ITM Web of Conferences **80**, 04006 (2025) https://doi.org/10.1051/itmconf/20258004006
ACAAI 2025

Multidimensional Investor Sentiment and Stock Market Volatility in China: Measurement, Mechanisms and Applications

Mingxuan Kang[1*]

[1]Starriver Bilingual School, 200083 Shanghai, China

Abstract. This paper systematically investigates the impact of investor sentiment on stock market volatility in China. The rational hypothesis in traditional financial theory is difficult to explain many anomalies in the real market, so behavioral finance provides an important perspective. This paper firstly combed the investor sentiment of measure (direct index method, text, data analysis, multi-source data fusion method) and its correlation model, such as Vector Autoregression (VAR), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Time Varying Parameter Vector Auto-regression (TVP-VAR). On the basis of pointing out the limitations of existing research in terms of data sources, model causality and long-term rule capture, this paper proposes to deepen the mechanism analysis by constructing multi-dimensional sentiment indicators and combining long-term data and behavioral finance theory. This study not only provides a new perspective for understanding market volatility, but also provides practical enlightenment for regulators' risk warning and investors' strategy optimization. Future research should focus on integrating alternative data sources and developing cross-market sentiment contagion models to further enhance predictive accuracy and practical application value.

1 Introduction

In traditional financial theory, "rational investor hypothesis" and "efficient market hypothesis" have long been dominant [1]. Their core logic holds that market prices have fully reflected all available information, and investors' behavior always follows the rational decision-making paradigm. However, in the real financial market, investors' behaviors are often not rational, and investors' decisions are often driven by irrational factors such as psychological factors and group behavior. Extensive investor sentiment fluctuations directly affect market price trends in many scenarios [2]. The development of behavioral finance is based on this observation, emphasizing the influence of "herding effect", "anchoring effect", "overconfidence" and other irrational factors on market prices [3]. This feature is particularly prominent in China's capital market: on the one hand, the high proportion of retail investors in China's capital market and the rapid dissemination of information lead to an extremely significant impact of investor sentiment on market fluctuations [4]. On the other hand, the

* Corresponding author: mingxuankang983@gmail.com

immediacy of information dissemination on the Internet further accelerates the diffusion and amplification of sentiment [5]. The extreme market in recent years, such as the 2015 crash, during the outbreak of market volatility, show that investor sentiment has become an important market influence ingredient. Therefore, it is necessary to quantitatively analyze the mechanism of investor sentiment on stock market volatility based on systematic review of historical research and explore its far-reaching impact on market stability and investment prediction.

There have been relevant studies on the relationship between investor sentiment and stock market volatility in the academic world. Some studies are based on sentiment data from specific channels, which is difficult to fully reflect the overall market sentiment. For example, studies based on Baidu search index found that sentiment has a leading effect during periods of intense market fluctuations. Other scholars use Oriental wealth stock review build sentiment index, prove that investor sentiment has significant predict effect to stock fluctuations, but the sample is confined to a single platform, and effectively eliminate the noise in the comments, lead to emotional measurement accuracy than. And existing research more focused on short-term samples or specific event period, elusive long-term regularity and the phase difference [6]. For example, through Weibo sentiment analysis, the predictive power of sentiment on the characteristics of China's stock market has been verified, but the ability of the model to explain the economic mechanism is weak. The research combined with BERT and TVP-VAR model reveals the time-varying impact of sentiment on volatility and liquidity, especially in the downward market [7, 8]. However, the short sample period makes it difficult to cover long-term structural changes. In addition, some studies have found that sentiment can significantly improve the predictive power of stock price crash risk, but most of them focus on extreme events and ignore the fluctuations in daily situations [9].

In general, the existing literature proves the important role of investor sentiment, but there are generally the following limitations: single source of sentiment data, lack of multi-dimensional coverage; The short sample period makes it difficult to capture the rules of long-term and different market stages; The research method is biased towards empirical prediction, and the mechanism analysis is insufficient. The systematic research on the Chinese market is relatively limited. The main contribution of this paper is that it is the first time to fuse multi-dimensional data to construct Chinese market sentiment indicators. It covers a longer time period and multiple market stages to enhance the robustness of the results. The empirical results are combined with behavioral finance theory to explain how emotions drive volatility; And put forward the idea of sentiment monitoring and risk management that can be used by regulators and investors [10].

