A Method for Analyzing Stock Price Volatility Based on Market Sentiment

Xinjian Ji
Dalian University of Technology, Business School

Tong Li
Dalian University of Technology, Business School

Abstract. In recent years, the Chinese stock market has exhibited frequent abnormal fluctuations due to major capital operations and shifts in international economic policies, posing significant challenges to investors' decision-making. This paper introduces a method for analyzing stock price volatility grounded in system dynamics principles. By constructing a large-scale dataset of Chinese stock commentary, we developed a precise sentiment analysis model tailored to Chinese stock reviews and introduced a novel public opinion influence coefficient model, significantly enhancing the correlation between stock price volatility and public sentiment data. In the realm of natural language processing, our approach represents a key advancement in understanding the impact of market sentiment on stock market behavior, offering robust theoretical and empirical support.

Additionally, to promptly detect abnormal stock price fluctuations, this paper proposes a warning mechanism that integrates engineering statistics methods, culminating in a comprehensive abnormal fluctuation warning model. This approach delivers a systematic solution encompassing "data processing, prediction, warning, and analysis," and empirical testing has confirmed its accuracy and reliability in practical applications. This research successfully bridges theoretical insights with practical implementation, offering new perspectives and methodologies for stock price volatility analysis.

Key words: Stock volatility analysis, stock price prediction, stock price warning, Market sentiment

1. Introduction

In rapidly changing global financial markets, stock price volatility not only directly influences investors' decisions but also significantly impacts market liquidity and policy effectiveness (Li et al. 2023). Consequently, accurate analysis of stock price volatility is crucial for individual investors, corporations, and policymakers. However, with the increasing uncertainty in the current international environment, effectively understanding stock price volatility has become an ever-greater challenge (Hashmi et al. 2021).

The primary challenge lies in the complexity of market microstructure, especially the behavior of large institutional investors (Madhavan 2002). These large and agile funds, through carefully crafted trading strategies—such as targeted inflows and outflows of substantial capital—can effectively manipulate market sentiment (Liu et al. 2019), prompting retail investors to exit prematurely or chase rising prices, thereby reaping excess profits from price movements (Koesrindartoto et al. 2020). This process often exacerbates market uncertainty, making it difficult for ordinary investors to accurately predict market trends (Wei et al. 2020).

Additionally, global political and economic instability, particularly the asymmetrical adjustments in financial policies among nations, further complicates the external environment affecting stock price volatility (Li et al. 2019). The implementation of adverse financial policies, such as trade sanctions and capital controls, not only triggers sudden inflows or outflows of foreign capital (Chang et al. 2015) but also causes market sentiment to deviate sharply from fundamental economic conditions, leading to market panic and significant disruptions in the Chinese stock market (He et al. 2020).

As the Chinese stock market increasingly exhibits abnormal price volatility that appears decoupled from investor sentiment, the urgent issue that needs to be addressed is how to accurately analyze stock price volatility while precisely identifying these abnormal fluctuations.

1.1. Related Research

To analyze trends in stock price fluctuations, early scholars extensively utilized econometric methods in finance for in-depth research.

In the early 1980s, Granger and Hosking, building on fractional differencing and integration, proposed the ARIMA model, which was effective in short-term stock price prediction, capturing short-term fluctuations and trends in stock prices (Granger and Joyeux 1980, Hosking 1981). Subsequently, Engle and others introduced the ARCH model, which effectively identified the phenomenon of volatility clustering in the stock market (Engle 1982). Building on this, many researchers proposed improved models based on the ARCH framework, such as the GARCH model introduced by Bollerslev, which further enhanced the ability to capture volatility (Bollerslev et al. 1994).

In the 21st century, hybrid models have been widely adopted in financial time series forecasting. For instance, Babu and colleagues utilized a partitioned difference model based on ARIMA-GARCH to forecast price fluctuations in the Indian stock market (Babu and Reddy 2015). Xiang and co-authors combined ARIMA (1,1,0) and GARCH (1,1) models to more accurately predict short-term fluctuations in oil prices (Xiang 2022). Despite the significant success of these models in predicting stock market volatility and trends, they exhibit certain limitations. The primary issue is that these models heavily rely on the statistical characteristics of historical fundamental data, while overlooking the core factor driving stock price fluctuations, namely, investor behavior. Additionally, factors related to public opinion, such as social news and public policies, also significantly influence investor sentiment, which in turn indirectly affects stock prices.

To address these limitations, recent studies have increasingly explored the application of natural language processing (NLP) technology in financial markets, particularly through the analysis of online public opinion data to measure market sentiment, incorporating it as a key input variable in the analysis of stock price fluctuations.

In 2005, Xun Liang and Rong-Chang Chen utilized a language parser based on Chinese natural language processing technology to analyze online stock news and demonstrated through neural network experiments a significant relationship between this news and stock returns (Liang and Chen 2005). Subsequently, Vivek Sehgal and Charles Song introduced a stock market prediction approach leveraging internet users' sentiment, which predicted stock price fluctuations by scanning financial forums and extracting the expressed emotions (Sehgal and Song 2007). WeiLi Xia and LinLin Guo conducted an empirical analysis of the Chinese Growth Enterprise Market (GEM), revealing that investor sentiment significantly influences short-term market returns and volatility (Xia and Guo 2015). A. Romanowski and Michal Skuza predicted Apple Inc.'s stock price by classifying sentiments expressed in Twitter data, showing that sentiment classification can effectively forecast future stock price trends (Romanowski and Skuza 2017). Similarly, J. Deveikyte and colleagues examined the correlation strength between sentiment indicators on a given date and market volatility and returns on the following day, concluding that there is evidence supporting a relationship between sentiment and stock market trends (Deveikyte et al. 2022).

Despite the substantial progress in utilizing market sentiment for stock price prediction, several critical issues remain unresolved. First, although market sentiment indices are widely employed in stock price prediction, their range of fluctuation is limited, whereas stock prices often display greater long-term volatility. This discrepancy results in a strong short-term correlation between market sentiment indices and stock prices, but this correlation diminishes over the long term, reducing predictive effectiveness. Second, the relationship between sentiment data and stock price trends in previous studies is often weak and, in some cases, even negative, which diverges from the predictions of market sentiment theory. Lastly, although many studies have successfully utilized online public opinion data to forecast stock fluctuations, there remains a lack of comprehensive theoretical analysis and empirical investigation into the underlying causes of the weak correlation between public opinion data and stock price trends.

1.2. Contributions of This Study

This study introduces a method for analyzing stock price volatility based on system dynamics. First, to enhance the accuracy and efficiency of stock sentiment analysis, we developed a large-scale dataset for classifying sentiments in Chinese stock comments and applied the BERT-wwm-ext-Chinese model for deep learning. Next, to address the significant disparity between public opinion data and stock price volatility, we investigated the relationship between public opinion influence and stock price volatility and proposed a coefficient model for public opinion influence. This model significantly strengthens the correlation between public opinion influence and stock price changes. Finally, in response to the frequent abnormal fluctuations in the current stock market, we designed and constructed a stock price early warning model. This model performs accurate and timely early warning analysis of stock price fluctuations by conducting real-time monitoring of abnormal fluctuations in company stock prices, combined with recent news events, stock market fund flow data, and the correlation between public opinion influence and stock price volatility.

The main contributions of this paper are:

(1) Unlike previous studies that focus on predicting stock prices using market sentiment, we establish a link between market sentiment and stock price volatility, proposing a novel approach to stock price volatility analysis based on system dynamics. This approach significantly enhances the correlation between these factors, greatly improving the long-term relationship between them, thereby increasing the feasibility and accuracy of using online public opinion data to predict stock price fluctuations.

(2) To further assist in identifying abnormal stock price fluctuations, we designed and constructed a stock price early warning model. This model helps identify abnormal fluctuations by analyzing the underlying factors driving stock price changes, enabling more accurate judgments. Experimental results indicate that the stock price early warning software developed using this model demonstrates exceptional sensitivity, providing valuable support for contemporary financial market risk management.

