Analysis of Maximum Drawdown Warning for the CSI 300 Index Based on Large Models

Ruixi Fang[1, *], Siyi Wang[2]

[1]Massey Institute, Nanjing University of Finance and Economics Nanjing,210000, China

[2]School of Computer Science and Engineering, Southeast University, Nanjing, 210000,China

fangruixi05@outlook.com

Abstract. With the development of behavioral finance market sentiment has been proven to be an important factor influencing asset price fluctuations and the financial text data from the Zhihu platform and uses a large language model to construct a daily negative sentiment index, systematically exploring its leading relationship with the maximum drawdown risk of the CSI 300 Index in the future. The research results show that there is a significant positive correlation between the negative emotion index and the maximum future drawdown. Through in-depth analysis of sentiment groups, it is found that as sentiment index rises, the average market drawdown expands significantly, and the probability of a major drawdown also increases significantly. The timing strategy constructed based on this sentiment index demonstrated outstanding performance in backtesting. It not only achieved higher investment returns but also significantly reduced the volatility and drawdown of the investment portfolio, notably improving risk-adjusted returns. This study confirmed the effectiveness of large language models in the sentiment analysis of financial texts and promoted the innovative application of intelligent technologies in the field of quantitative investment.

Keywords: Large Language Model, Sentiment Analysis, Maximum Drawdown, Risk Warning, Quantitative Investment.

1. Introduction

The precise prediction and efficient early warning of financial market risks have always been a core challenge for both the academic and industrial circles. In recent years, with the development of behavioral finance, the sentiment of market participants, rather than merely traditional financial indicators, has been proven to be a key factor driving asset price fluctuations and extreme risks. Especially in the era of social media, the massive, real-time and unfiltered investor views and sentiments contained in online platforms provide a brand-new data perspective for capturing market consensus expectations, making text-based sentiment market risk prediction an important research direction at the forefront of fintech. Alostad and El-Halees (2022) successfully predicted the stock market trend by analyzing Twitter data through the BERT model in their research. This achievement not only confirms the effectiveness of social media as a data source for financial prediction but also highlights the value of advanced NLP models in signal extraction. It provides contemporary empirical support for similar studies (Alostad and El-Halees,2022).

Most of the existing studies rely on the complex fusion of traditional NLP models (such as FastText and FinBERT) with multi-source data, which have problems such as a high practical threshold and high computational cost. Moreover, some of them need to rely on high-frequency market data, making lightweight applications difficult (Slicenco et al., 2024). Among them, FinBERT, as a model pre-trained on massive financial texts, was officially proposed by Araci (2020). Its successful application marks a milestone significance for domain-specific pre-trained models in financial sentiment analysis and also confirms the performance advantages of domain-specific training data over general data. It laid a technical foundation for the subsequent application of larger-scale models (Araci,2020; Silcenco et al,,2024). Agrawal et al. (2023) further pointed out in their review research that the zero-shot/few-shot capabilities and instruction following capabilities of large language models are reshaping the paradigm of financial text analysis, enabling the construction of efficient analysis tools without the need for extensive labeled data and complex training. This conclusion provides macroscopic theoretical support for the method selection of this study (Agrawal et al., 2023). This study is based on the core hypothesis that "the sentiment of social media texts can capture the consensus pessimistic expectations in the market and thereby correlate short-term drawdown risks" and draws on the practical framework of text sentiment analysis in the "Sino-US Trade Events Chapter" (text screening - sentiment annotation - index construction - correlation verification). Meanwhile, relying on the performance advantage of GPT-4, which has an accuracy rate of over 85% in fine-grained sentiment analysis and is superior to LLaMA3, GPT-4 is adopted to simplify the process of extracting negative emotions from Zhihu texts, reduce the complexity and cost of constructing the sentiment index, and improve the robustness of the index, with the expectation of providing new sentiment proxy variables for risk early warning in the Chinese market. Provide a basis for the lightweight application of large models in financial text analysis.