Aiming at the existing problem of the study, the innovative contributions of this paper is mainly two aspects in the theory and practice. Theoretical level, through the study of the quantitative measure of investor sentiment, for the application of behavioral finance in China's capital market situation provides new empirical support, help rich emotions drive share price fluctuation theory. At the practical level, through the analytical framework and conclusions proposed in this paper, regulators can establish a risk early warning mechanism with the help of sentiment monitoring, so as to intervene in time at the beginning of systemic risk brewing. Of course, investors can also optimize their investment decisions based on the rules between sentiment and volatility, which can not only avoid the risks caused by group irrationality but also find potential market opportunities.

In summary, this paper aims to reveal the internal relationship between investor sentiment and market volatility through data-driven methods and provide a new decision-making reference perspective for market participants.

2 Related work

The research on the relationship between investor sentiment and stock market volatility is mainly based on the effective quantification of the abstract psychological concept of investor sentiment. The academic community has developed a variety of measurement methods, and on this basis, various models have been used to analyze the relationship between investment sentiment and market volatility.

2.1 Measurement methods of investor sentiment

Existing research mainly uses three types of methods to construct sentiment indicators, each of which has its own advantages and disadvantages. This section introduces the three types of methods.

First, direct indicator method. This method uses market transaction data as a proxy variable for sentiment, which is widely used because of its easy access and strong objectivity. Common indicators include turnover rate, reflect market trading activity and investors' attention; The discount rate for closed-end funds, traditionally seen as a direct measure of retail investor sentiment; New shares listed on the yield and the number of ipos, the market sentiment, increase new issues and premium levels. However, as an indirect reflection of sentiment, such indicators contain more noise and are vulnerable to the interference of other market factors, such as liquidity and policy changes, so it is difficult to purely represent psychological expectations.

The second is text big data analysis. With the information explosion in the Internet era, this method has become the mainstream of current research. It conducts sentiment analysis on massive unstructured text data, such as news, social media posts, stock bar comments, to directly capture investors' opinions. The specific technical path can be divided into: (1) dictionary method: relying on the pre-built financial sentiment dictionary, such as Dalian Institute of Technology sentiment vocabulary, the text is segmented, matched and weighted by sentiment value, to generate sentiment index. This method is computationally efficient, but it is difficult to cope with complex semantics such as metaphor and irony in financial texts, such as "cutting meat" means loss, and "eating noodles" means bleak market. (2) Machine learning method: Supervised learning model, such as Support Vector Machine (SVM), Long and Short-Term Memory network (LSTM), and pre-training model are used to train the annotated emotion text, to build a classifier to identify the emotion of new text. For example, Bidirectional Encoder Representations from Transformers (BERT) model is used to deeply mine text semantics, which significantly improves the accuracy of emotion measurement. Although this method has high accuracy, its model is like a "black box", with weak interpretation of economic meaning and highly dependent on a large amount of high-quality annotated data.

Third, multi-source data fusion method. In order to overcome the bias of a single data source and comprehensively capture the overall market sentiment, scholars often use statistical methods such as principal component analysis (PCA) or factor analysis to reduce the dimensionality of the above direct and indirect indicators. To synthesize a Composite Investor Sentiment Index. For example, the BW index constructed by Baker and Wurgler. For the Chinese market, there are also studies that attempt to integrate web search indexes, such as Baidu index, social media sentiment, and market transaction data to construct more robust sentiment proxy variables.

2.2 Research methods on the relationship between investor sentiment and market volatility

Based on quantitative sentiment, researchers use a variety of econometric models to test the association between sentiment and market volatility.

The Granger causality test is often used to preliminarily determine whether sentiment is a leading indicator of market volatility. The vector autoregressive (VAR) model and its impulse response function (IRF) and variance decomposition (VD) can analyze the dynamic interaction and contribution between them. To better describe the volatility agglomeration of financial time series, GARCH family models (such as Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) and Generalized Autoregressive Conditional Heteroskedasticity in Mean (GARCH-M)) are widely used. By introducing sentiment variables into the conditional variance equation, the predictive ability of sentiment on volatility can be directly tested.

To capture the time-varying characteristics of the relationship, the time-varying parameter vector autoregressive model (TVP-VAR) is introduced. For example, the research combining BERT and TVP-VAR reveals that the impact of investor sentiment on market volatility and liquidity is stronger during bear markets. Considering that there may be regime switching in the market, the Markov regime switching model (MS-VAR) can be used to study the difference in the effect of sentiment in different market states (bull market, bear market and shock market). In addition, quantile regression can reveal the asymmetry of the influence of sentiment under different levels of market volatility, such as extreme crashes or stable periods.