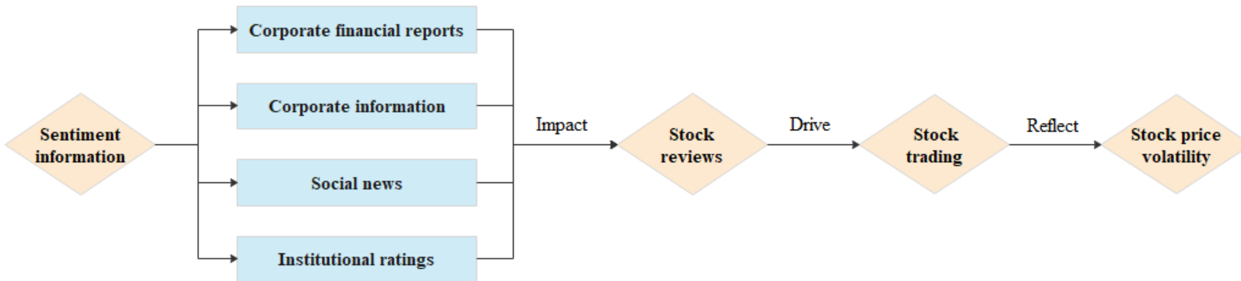

Figure 1 **Relationship between Sentiment Information and Stock Price Volatility.**

2. Stock Price Volatility Prediction Models

2.1. Data Sources and Preprocessing

Sentiment information that influences stock price trends encompasses financial reports, news articles, institutional ratings, and other related content (Audrino et al. 2020). This information shapes investors' perceptions and decision-making, driving their trading behavior and, through cumulative impacts, significantly influences stock price volatility Li et al. (2014). Given that investors are central to stock price movements, their comments are a rich source of sentiment information (Ren et al. 2019). This study selects stock comments from Eastmoney Stock Forum as the data source and constructs a sentiment influence coefficient model to explore the relationship between sentiment information and stock price volatility, as illustrated in Figure 1.

This study uses the Eastmoney Stock Forum as the primary data collection source. As a leading stock forum in China, the platform accumulates a vast amount of real-time discussions and exchanges on individual stocks, with a high volume of new posts and replies daily, covering a broad spectrum of markets and sectors. This ensures the data's timeliness and accuracy, effectively reflecting investors' immediate views and sentiment shifts in the stock market.

For the case study, we selected Sany Heavy Industry (SH600031), a leading company in the machinery industry, as the primary focus. To test the generalizability of the method, we also included Seres (SH601127), a leader in the new energy vehicle industry, as an additional research subject. Both companies have market capitalizations in the hundreds of billions and attract substantial comment activity, reflecting significant market attention and in-depth discussion.

Additionally, we utilized the Python libraries from Tushare and Akshare to collect daily stock price and market capitalization data for each company.

Given that the performance of sentiment analysis models heavily depends on the quality of the text data (Mishev et al. 2020), and recognizing the scarcity of sentiment analysis datasets tailored to the financial sector in Chinese, we developed a specialized Chinese dataset for sentiment classification

Table 1 Sentiment Analysis Model Accuracy

Model	Classification Accuracy
LSTM Sentiment Analysis Model	0.88
BERT-base Chinese Sentiment Analysis Model	0.96
BERT-wwm-ext-Chinese Sentiment Analysis Model	0.98

Table Notes.

of stock comments. Specifically, we used web scraping to collect 8,713 historical comments across multiple stocks. These comments were then manually labeled as either positive (4,608 comments) or negative (4,105 comments) to form the training dataset for our model.

2.2. Development of the Chinese Stock Commentary Sentiment Analysis Model

We evaluated the performance of three sentiment analysis models: an LSTM-based model (Shi et al. 2015), a BERT-base-Chinese model (Devlin 2018), and a BERT-wwm-ext-Chinese model (Cui et al. 2021). The dataset was randomly split into training and test sets in a 4:1 ratio. The labeled commentary data were then used to train each of the three models. The training results are presented in Table 1.

The results show that the BERT-wwm-ext-Chinese sentiment analysis model achieved the highest accuracy, significantly enhancing classification precision and efficiency, particularly outperforming the LSTM model.

2.3. Development of the Public Opinion Influence Coefficient Model

To explore the relationship between online sentiment and stock prices, we developed an online sentiment indicator at a microeconomic level, systematically examining its specific effects on stock market volatility. In doing so, we introduced the Public Opinion Influence Coefficient Model.

$$\sum_{1}^{i} \left(f_i^{+} - f_i^{-}\right) = h\left(m_i \times \frac{v_i - v_{i-1}}{v_{i-1}}\right) \tag{1}$$

It utilizes several key variables: $f+$ and $f-$ capture the volume of positive and negative comments, respectively, about a specific stock on the forum between consecutive trading days. The Public Opinion Influence Coefficient (F), calculated as the difference between $f+$ and $f-$, provides a measure of the overall sentiment towards the stock based on these discussions.

m represents the company's outstanding shares, reflecting the number of shares available for public trading (Hong et al. 2006). This variable is included because companies with a larger float tend to exhibit greater stock price stability. A larger float facilitates enhanced corporate governance through increased external oversight and mitigates price volatility, as a greater number of shares are available to absorb market fluctuations (Wang and Xu 2004). v denotes the stock price, while a captures the rate of stock price change as a percentage difference between consecutive trading days. By analyzing the interplay of these variables, the study aims to shed light on how online sentiment may influence stock price fluctuations.

By simplifying the above equation, we derive the following equation (2):

$$\sum_{1}^{i} F_i = h\left(m_i a_i\right) \tag{2}$$

α denotes the rate of stock price change, $a = \frac{v_i - v_{i-1}}{v_{i-1}}$, calculated as the ratio of the difference between stock prices on two consecutive trading days to the stock price on the earlier trading day.

2.4. Time Period Adjustment for the Public Opinion Influence Coefficient

We subsequently collected stock commentary data for Sany Heavy Industry from May 1, 2024, to July 1, 2024, and applied the established sentiment analysis model to assess the sentiment orientation of the collected comments. Using the Public Opinion Influence Coefficient Model (1), we calculated the daily public opinion influence coefficients F and ma for this period and visualized the findings.

The analysis reveals a weak positive correlation between the public opinion influence coefficients F and ma, with F exhibiting a noticeable time lag behind ma in the time series. This observation is consistent with the low correlation and time lag issues widely discussed in existing literature (Moews et al. 2019, 2018). The results are presented in Figure 2.

Given that the daily trading sessions of the Chinese stock market run from 9:30 AM to 3:00 PM, we assume that stock commentary made after the market closes at 3:00 PM until before the next day's opening (i.e., before 3:00 AM) potentially impacts the following day's stock price, with no direct effect on the same day's stock price. Additionally, sentiment accumulated during non-trading days (holidays) affects the stock price on the first trading day thereafter.

To resolve the inconsistency between the trends of F and m, we adjusted the time window for F, redefining it as the period from after the close of the previous trading day (3:00 PM) to before the close of the current trading day (3:00 PM). For the first trading day following a holiday, F is treated as the cumulative influence of public opinion during the holiday period.

Figure 3 illustrates the positive correlation between the rate of change in the selected stock's price and its public opinion information. After adjustment, the trend of the F curve aligns more closely with that of the MA curve, indicating improved synchronization.