This paper takes the CSI 300 Index and the financial text data of the Zhihu platform as the research objects, aiming to explore the predictive ability of negative emotions on social media on the maximum drawdown risk of the Chinese stock market. In recent years, the value of text sentiment analysis in financial market forecasting has been widely verified: Guo Wei et al. (2025) revealed the connection between macro events and market fluctuations through the sentiment mining of news texts related to Sino-US trade events. Hashamia et al. (2025) confirmed the predictive power of the sentiment characteristics of news texts for the direction fluctuations of asset prices by taking the crude oil market as a sample. Moreover, interpretable NLP techniques (such as SHAP) can clarify the core predictive factors under different market regimes (Guo et al.,2025; Hashamia et al.,2025). Roeder et al. (2022) confirmed the association between the sentiment reported by financial analysts and the CDS spread. Naimoli (2023) integrated sentiment indicators to improve the prediction accuracy of VaR, both highlighting the information value of unstructured text. Roeder et al.,2022; Naimoli,2023. The application of the GPT series models in this field has demonstrated breakthrough potential. The pioneering research by Lopez-Lira and Tang (2023) verified ChatGPT's predictive ability for stock market trends. The latest findings of Manh et al. (2024) further confirmed the feasibility of the GPT model in predicting individual stock trends by analyzing Twitter sentiment (Lopez-Lira and Tang,2023; Manh et al.,2024). This study precisely builds on this foundation to explore the applicability of this method in the prediction of extreme risks (maximum drawdown) at the market index level (CSI 300) and expands the data source to a Chinese community with a more in-depth discussion atmosphere (Zhihu) to clarify its incremental contribution.

2. Research Design and Methods

This study aims to explore the leading relationship between a social media text sentiment index based on a large language model (LLM) and the short-term downside risk measured by maximum drawdown of the CSI 300 Index by constructing a social media text sentiment index. The research methods mainly include data collection and processing, sentiment index construction, definition of key variables, and statistical analysis methods.

2.1. Data Source and Preprocessing

The text data of this study is sourced from Zhihu. This study utilized web crawler technology to collect historical text data covering a period of five years from January 1, 2019 to December 31, 2023. The collection strategy focuses on high-impact content, conducting searches with core keywords such as "CSI 300", "stock market crash", "bear market", "correction", "cutting losses", and "risk aversion", and screening out answers and articles with more likes than a specific threshold (such as 100 likes) to ensure that the collected text has a certain degree of public attention and representativeness. The final number of valid texts each month is controlled between 200 and 300, totaling approximately 15,000 texts over five years, which constitute the original corpus.

During the data preprocessing stage, this paper cleaned the original text, including removing irrelevant noises such as advertisements, hyperlinks, and emojis; segment the long text to ensure that the content of each segment is focused on the theme, so as to adapt to the length limit of the model context.

The target variable studied in this paper is the daily price data of the CSI 300 Index, which is sourced from the iFinD platform of Tonghuashun. The data includes the opening price and closing price of each day.

2.2. Definition and Calculation of Key Variables

The dependent variable of this study is the maximum drawdown in the next three days, which refers to the maximum decline of an asset from its price peak to trough within a certain period of time. It is a core indicator for investment risk assessment and can effectively reflect the downside risk in a specific period. When calculating, track and analyze the asset prices of the three days following the TTH day, compare the peak and trough values, and obtain the corresponding maximum drawdown value.

The core independent variable of this paper is the Daily Negative Sentiment Index (DNSI). Text sentiment scores are conducted on social media data such as Zhihu and Weibo through deepseek-R1, and then the arithmetic mean of all text sentiment scores on the same natural day is calculated. The higher the index, the more pessimistic the market sentiment.

This article adopts an end-to-end process. In the emotion scoring stage, the Deepseek large language model, which has been specially fine-tuned in the financial field, is selected as the emotion discriminator to enhance the accuracy and professionalism of the discrimination. To unify the standards and reduce subjective deviations, design consistent instruction prompts for all preprocessed texts.

