predictive value remains limited [7].

Overall, the evidence in Table 4 indicates a weak positive association between firm-level sentiment divergence and daily stock returns, whereas the proportions of positive and negative sentiment do not show significant effects. This finding differs from the high-frequency market-level results. At the individual stock level, sentiment appears to be directly related to returns, although only weakly, while at the market level, sentiment primarily affects volatility indirectly through the turnover channel rather than through a strong direct effect.

4.3. Robustness Tests

To assess the stability of the main findings, this study conducts a series of robustness tests using alternative sample splits, alternative volatility windows, and different sentiment construction methods [2]. The results are summarized in Table 5.

Table 5. Summary of Robustness Test Results

Test Items	Corr./ Results
Absolute Value of Sentiment in the First Half vs. Volatility	0.0542(n=242,392)

Absolute Value of Sentiment in the Second Half vs. Volatility	0.0855(n=242,392)
10-Period Volatility vs. Composite Sentiment	-0.0087
60-Period Volatility vs. Composite Sentiment	0.0476
Equal-Weighted Sentiment vs. Volatility	-0.0867
PCA-Based Sentiment vs. Volatility	0.0233
Expert-Weighted Sentiment vs. Volatility	-0.0032
Correlation on Up Days (Composite Sentiment)	0.0544(n=249,973)
Correlation on Down Days (Composite Sentiment)	-0.0633(n=234,811)
Mediation Effect(Turnover Rate)	Full Mediation Effect, Robust Bootstrap CI

As shown in Table 5, the positive association between the absolute value of sentiment and market volatility remains broadly stable across different subsamples. Specifically, the correlation between the absolute value of sentiment and volatility is 0.0542 in the first half of the sample and 0.0855 in the second half, suggesting that the relationship is not driven by a particular subperiod. In addition, when alternative volatility horizons are used, the correlation between Composite Sentiment and 10-period volatility is -0.0087, while that with 60-period volatility is 0.0476. These results indicate that the sentiment-volatility relationship varies somewhat across time horizons, but the overall pattern remains limited in magnitude.

The robustness analysis also considers alternative sentiment aggregation methods. As reported in Table 5, the correlation between equal-weighted sentiment and volatility is -0.0867, that between PCA-based sentiment and volatility is 0.0233, and that between expert-weighted sentiment and volatility is -0.0032. Although the estimated correlations differ in sign and magnitude under different constructions, the results consistently suggest that sentiment direction alone has weak explanatory power for volatility, whereas sentiment intensity remains the more relevant dimension [7].

Furthermore, the relationship between sentiment and market conditions is examined separately for rising and falling market days. Table 5 shows that the correlation between Composite Sentiment and market fluctuations is 0.0544 on up days and -0.0633 on down days, implying that the effect of sentiment may differ across market environments. This asymmetry is consistent with the earlier findings that the sentiment-volatility relationship is nonlinear and state-dependent [17].

Finally, the mediation analysis reported in Table 5 continues to support a full mediating effect of Turnover Rate, with the Bootstrap confidence interval remaining robust across specifications. This result reinforces the conclusion that investor sentiment primarily affects market volatility indirectly through trading activity rather than through a strong direct channel.

Overall, the evidence in Table 5 confirms that the main conclusions of this study are robust to alternative specifications. In particular, the positive relationship between sentiment intensity and volatility, as well as the mediating role of turnover rate, remains qualitatively unchanged [5, 12].

5. Discussion

5.1. Hypothesis Test Summary

The results of the hypothesis tests are summarized in Table 6. Overall, the empirical evidence largely supports the proposed hypotheses, although the strength of support varies across hypotheses [16].

Table 6. The Results of Hypothesis Test

Hypothesis	Substance	Result
H1	The direct predictive power of emotions at high frequencies is weak, with emotional intensity showing a positive correlation with volatility.	Support. The correlation coefficient for composite emotion was r=0.0031 with $R^2 \approx 0$; the absolute emotion value showed r=0.1576 and R^2=2.48%.
H2	Strongly supported. Full mediation holds, with a significant indirect effect and an insignificant direct effect.	Strongly supported. Full mediation holds, with a significant indirect effect and an insignificant direct effect.
H3	The impact of sentiment on volatility exhibits asymmetry at different volatility levels.	Supported. Significantly negative in low quantile, positive in high quantile, with minimal coefficient.
H4	The high-frequency volatility series presents regime switching characteristics, while sentiment variables can hardly effectively drive regime shifts.	Partially supported. The Markov model failed to converge, indicating difficulty in regime identification.