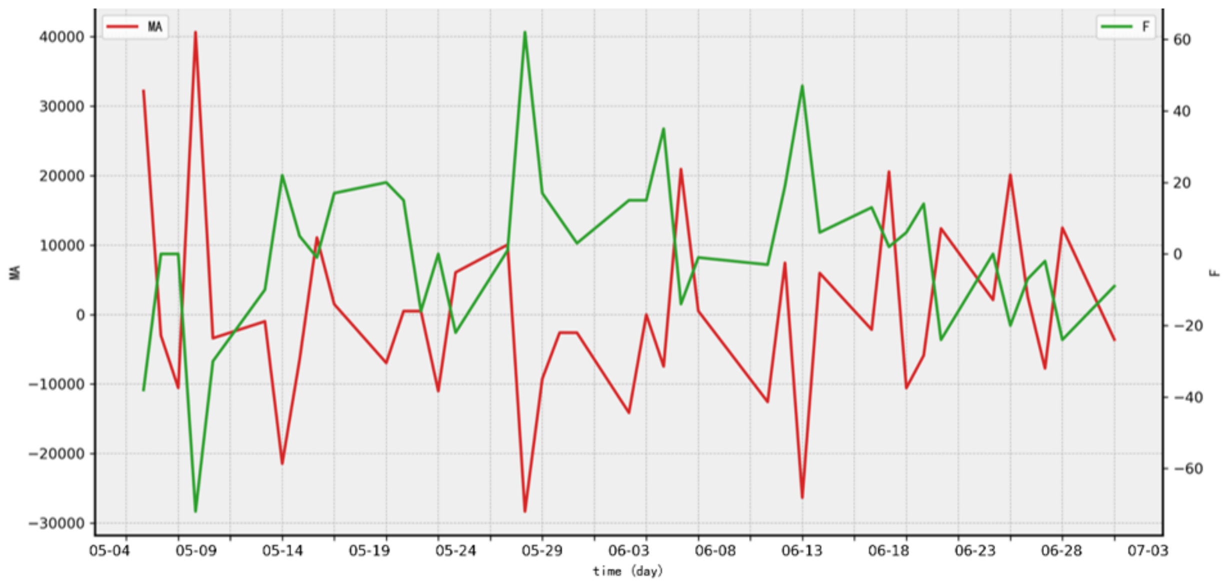

Figure 2 Public Opinion Influence Coefficient F Prior to Time Period Adjustment.

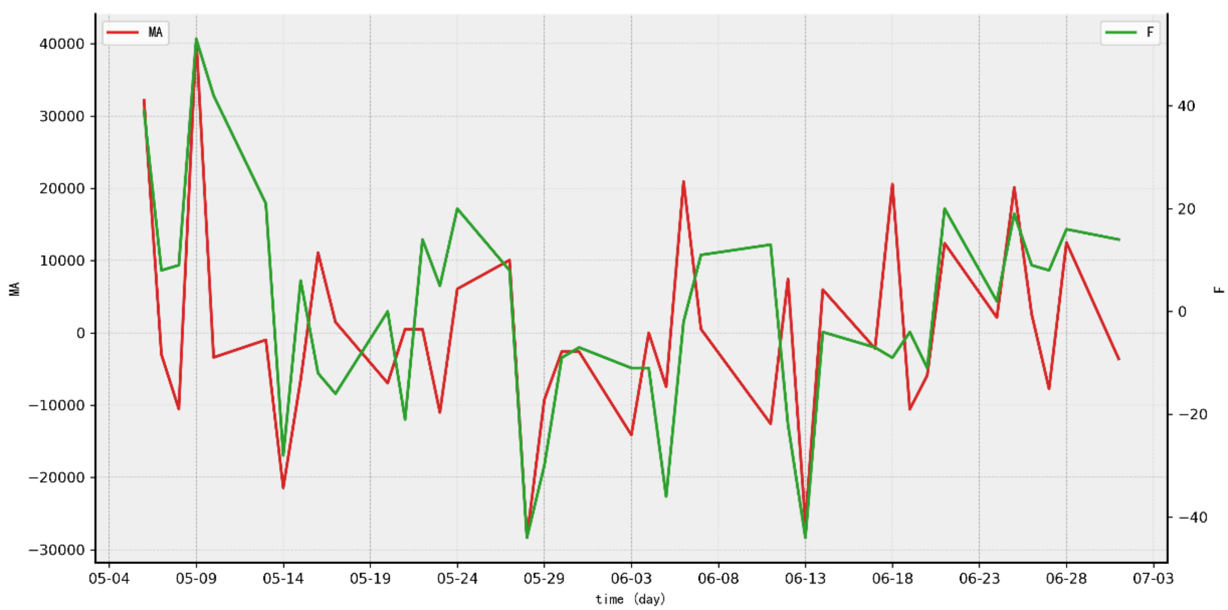

Figure 3 Public Opinion Influence Coefficient F After Time Window Adjustment.

To quantitatively assess the linear correlation between the F curve and the MA curve, we employ the correlation coefficient as the statistical measure. The correlation coefficients between the public opinion influence coefficient F (both before and after adjustment) and the stock price change rate were calculated, and the results are presented in Table 2. Table 2 demonstrates that the positive correlation coefficient between the adjusted public opinion influence coefficient F and MA has

Table 2 Correlation Coefficient between F and MA

Time Period Adjustment	Correlation Coefficient
Adjusted	0.6078
Unadjusted	-0.7685

Table Notes.

Table 3 Prediction Accuracy Across Different Input Time Steps

Input Time Steps	MSE
1 day	0.01408
2 days	0.01371
3 days	0.01045
4 days	0.00548
5 days	0.00786
6 days	0.00916
7 days	0.01291

Table Notes.

significantly increased. This result suggests that the adjusted F curve more accurately reflects the impact of public sentiment on stock price volatility.

3. Stock Price Volatility Analysis

3.1. Stock Price Volatility Forecasting

Predicting stock price movements requires accounting for the persistent effects of sentiment information, indicating that sentiment influences not only the immediate stock price but also continues to affect it over subsequent periods. Therefore, this study employs the XGBoost machine learning framework (Chen and Guestrin 2016) to optimize and fit model, using several days of historical sentiment data F as input features and the current market price MA as the target variable. Given the significant difference in scale between F and MA, normalization was applied to ensure data consistency. By adjusting the input time steps, we obtained the prediction accuracy results shown in Table 3.

Through multiple training iterations, we examined the prediction accuracy across different input time steps and found that a four-day input time step yields the best predictive performance. This result highlights the significant persistence of sentiment information's influence on stock prices, typically lasting around four days.

To further enhance the predictive performance of the XGBoost model, we introduced Bayesian optimization (Snoek et al. 2012) to accurately search for and identify the optimal hyperparameter configuration. Key XGBoost hyperparameters considered include the number of trees ($n\_estimators$), maximum tree depth (max_depth), minimum child weight of leaf nodes

Table 4 Model Hyperparameters and Their Ranges

Hyperparameter	Range
n_estimators	[50, 200]
max_depth	[3, 15]
min_child_weight	[1,10]
learning_rate	[0.01, 0.3]
Subsample	[0.6, 1.0]
colsample_bytree	[0.6, 1.0]

Table Notes.

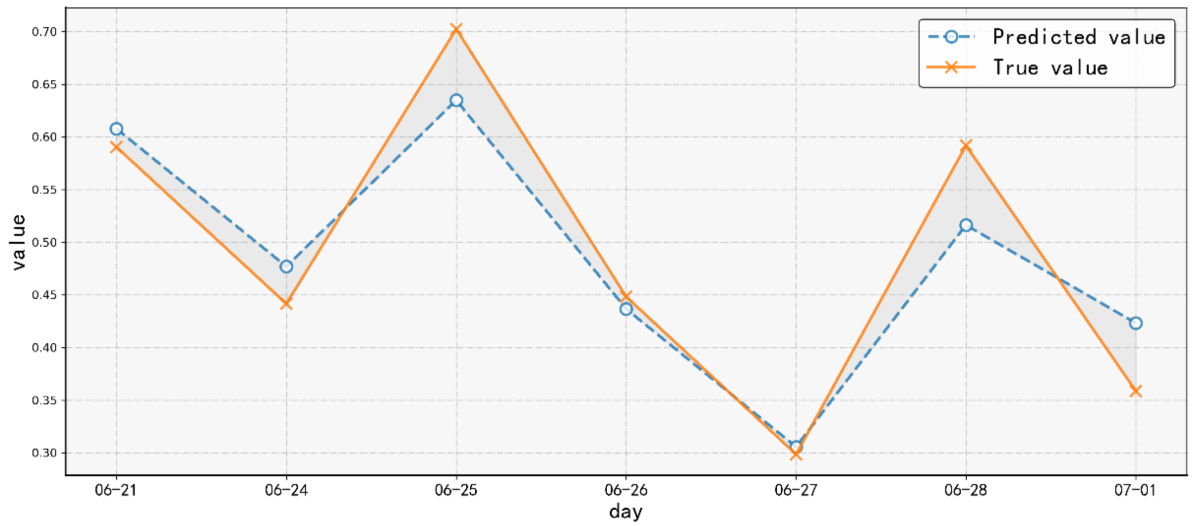

Figure 4 Predicted and Actual Trends of Sany Heavy Industry Stock Price Volatility.