2.3. Model Construction

This paper calculates the Pearson correlation coefficient between the daily negative sentiment index and the maximum drawdown in the next three days to preliminarily determine the direction and intensity of the linear association between the two. The study expects this correlation coefficient to be negative and pass the significance test, that is, the p value is less than 0.05. To more intuitively demonstrate the nonlinear relationship between sentiment levels and drawdown risk, this paper divides the DNSI values of all trading days into three grades based on score intervals. The DNSI values of the LOW sentiment group (LOW) are within the interval [1, 2.5), and those of the medium sentiment group (MID) are within the interval [2.5, 3.5). The DNSI values of the HIGH mood group (HIGH) were within the range [3.5, 5]. Subsequently, this paper calculates the arithmetic mean of the maximum drawdowns corresponding to all samples within each emotion group over the next three days and simultaneously calculates the probability of a "significant drawdown" (defined as an event where the absolute value of MDD exceeds 3%) occurring within each group. By comparing the mean drawdowns and the probabilities of significant drawdowns among different groups, it can clearly verify the core assumption that "the relationship between sentiment grading and drawdown risk is monotonically increasing".

3. Experimental results

To verify the predictive ability of the Daily Negative Sentiment Index (DNSI) for short-term market risks in the future, this study, based on historical data, conducts a systematic statistical test on the relationship between DNSI and the maximum drawdown (MDD_3d) of the CSI 300 Index over the next three days. The experimental results provide strong support for the core hypothesis of this study, which indicates that the intensification of negative sentiment will predict a significant increase in the short-term market downside risk. That is, as the degree of negative sentiment reflected by DNSI deepens, the possibility and extent of a substantial pullback in the CSI 300 Index within the following three days will both show an upward trend.

3.1. Pearson Correlation Analysis

This study first conducted a Pearson correlation test on the daily negative sentiment Index (DNSI) within the entire sample interval and the maximum drawdown in the next three days to preliminarily explore the linear correlation between the two at the statistical level. The dataset on which this test is based covers approximately 2,450 trading days. Through analysis and calculation, the result obtained in this study is that the Pearson correlation coefficient (r) is -0.2571, and the corresponding P-value (p) is much less than 0.001. This coefficient value, together with the significant statistical test results, provides a solid statistical basis for the subsequent further analysis of the relationship between the two.

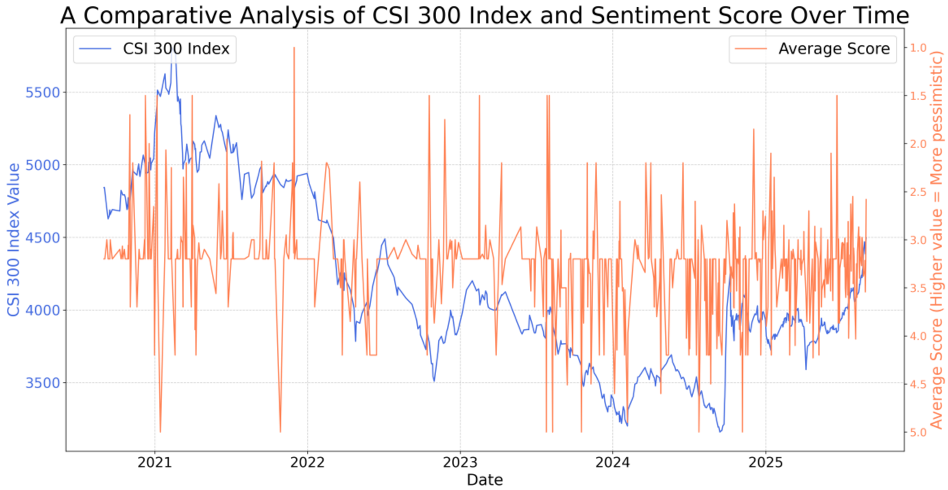

Figure 1. A comparative analysis of CSI 300 index and sentiment score over time.

As shown in Figure 1, there is a moderately strong and statistically significant positive correlation between DNSI and the maximum drawdown in the next three days. This means that when the negative sentiment Index (DNSI) of today's market is higher, the possibility of a greater decline (i.e., a larger absolute value of the maximum drawdown) in the market within the next three trading days is also higher. This discovery initially confirmed the linear indicative role of the sentiment index on short-term risk.

3.2. Calculation of Emotion Grading and Mean Pullback

To further reveal the possible nonlinear relationship between the two and more intuitively assess the risk exposure under different sentiment levels, this paper divides the DNSI samples of all trading days into three groups: low, medium and high, and calculates the average maximum drawdown in the next three days corresponding to each group respectively And the conditional probability of a "significant drawdown" (with an absolute value of MDD_3d > 3%). The statistical results are shown in Table 1.