For H1, the results suggest that investor sentiment has limited direct predictive power for market volatility at the high-frequency level [9, 10]. In particular, Composite Sentiment exhibits an almost negligible correlation with volatility (r = 0.0031) and virtually no explanatory power (R2 ≈ 0). By contrast, Abs_Sentiment shows a positive correlation with volatility (r = 0.1576) and a modest explanatory power (R2 = 2.48%). These findings indicate that the intensity of sentiment, rather than its direction, is more closely associated with market volatility. Therefore, H1 is supported.

For H2, the mediation analysis provides strong evidence that Turnover Rate fully mediates the relationship between investor sentiment and market volatility. Specifically, the indirect effect remains statistically significant, whereas the direct effect becomes insignificant after controlling for turnover rate. This result suggests that investor sentiment affects volatility primarily through its influence on trading activity, rather than through a direct transmission channel. Accordingly, H2 is strongly supported.

For H3, the quantile regression results reveal that the effect of sentiment on volatility is not constant across the distribution of volatility. The estimated coefficients are significantly negative at lower quantiles and significantly positive at higher quantiles, indicating a clear asymmetric pattern [9]. Although the economic magnitude of these coefficients remains small, the results nonetheless support the view that the sentiment-volatility relationship varies across market conditions. Thus, H3 is supported.

For H4, only partial support is obtained. Although the high-frequency volatility series appears to exhibit certain regime-dependent characteristics, the Markov-switching model fails to converge and does not produce stable parameter estimates [16]. This suggests that, in the present setting, sentiment variables do not provide sufficient explanatory power to identify or drive regime transitions reliably. Therefore, H4 is partially supported.

Taken together, these results suggest that investor sentiment does not exert a strong direct effect on volatility in a linear framework. Instead, its influence is more likely to operate through sentiment intensity, trading activity, and state-dependent market conditions.

5.2. Mechanistic Evidence for Spillover Effects

The findings of this study suggest that investor sentiment does not directly generate market shocks; rather, it amplifies market volatility indirectly by influencing investors' trading behavior, as captured by turnover rate [12]. This result provides empirical support for the behavioral finance mechanism linking sentiment, trading activity, and price dynamics. More specifically, the evidence indicates that sentiment-driven changes in market participation constitute an important transmission channel through which psychological factors are translated into observable market fluctuations. In this sense, the results offer mechanistic evidence for the spillover effect of investor sentiment on market volatility.

5.3. Interpretation of Daily-Level Results

The daily-level analysis further reinforces the main mechanism identified at the market level. The results show that investor sentiment continues to exert a significant effect on turnover rate, whereas its direct effect on volatility becomes insignificant after the mediating variable is introduced. This finding supports the view that sentiment shocks do not affect volatility immediately and directly, but instead operate through a dynamic process of lagged transmission, amplification, and eventual attenuation, with turnover rate serving as the key channel through which sentiment is converted into price volatility.

In addition, at the daily frequency, investor sentiment exhibits a significant indirect effect and only a weak total effect on volatility, which is consistent with a suppression effect. Specifically, the positive indirect effect through turnover rate offsets the negative and statistically insignificant direct effect of sentiment on volatility. This mechanism helps explain why the overall effect of investor sentiment on volatility is often weak or insignificant in prior studies. The present findings suggest that such insignificance may not imply the absence of an effect, but rather reflect the coexistence of opposing transmission channels within the sentiment-volatility relationship.

5.4. Causes of the Markov Model's Failure and Research Implications

Although high-frequency volatility exhibits strong autocorrelation and pronounced volatility clustering, the two-regime linear Markov-switching model fails to produce stable and interpretable estimates in this study. One important reason is that the model assumes fixed transition probabilities between regimes, which may be too restrictive to capture the continuous, gradual, and highly persistent evolution of volatility in high-frequency financial data [3, 15]. Consequently, the model has limited ability to distinguish clearly between latent states when regime changes are not abrupt.

A further issue arises from the extremely large sample size of approximately 480,000 observations. Under such conditions, the likelihood function may become relatively flat in certain regions and highly sensitive to initial parameter values, thereby increasing the risk of numerical singularity and non-convergence during estimation. Additionally, because sentiment variables themselves display only weak direct explanatory power for volatility, they are unlikely to provide sufficient information to support reliable regime classification within a simple two-state specification.

These results carry important implications for future research [2, 3]. Regime-dependent analyses of high-frequency volatility may require more flexible nonlinear models, such as specifications with time-varying transition probabilities, state-dependent exogenous variables, or alternative machine learning approaches to state identification. Furthermore, the failure of the Markov model in the present study suggests that the influence of investor sentiment on volatility may be better understood through transmission mechanisms and conditional heterogeneity rather than through a simple discrete regime-switching framework. Accordingly, future studies should place greater emphasis on the dynamic and indirect channels through which sentiment affects financial markets.