(*min_child_weight*), learning rate (*learning_rate*), subsample ratio (*subsample*), and feature subsample ratio (*colsample_bytree*). Based on prior experience and a comprehensive literature review, we set reasonable search intervals for each hyperparameter (detailed in Table 4) to ensure a thorough and effective optimization process.

We then applied the Bayesian optimization algorithm, performing 30 iterations to determine the optimal hyperparameter configuration for the XGBoost model. The final hyperparameter settings were as follows: the number of trees was set to 90, the maximum tree depth to 7, the minimum child weight to 5, the learning rate to 0.1383, the feature subsample ratio to 0.9866, and the subsample ratio to 0.9689. Using this configuration, we trained the XGBoost model and applied it to the stock price prediction task. The prediction results are presented in Table 5 and Figure 4.

Based on a comprehensive evaluation using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) metrics, the model's predicted values demonstrate strong alignment with the actual stock prices. This strongly validates the model's effectiveness in

Table 5 XGBoost Model Prediction Performance Metrics

Metric	Value
MSE	0.00231
RMSE	0.04811
MAE	0.03990

Table Notes.

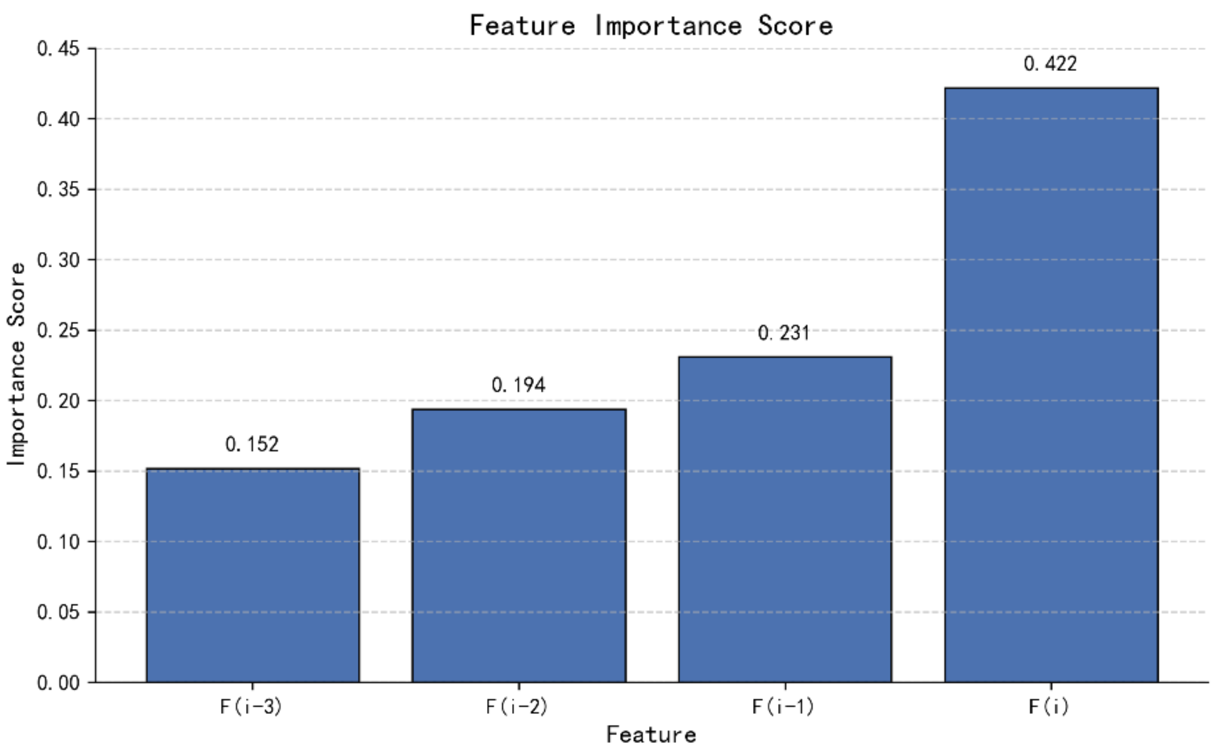

Figure 5 Public Opinion Influence Coefficient Weights.

capturing the influence of market sentiment on stock price movements. Consequently, under normal market conditions, sentiment information serves as a critical factor in accurately predicting the trend and magnitude of stock price fluctuations.

3.2. Public Opinion Influence Weight Analysis

Additionally, by leveraging the feature_importances_ function of the XGBoost model, we quantified the contribution of daily public opinion influence within the input time steps to the model's predictions, represented by importance scores, as shown in Figure 5.

As illustrated in the figure, the weights corresponding to day $i-3$, day $i-2$, day $i-1$, and day i are 0.172, 0.204, 0.241, and 0.382, respectively. This distribution pattern clearly demonstrates the temporal dynamics of public opinion's impact on stock prices: the influence of the current day's public opinion is the most pronounced, while the influence of public opinion from previous days gradually diminishes. This finding not only illustrates the natural decline of public opinion influence over time but also strongly supports the study's conclusion that the impact period of public opinion on stock prices is approximately four days.

3.3. Stock Price Volatility Alert

Accurate prediction of stock price fluctuations requires a thorough understanding of normal price movements. To effectively detect abnormal fluctuations in stock prices and provide timely decision support for investors and regulatory agencies, this study introduces the individual moving range chart (X-MR chart) (Shewhart 1931) from engineering statistics into the development of a stock price volatility alert model. The X-MR chart precisely monitors data distribution and promptly identifies and responds to any changes in the process, making it particularly well-suited for forecasting abnormal stock price volatility.

The individual moving range chart model is structured as follows:

The mean of the individual measurements x is calculated according to equation (3):

$$\tilde{x} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{3}$$

Where x_j represents the jth observation, and n denotes the total number of observations. The mean of the moving range $\overline{R_s}$ is calculated according to equation (4):

$$\overline{R_s} = \frac{\sum_{j=1}^{n-1} R_{si}}{n-1} \tag{4}$$

Where R_{sj} is the moving range of the jth observation, calculated as $|\boldsymbol{x}_j - \boldsymbol{x}_{j-1}|$.

The center line (CL), upper control limit (UCL), and lower control limit (LCL) for the control charts are computed as follows:

(1)X chart: $\mathrm{CL} = \bar{x}$; $UCL = \bar{x} + 2.66\overline{R_s}$; $\mathrm{LCL} = \bar{x} - 2.66\overline{R_s}$

(2)R chart: $\mathrm{CL} = \overline{R_s}$; $\mathrm{UCL} = 3.27\overline{R_s}$; LCL is not applicable

This model detects the following eight types of signals:

Type 1 Warning: A single point exceeds the control limits.

Type 2 Warning: Two out of three consecutive points fall in the outer third of the control limits.

Type 3 Warning: Four out of five consecutive points fall in the outer half of the control limits.

Type 4 Warning: Nine consecutive points fall on the same side of the center line.

Type 5 Warning: Seven consecutive points show a consistent increase or decrease.

Type 6 Warning: Eight consecutive points fall in the outer third of the control limits.

Type 7 Warning: Fifteen consecutive points fall in the middle third of the control limits.

Type 8 Warning: Fourteen consecutive points alternate between rising and falling.

3.4. Empirical Evaluation of the Model

To assess the sensitivity and effectiveness of the stock price volatility warning model, we analyzed the stock price data of leading publicly traded companies Sany Heavy Industry and Seres. This analysis was supplemented with news events and capital flow data during abnormal periods to provide a comprehensive assessment of the detected anomalies.