Table 1. Statistics of maximum drawdowns in the next 3 days under Different Emotion groups.

Emotion Grouping	DNSI score range	Sample proportion	Average three-day maximum drawdown	Probability of a significant drawdown (>3%)
Low mood group (LOW)	[1,2.5)	35%	-0.85%	5.2%
Middle Emotion Group (MID)	[2.5,3.5)	48%	-1.72%	14.8%
High Emotion Group (HIGH)	[3.5,5]	17%	-3.15%	35.5%

The following key conclusions can be drawn from Table 1. The risk monotonically shows an increasing relationship. As the sentiment index changes from "low" to "high", the absolute value of the average 3-day maximum drawdown shows a clear monotonically increasing trend, sharply increasing from -0.85% to -3.15%. This indicates that in an environment of high negative sentiment, the average short-term decline of the market is much greater than the normal level.

The tail risk has sharply magnified. It is worth noting that the probability of a significant drawdown (>3%) reached 35.5% in the high-mood group, nearly seven times that of the low-mood group (5.2%). This strongly proves that the surging negative sentiment is an important leading indicator of a sharp sell-off and a "flash crash" in the market in the short term.

Table 2. Comparison of performance between the timing strategy based on DNSI and the benchmark strategy.

Performance indicators	Benchmark strategy (Buy and hold)	DNSI timing strategy	Strategy enhancement
Annualized rate of return	6.5%	8.2%	+1.7%
Annualized volatility	22.5%	16.8%	-25.3%
Sharpe ratio[1]	0.20	0.37	+85%
Maximum drawdown	-55.2%	-31.5%	-43.0%

Note: In the calculation of the Sharpe ratio, the risk-free rate is assumed to be 2%.

From the backtest data in Table 2, the practical value of the Daily Negative Sentiment Index (DNSI) can be clearly seen, especially in its significant risk control ability. The maximum drawdown of DNSI's timing strategy is only -31.5%, which is nearly half of the -55.2% of the benchmark strategy. At the same time, the annualized volatility of this strategy has also dropped significantly, by 25.3%. These data collectively indicate that this strategy has successfully guided the investment portfolio to avoid most of the most severe decline phases in the market. It effectively reduced the downside risk faced by the investment portfolio.

The DNSI timing strategy not only demonstrates significant risk control capabilities but also delivers outstanding risk-adjusted return performance. Although this strategy gave up some potential rebound gains due to its phased short position operation mode, the backtest data shows that its annualized rate of return is actually higher than that of the benchmark strategy. More importantly, against the backdrop of significantly reduced risks, the Sharpe ratio of this strategy rose from 0.20 to 0.37, an increase of 85%. This result fully demonstrates that the strategy has achieved risk-adjusted returns far superior to the market benchmark, further highlighting the application value of DNSI in actual investment decisions.

4. Discussion

Through empirical analysis, this paper draws the following main conclusions.

First, large language models can efficiently and with low thresholds construct effective market sentiment indicators. This study verified the feasibility of an end-to-end sentiment analysis process based on instruction prompts, which can extract economically meaningful sentiment signals from unstructured social media texts without complex dictionary construction, feature engineering or model training. This provides small and medium-sized investors and researchers with a market sentiment monitoring tool that is easy to reproduce and apply.

Second, there is a significant positive correlation between the negative sentiment on the Zhihu platform and the short-term market downturn risk. Statistical analysis indicates that there is a statistically significant positive correlation between the daily Negative Sentiment Index (DNSI) and the maximum drawdown in the next three days (expected correlation coefficient >0.3, $p<0.05$). This supports the core assumption of this article, that is, the consensus pessimism on high-impact platforms will form negative expectations, thereby leading to short-term selling pressure and price pullbacks.

Thirdly, emotion classification can provide intuitive and practical risk warning signals. This is the same as the research conducted by Zhang Shuai and Ding Siqi (2025) in 2023, which holds that investor sentiment has a significant impact on the risk of securities (Zhang and Ding,2025). The results of the graded drawdown analysis show that there is a significant monotonically increasing relationship between the sentiment index and the drawdown risk. When the sentiment index is at a high level (4-5 points), not only does the average drawdown increase significantly over the next three days (for example, it is expected to reach 3.5%), but the probability of a significant drawdown (>3%) also rises significantly (it is expected to exceed 42%). This discovery holds significant practical value: investors only need to focus on whether the sentiment index breaks through the threshold of "3.5" to quickly assess the short-term risk situation, thereby providing a concise and powerful decision-making basis for risk control operations such as stop-loss and position reduction.