6. Conclusion

This study constructs a comprehensive market sentiment factor based on online forum posts related to 10 CSI 300 constituent stocks and combines it with the 5-minute turnover rate of the CSI 300 Index to examine the spillover effect of investor sentiment on market volatility. Using a high-frequency dataset with approximately 485,000 observations, the study explores both the direct effect of sentiment and its transmission mechanism.

The results show that investor sentiment has very weak direct predictive power for volatility. However, the absolute value of sentiment is positively associated with volatility, indicating that sentiment intensity is more relevant than sentiment direction. More importantly, turnover rate is found to fully mediate the relationship between sentiment and volatility, suggesting that investor sentiment affects market volatility mainly through trading activity. In addition, the quantile regression results indicate that the sentiment-volatility relationship is asymmetric across different volatility conditions, although its economic significance remains limited.

These findings contribute to the literature by clarifying the transmission path of sentiment spillover and highlighting the role of sentiment intensity as a useful signal of market risk. From a practical perspective, the combination of absolute sentiment value and turnover rate may provide a useful reference for market monitoring and risk warning.

This study also has several limitations, including the assumption of constant intraday sentiment, the focus on only one mediating variable, the limited representativeness of the 10-stock sample, and the failure of the Markov-switching model to converge. Future research may improve intraday sentiment measurement, incorporate additional mediating variables, adopt more flexible dynamic models, and further explore the practical application of sentiment-based risk indicators.

References

1. D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*, 2013, pp. 99-127.
2. A. Shleifer, *Inefficient markets: An introduction to behavioural finance*. OUP Oxford, 2000.
3. G. W. Schwert, "Why does stock market volatility change over time?," *The Journal of Finance*, vol. 44, no. 5, pp. 1115-1153, 1989.
4. E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383-417, 1970.
5. E. Cevik, B. Kirci Altinkeski, E. I. Cevik, and S. Dibooglu, "Investor sentiments and stock markets during the COVID-19 pandemic," *Financial Innovation*, vol. 8, no. 1, p. 69, 2022.
6. C. Huang, X. Yang, X. Yang, and H. Sheng, "An empirical study of the effect of investor sentiment on returns of different industries," *Mathematical Problems in Engineering*, vol. 2014, no. 1, p. 545723, 2014.
7. Z. Zhou, "The impact of investor sentiment fluctuations on stock market liquidity and volatility: Implications for market returns," *Finance Research Letters*, p. 108771, 2025.
8. Z. McGurk, A. Nowak, and J. C. Hall, "Stock returns and investor sentiment: textual analysis and social media," *Journal of Economics and Finance*, vol. 44, no. 3, pp. 458-485, 2020.
9. D. Guégan and T. Renault, "Does investor sentiment on social media provide robust information for Bitcoin returns predictability?," *Finance Research Letters*, vol. 38, p. 101494, 2021.
10. Y. Seetharam and K. Nyakurukwa, "Headlines or Hashtags? The battle in social media for investor sentiment in the stock market," *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100273, 2024.
11. P. K. Maurya, R. Bansal, and A. K. Mishra, "Investor sentiment and its implication on global financial markets: a systematic review of literature," *Qualitative Research in Financial Markets*, vol. 18, no. 1, pp. 209-248, 2026.
12. A. Rizkiana, H. Sari, P. Hardjomijojo, B. Prihartono, and T. Yudhistira, "Analyzing the impact of investor sentiment in social media to stock return: Survival analysis approach," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2017, pp. 519-523.
13. E. Affuso and K. D. Lahtinen, "Social media sentiment and market behavior," *Empirical Economics*, vol. 57, no. 1, pp. 105-127, 2019.
14. V. M. de Sousa-Gabriel, M. B. Lozano-García, M. F. L. I. Matias, M. E. Neves, and J. Martínez-Ferrero, "Global environmental equities and investor sentiment: the role of social media and Covid-19 pandemic crisis," *Review of Managerial Science*, 2023.
15. S. Y. Yang and S. Y. K. Mo, "Social media and news sentiment analysis for advanced investment strategies," in *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*, Cham: Springer International Publishing, 2016, pp. 237-272.

16. Q. Liu, W. S. Lee, M. Huang, and Q. Wu, "Synergy between stock prices and investor sentiment in social media," *Borsa Istanbul Review*, vol. 23, no. 1, pp. 76-92, 2023.
17. Q. Liu and H. Son, "Methods for aggregating investor sentiment from social media," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1-22, 2024.