In this empirical evaluation, we classified stock price anomalies into two categories: ”true anomalies” and ”false anomalies”:

True anomalies: These occur when the warning model accurately detects abnormal stock price fluctuations and issues a warning, but the correlation between public opinion influence and stock price volatility is low. Such instances often suggest the possibility of market manipulation, institutional misconduct, or other irregular activities.

False anomalies: These occur when the warning model also detects and warns of abnormal stock price fluctuations, but the correlation between public opinion influence and stock price volatility is high. This is typically due to significant changes in company financial performance, which trigger widespread market discussion and result in substantial stock price movements influenced by public opinion.

3.4.1. Empirical Analysis of Stock Price Volatility Warnings for Sany Heavy Industry

(1)Analysis of False Anomalies

First, we tested the feasibility of the public opinion influence coefficient model in predicting stock price volatility under normal market conditions using the stock price warning model. Specifically, we analyzed stock price data for Sany Heavy Industry (SH600031) from May 1, 2024, to July 1, 2024, by inputting this data into the warning model. The results are presented in Figures 6 and 7.

The analysis shows that during the observation period, the X-chart did not display any abnormal fluctuations. However, the MR-chart triggered a Type 2 warning on May 8-9, characterized by two out of three monitoring points falling in the outer third of the control limit. This anomaly corresponds to the significant rise in the MA curve on May 9th in the F-MA trend chart shown in Figure 2.

Further investigation revealed that on May 9, northbound capital made a net purchase of Sany Heavy Industry stock totaling 212 million RMB. Additionally, the company’s first-quarter financial

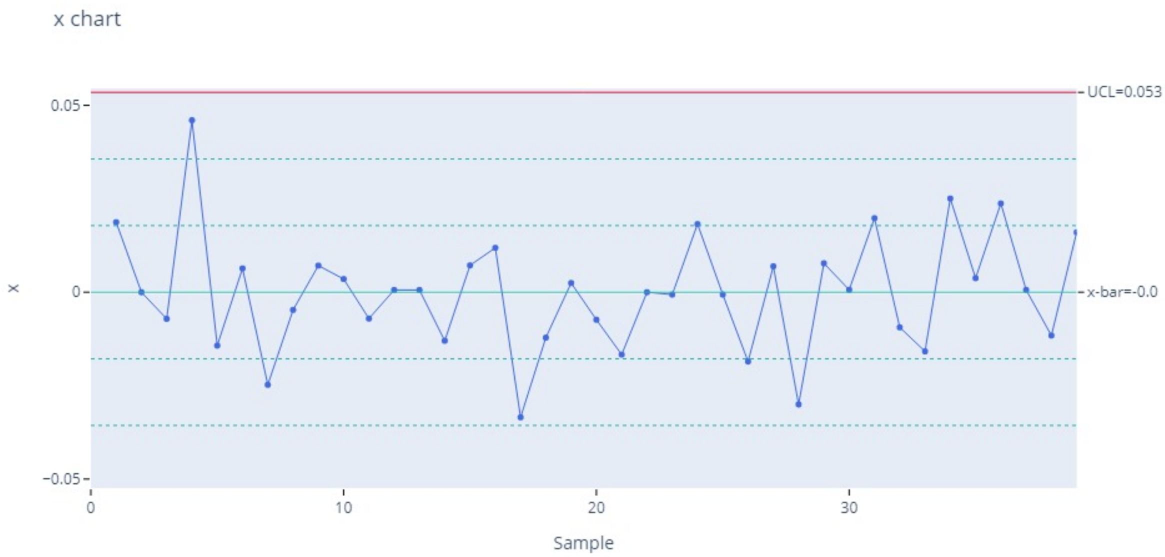

Figure 6 X-Chart for Daily Stock Price Volatility Warnings of Sany Heavy Industry.

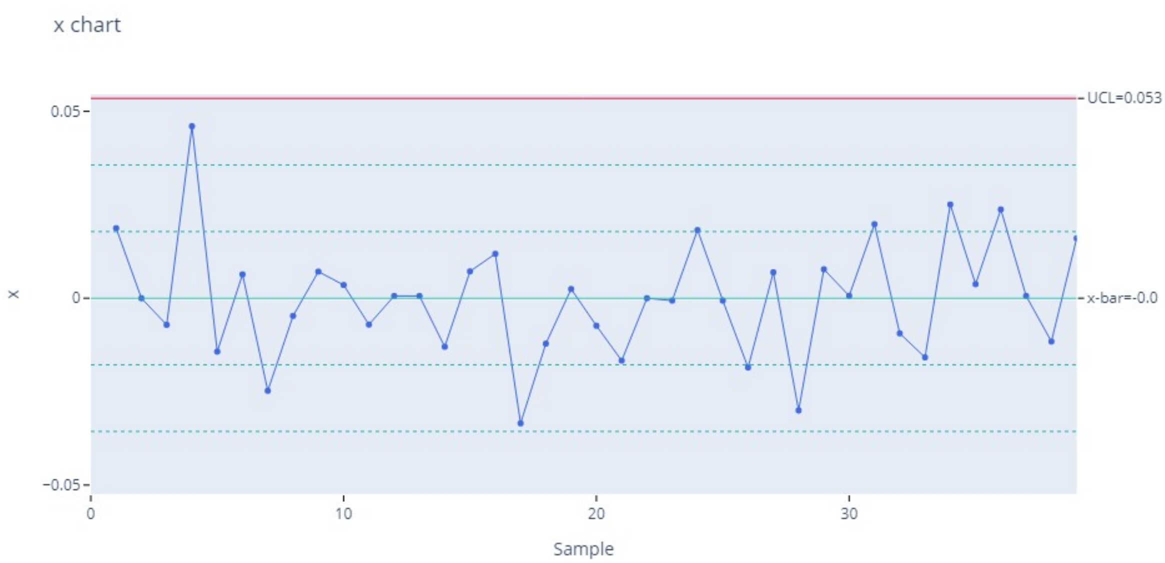

Figure 7 MR-Chart for Daily Stock Price Volatility Warnings of Sany Heavy Industry.

report for 2024, released before the May Day holiday, showed a net profit of 1.58 billion RMB, a year-over-year increase of 4.21%. Furthermore, over 10 institutions provided positive ratings, forecasting that the company's net profit for 2024 would increase by over 30%. These favorable news events significantly boosted investor confidence.

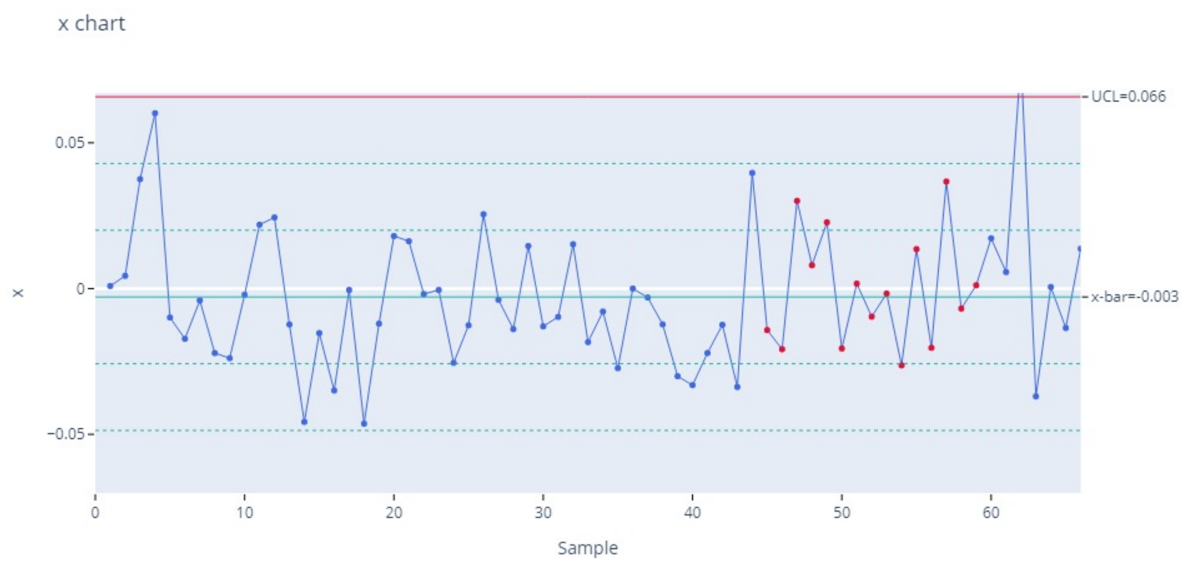

Figure 8 X-Chart for Daily Stock Price Volatility Warnings of Sany Heavy Industry.