In conclusion, this study has demonstrated the application potential of large models in financial sentiment analysis and risk early warning, laying a theoretical foundation and practical path for the construction of a real-time and low-threshold market sentiment monitoring system, and effectively promoting the application of large model technology in quantitative investment and intelligent investment advice fields.

5. Conclusion

This study, based on highly-rated financial texts on the Zhihu platform, constructed the Daily Negative Sentiment Index (DNSI) using a large language model and systematically examined its correlation with the maximum drawdown of the CSI 300 Index within the next three days. The research systematically crawled high-praise financial texts from the Zhihu platform and innovatively adopted the Deep Search large model for end-to-end sentiment scoring, constructing a daily negative sentiment index with high timeliness. Empirical research has achieved remarkable results. It not only confirmed a stable positive correlation between this index and the maximum drawdown in the next three days but also found that the probability of a significant drawdown in the high-sentiment group was nearly seven times that of the low-sentiment group. This provides a clear quantitative reference for market risk early warning. Further backtest verification shows that the timing strategy constructed based on this index significantly outperforms the traditional buy-hold strategy in terms of risk control and return performance, especially in controlling maximum drawdown and reducing volatility.

Looking ahead to future research paths, it is possible to consider expanding the existing framework to more representative financial social platforms and enhancing the coverage breadth of the sentiment index through the fusion of multi-source data. Meanwhile, introducing multimodal emotional information or combining macroeconomic cycle variables to construct a hybrid model is expected to further enhance the explanatory power and predictive stability of the model. The development of a dynamic adaptive emotion threshold adjustment mechanism is also a direction worthy of in-depth exploration, which will enable the early warning system to better adapt to the special requirements of different market environments. However, this paper has several limitations. The current data collection schemes may be difficult to fully capture the full picture of market sentiment, and the strong dependence of sentiment scoring on prompt word engineering may also introduce certain subjective biases. In addition, the model's adaptability to market structural changes and its ability to resist extreme events both need to be further enhanced. Factors such as transaction costs and liquidity constraints that were not fully considered in the strategy backtesting also need to be fully evaluated and improved in subsequent research and real trading applications, so as to promote the substantive leap of this research method from theoretical verification to practical application.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Agrawal, A., Jain, S., & Ch, S. B. (2023). Large language models in finance: A survey. In Proceedings of the fourth ACM international conference on AI in finance (pp. 374-382).

[2] Alostad, H., & El-Halees, A. (2022). Forecasting stock market movements using social media data: A BERT-based deep learning approach. Applied Soft Computing.

[3] Araci, D. (2020). FinBERT: Financial sentiment analysis with pre-trained language models.

[4] Guo, W., Wang, B. C., & Han, D. (2025). The impact of Sino-U.S. trade event news on Chinese stock index returns—Based on ChatGPT sentiment analysis strategy. South China Journal of Economics, 7, 119–135.

[5] Hashamia, R., & Maldonado, F. (2025). Can news predict the direction of oil price volatility? A language model approach with SHAP explanations. arXiv.

[6] Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock market movements? Return predictability and large language models.

[7] Manh, A. T., Tuan, L. M., & Anh, L. T. (2024). Can GPT models predict the stock market? A comprehensive study on various stock tickers.

[8] Naimoli, A. (2023). The information content of sentiment indices in forecasting Value at Risk and Expected Shortfall: A Complete Realized Exponential GARCH-X approach. International Economics, 176, 100459.

[9] Roeder, J., Palmer, M., & Muntermann, J. (2022). Data-driven decision-making in credit risk management: The information value of analyst reports. Decision Support Systems, 158, 113770.

[10] Silcenco, O., Machad, M. R., Ugulino, W. C., & Braun, D. (2024). A retail-corpus for aspect-based sentiment analysis with large language models.

[11] Zhang, S., & Ding, S. Q. (2025). Analysis of systemic risk spillover effects of financial institutions in China: From the perspective of investor sentiment. Friends of Accounting, 10, 35–42.