Although the stock price exhibited short-term abnormal fluctuations, the strong first-quarter performance led to a substantial increase in the public opinion influence coefficient. Therefore, this fluctuation can be regarded as a normal market reaction to positive signals, qualifying it as false anomalies. Consequently, the public opinion influence coefficient effectively indicated stock price trends during this period, validating its feasibility in predicting stock price volatility under normal conditions.

(2)Analysis of True Anomalies

To further validate the effectiveness of the stock price warning model in identifying true anomalies in stock price fluctuations, we applied the model to historical stock price data for Sany Heavy Industry. The analysis identified abnormal data from mid-March to April 2022. The results are shown in Figures 8 and 9.

As shown in the figures, the X-chart triggered a Type 8 warning signal between March 17 and April 1, 2022, characterized by 14 consecutive alternating upward and downward data points. This phenomenon is visually reflected in the F-MA trend chart in Figure 10, where the F curve starts to lag behind the MA curve, and their trends diverge.

To objectively determine whether the stock price fluctuation was abnormal, we compared the correlation coefficients between the F curve and the MA curve before and after the warning. The results are presented in Table 6, showing that after the warning (March 14 to April 1), the correlation coefficient between the stock price change rate and the public opinion influence

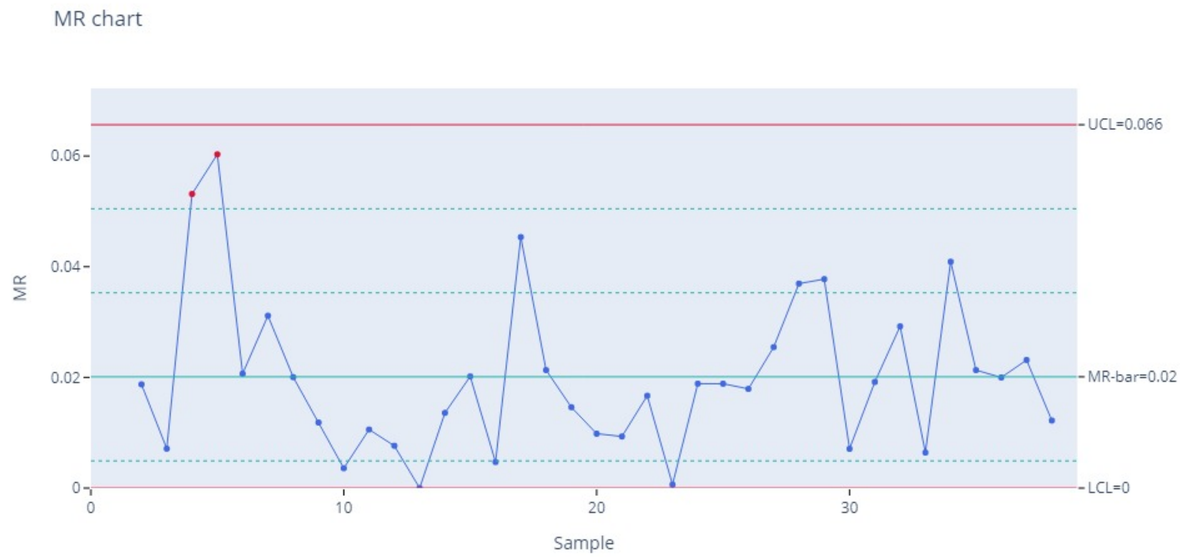

Figure 9 **MR-Chart for Daily Stock Price Volatility Warnings of Sany Heavy Industry.**

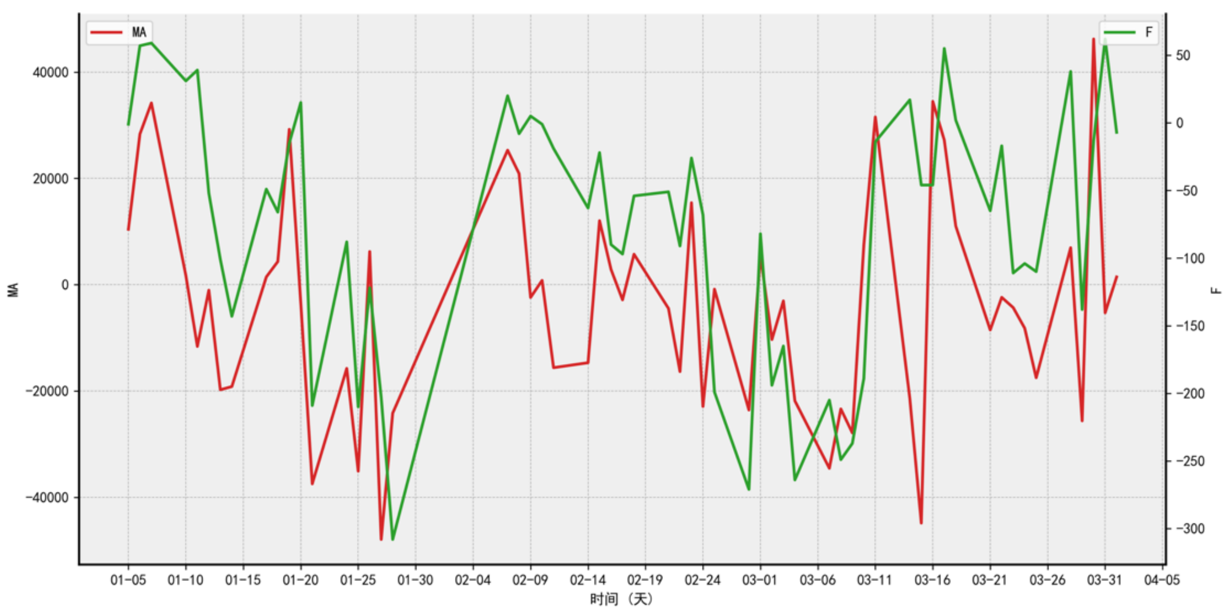

Figure 10 **F-MA Trend Chart of Sany Heavy Industry.**

coefficient significantly decreased. In contrast, the pre-warning correlation coefficient was as high as 0.699, indicating significant anomalies in stock price fluctuations.

To explore the causes of these abnormal stock price fluctuations, we analyzed relevant news, announcements, and stock data. The results showed no significant favorable or unfavorable news or announcements and no disclosure of large-scale selling by major institutions. However, according

Table 6 Correlation Coefficients for Stock Price Warnings of Sany Heavy Industry

Period	Correlation Coefficient
Warning Period (2022.03.14 - 2022.04.01)	0.379
Pre-warning Period (2022.01.05 - 2022.03.13)	0.699

Table Notes.

to the company's stock K-line chart and capital flow data, on March 14, 2022, major funds suddenly withdrew approximately 100 million RMB. Despite this, market investor sentiment remained optimistic, leading to a divergence between public opinion trends and stock price movements.

In the following trading days, the stock price continued to fluctuate around a certain level, with major funds frequently engaging in large-scale inflows and outflows, resulting in a sideways oscillation. This pattern often reflects the investment strategies of major institutions, possibly involving tactics like buying low and selling high to induce less confident investors to sell prematurely, thereby capturing the spread and reducing overall costs, setting the stage for future capital operations.

In this case, the divergence between the public opinion influence coefficient and stock price volatility represents true anomalies. Therefore, it is likely that the trading strategies of major institutions were a key factor leading to the series of anomalies in the company's stock price.

3.4.2. Empirical Analysis of Stock Price Warnings for Seres To verify the generalizability of the stock price volatility warning model, this study conducted an analysis using Seres as a case example. Recently, Seres has gained significant attention due to its AITO series of vehicles, with its market capitalization rising to 111.8 billion RMB, positioning it as a leader among new automotive companies. Therefore, Seres serves as a typical case for analysis.

We first obtained the company's stock price data for June 2024 and input it into the stock price warning model. The results of the individual moving range control charts are shown in Figures 11 and 12.

The analysis shows that from June 3 to June 17, 2024, the MR-chart triggered a Type 4 warning, indicating nine consecutive monitoring points clustered on one side of the centerline. This suggests that the stock price fluctuated within a narrow range during this period. Correspondingly, the F-MA trend chart shows clear lateral oscillations in stock price volatility. Meanwhile, the public opinion influence coefficient rose significantly, while the stock price declined, indicating true anomalies.

Further investigation revealed that on June 2, 2024, Seres released its May production and sales report, showing that the company sold 32,202 vehicles in May, a year-over-year increase of 489.13%. On the same day, the company also held a nationwide delivery ceremony for its new vehicles. These

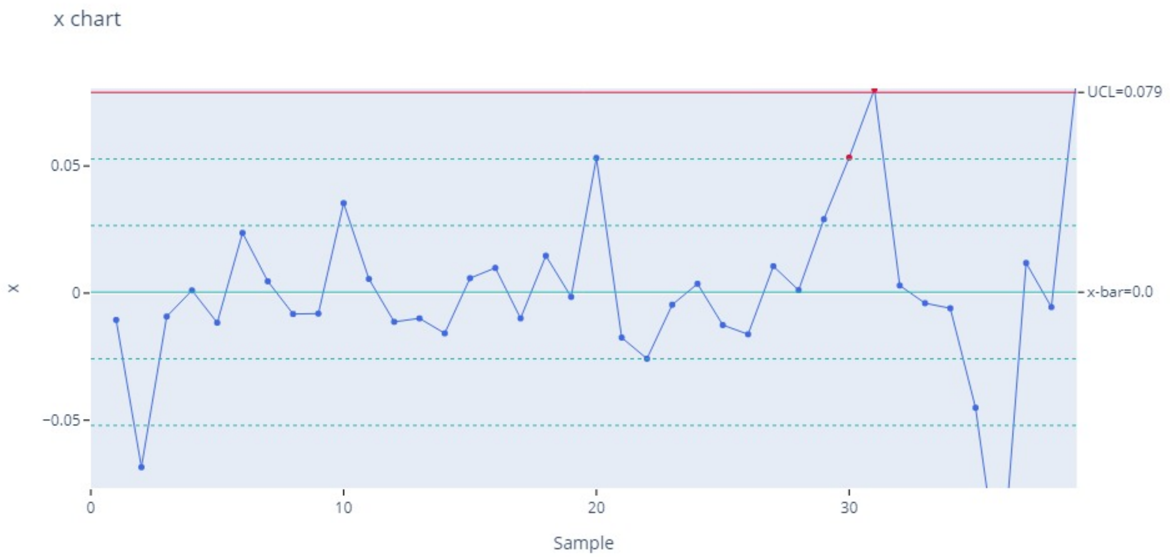

Figure 11 **X-Chart for Daily Stock Price Warnings of Seres.**

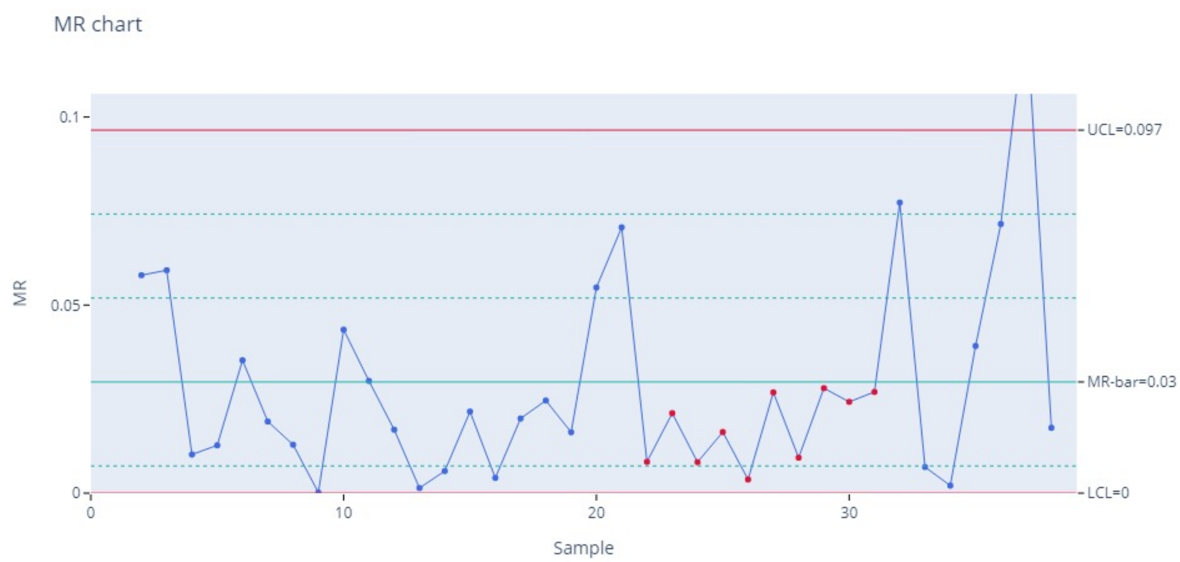

Figure 12 **MR-Chart for Daily Stock Price Warnings of Seres.**

positive developments boosted market sentiment, significantly raising the public opinion influence coefficient on June 3. However, major funds unexpectedly sold over 93 million RMB worth of stock that day, causing the stock price to open high and close low, falling short of expectations.

Despite the ongoing release of positive news and overall bullish sentiment, the stock price showed a sideways oscillation and gradually declined. This suggests that major market players may have

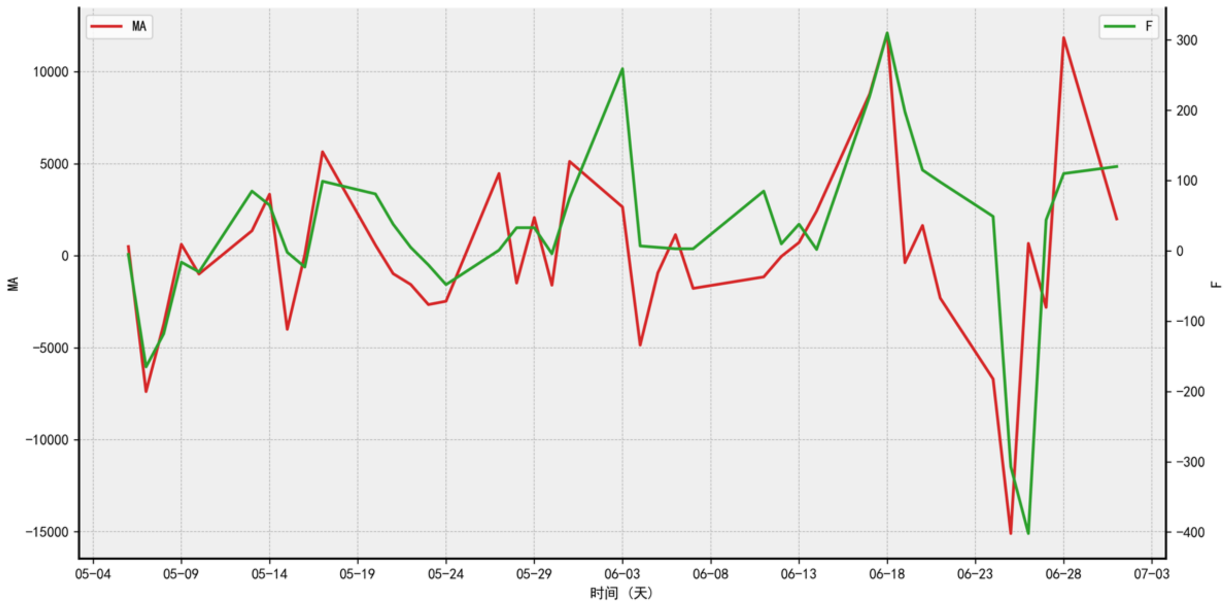

Figure 13 F-MA Trend Chart of Seres Stock.

been using a strategy to buy low and sell high, reducing their holding costs and inducing less confident investors to sell, preparing for a future price increase.

After the Type 4 warning, the stock price triggered a Type 1 warning signal on June 18, June 25, and June 28, with single monitoring points exceeding the preset control limits. Specifically, following the significant correction and limit down from June 18 to June 25, a large amount of low-priced stock became available. Major funds likely considered the stock price to have reached a reasonable level with potential for another increase. On June 28, substantial purchases by major funds drove the stock price significantly higher. Despite the price recovery, investor sentiment did not improve, indicating a lack of confidence in the stock's rebound and continued caution.

The accuracy of these empirical results validates the effectiveness of the proposed stock price warning model in identifying abnormal price fluctuations. The model enhances the ability of investors and regulators to detect abnormal volatility, providing strong support for more precise market analysis and decision-making.

4. Conclusion

This study integrates natural language processing techniques with financial market data to innovatively propose a stock price volatility analysis method grounded in system dynamics. This approach is designed to enhance the accuracy of stock price prediction and warning analysis in environments characterized by frequent abnormal fluctuations, thereby supporting more informed investment decisions and regulatory actions. We began by constructing a Chinese stock commentary sentiment analysis dataset and successfully developed a high-accuracy sentiment analysis model using the advanced BERT framework, which significantly improved the precision of sentiment detection. Furthermore, we introduced a public opinion influence model that substantially increased the correlation between public sentiment data and stock price volatility, leading to greater accuracy in our predictive models. Our analysis revealed that the influence of public opinion on stock prices persists for approximately four days before gradually diminishing, a finding that aligns with observed market behaviors.

Finally, our empirical analysis demonstrated the stock price warning model's exceptional ability to identify abnormal fluctuations. The model not only shows high sensitivity and accuracy in detecting the effects of market manipulation and major news events on stock prices but also provides critical early warnings to investors. This capability holds significant practical value in fostering a more stable and well-functioning market.

References

Audrino F, Sigrist F, Ballinari D (2020) The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting* 36(2):334–357.

Babu CN, Reddy BE (2015) Prediction of selected indian stock using a partitioning–interpolation based arima–garch model. *Applied Computing and Informatics* 11(2):130–143.

Bollerslev T, Engle RF, Nelson DB (1994) Arch models. *Handbook of econometrics* 4:2959–3038.

Chang C, Liu Z, Spiegel M (2015) Capital controls and optimal chinese monetary policy. *Journal of Monetary Economics* 74:1–15, URL `http://dx.doi.org/10.1016/J.JMONECO.2015.04.003`.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Cui Y, Che W, Liu T, Qin B, Yang Z (2021) Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29:3504–3514.

Deveikyte J, Geman H, Piccari C, Provetti A (2022) A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence* 5:836809.

Devlin J (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society* 987–1007.

Granger CW, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis* 1(1):15–29.

Hashmi SM, Gilal M, Wong W (2021) Sustainability of global economic policy and stock market returns in indonesia. *Sustainability* 13:5422, URL `http://dx.doi.org/10.3390/SU13105422`.

He F, Lucey B, Wang Z (2020) Trade policy uncertainty and its impact on the stock market -evidence from china-us trade conflict. *Finance Research Letters* 101753, URL `http://dx.doi.org/10.1016/j.frl.2020.101753`.

Hong HG, Scheinkman JA, Xiong W (2006) Asset float and speculative bubbles. *Journal of Finance* URL `http://dx.doi.org/10.1111/J.1540-6261.2006.00867.X`.

Hosking J (1981) Equivalent forms of the multivariate portmanteau statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 43(2):261–262.

Koesrindartoto DP, Aaron A, Yusgiantoro I, Dharma WA, Arroisi A (2020) Who moves the stock market in an emerging country – institutional or retail investors? *Research in International Business and Finance* 51:101061, URL `http://dx.doi.org/10.1016/J.RIBAF.2019.101061`.

Li J, Cheng L, Zheng X, Wang FY (2023) Analyzing the stock volatility spillovers in chinese financial and economic sectors. *IEEE Transactions on Computational Social Systems* 10:269–284, URL `http://dx.doi.org/10.1109/TCSS.2021.3134487`.

Li T, Ma F, Zhang X, Zhang Y (2019) Economic policy uncertainty and the chinese stock market volatility: Novel evidence. *Economic Modelling* 87:24–33, URL `http://dx.doi.org/10.1016/J.ECONMOD.2019.07.002`.

Li X, Xie H, Chen L, Wang J, Deng X (2014) News impact on stock price return via sentiment analysis. *Knowl. Based Syst.* 69:14–23, URL `http://dx.doi.org/10.1016/j.knosys.2014.04.022`.

Liang X, Chen RC (2005) Mining stock news in cyberworid based on natural language processing and neural networks. *2005 International Conference on Neural Networks and Brain*, volume 2, 893–898 (IEEE).

Liu C, Liu H, Nassar R, Li L (2019) Institutional information manipulation and individual investors' disadvantages: A new explanation for momentum reversal on the chinese stock market. *Emerging Markets Finance and Trade* 57:525 – 540, URL `http://dx.doi.org/10.1080/1540496X.2019.1593825`.

Madhavan A (2002) Market microstructure: A practitioner's guide. *Financial Analysts Journal* 58:28 – 42, URL `http://dx.doi.org/10.2469/faj.v58.n5.2466`.

Mishev K, Gjorgjevikj A, Vodenska I, Chitkushev LT, Trajanov D (2020) Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE Access* 8:131662–131682, URL `http://dx.doi.org/10.1109/ACCESS.2020.3009626`.

Moews B, Herrmann J, Ibikunle G (2018) Lagged correlation-based deep learning for directional trend change prediction in financial time series. *ArXiv* abs/1811.11287, URL `http://dx.doi.org/10.1016/j.eswa.2018.11.027`.

Moews B, Herrmann JM, Ibikunle G (2019) Lagged correlation-based deep learning for directional trend change prediction in financial time series. *Expert Systems with Applications* 120:197–206.

Ren R, Wu D, Liu T (2019) Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal* 13:760–770, URL `http://dx.doi.org/10.1109/JSYST.2018.2794462`.

Romanowski A, Skuza M (2017) Towards predicting stock price moves with aid of sentiment analysis of twitter social network data and big data processing environment. *Advances in Business ICT: New Ideas from Ongoing Research* 105–123.

Sehgal V, Song C (2007) Sops: stock prediction using web sentiment. *Seventh IEEE international conference on data mining workshops (ICDMW 2007)*, 21–26 (IEEE).

Shewhart WA (1931) Economic control of quality of manufactured product. .

Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo Wc (2015) Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28.

Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25.

Wang F, Xu Y (2004) What determines chinese stock returns? *Financial Analysts Journal* URL `http://dx.doi.org/10.2469/faj.v60.n6.2674`.

Wei P, Mao X, Chen X (2020) Institutional investors' attention to environmental information, trading strategies, and market impacts: Evidence from china. *Business Strategy and the Environment* URL `http://dx.doi.org/10.1002/bse.2387`.

Xia W, Guo L (2015) Empirical study on returns, volatility and investor sentiment of the growth enterprise market in china. *2015 3rd International Conference on Education, Management, Arts, Economics and Social Science*, 1290–1292 (Atlantis Press).

Xiang Y (2022) Using arima-garch model to analyze fluctuation law of international oil price. *Mathematical Problems in Engineering* 2022(1):3936414.