3 Limitations & Applications

Although the existing research has been advancing in theory and method has profoundly revealed the impact of investor sentiment on the stock market, there are still many limitations. A clear understanding of these limitations is the key to promote the deepening of research and application.

3.1 Limitations

In terms of the limitations of data, the first is the single source of data and the lack of representativeness. Most studies rely on data from a single platform, such as using only Baidu search index, or focusing on Weibo platform. Different groups have different channels of information acquisition and expression, so it is difficult to capture the sentiment of the whole market comprehensively and unbiased by a single source, and the universality of the conclusion is questionable. Secondly, the noise and semantic complexity of text data pose great challenges. Non-standard terms, irony, and context-dependent features in web texts limit the measurement accuracy of traditional dictionary-based methods, while deep learning models require high cleaning and annotation costs. Finally, the hybridity of emotion definition is often ignored. Studies often do not clearly distinguish whether emotions are specific to individual stocks or the overall market, and whether they are post-event complaints or pre-expectation, which may lead to confusion in causal inference.

In terms of the limitations of models and methods, "emphasizing correlation and ignoring causality" is the core problem. Although a large number of empirical studies have confirmed the significant association between emotions and fluctuations, it is still difficult to establish a strict causal relationship due to the problem of endogeneity. Most studies have a short sample period, or focus on extreme event Windows such as the 2015 stock market crash and the epidemic. Although strong effects can be found, it is difficult to capture long-term rules and structural changes across multiple bull and bear cycles, and the robustness of the conclusions needs to be strengthened. In addition, the explanatory power of the economic mechanism of the model is weak. Although the prediction performance of complex machine learning models is excellent, it is difficult to clearly explain whether emotions are transmitted to prices through specific behavioral channels such as "herding effect" or "overreaction", and the integration with behavioral finance theory is not deep enough.

In terms of the limitations of the connection between theory and application, the existing research is mostly focused on academic discussion, and how to transform into an effective systemic risk warning tool remains to be studie. How to set the threshold of sentiment index? How does it correlate with other macroeconomic indicators? There is a lack of relevant research on these issues related to regulatory practice. At the same time, the research mainly focuses on A-shares, and the exploration of the heterogeneous impact of cross-market (stocks, bonds and futures), cross-industry and different market value sectors is insufficient, which is difficult to guide refined investment.

3.2 Application situation

Despite the limitations, the research results in this field have gradually shown the potential to be translated into real productivity.

In the field of quantitative investment and asset management, many hedge funds and quantitative teams have incorporated sentiment indexes into their multi-factor models for asset allocation, industry turnover or market selection. Some financial data service providers (such as Wind and Flush) have also launched market sentiment dashboards based on online public opinion, providing professional investors with intuitive sentiment monitoring tools to assist them in judging the degree of short-term market overheating or panic.

In the field of risk early warning and financial supervision, the application prospect is particularly broad. In theory, regulators can build a multi-dimensional and high-frequency investor sentiment monitoring system, which can be incorporated into the macro-prudential assessment framework (MPA) or systemic risk early warning system. When the comprehensive sentiment index continues to be extremely optimistic or pessimistic, it can trigger early warning signals and provide data support for regulators to take counter-cyclical adjustment measures, such as risk warning and transaction supervision, to prevent systemic risks caused by irrational group behaviors. In addition, the research conclusions are also used for investor education, helping investors understand their own behavior patterns and promoting the formation of rational investment culture by revealing the relationship between cognitive bias and market volatility.

For listed companies, sentiment analysis also has value management significance. Investor relations departments of some listed companies can monitor the public opinion sentiment of their stocks on platforms such as stock bar and Snowball to timely understand the market's expectations and doubts about the company's strategy and performance, to more effectively extract information to manage market value fluctuations.

4 Suggestions

4.1 Deepening and expanding the data level

In the future, breakthroughs should be made in the data basis of emotion measurement to build a more comprehensive, accurate and efficient emotion recognition system.

(1) Promote multi-source data fusion: In order to alleviate the bias of single data source, it is suggested to integrate traditional market data, such as turnover rate, IPO information, text data (news, social media, research reports) and new behavioral data such as search index, APP usage data, to build a comprehensive emotion assessment framework. Multi-modal learning and other deep learning technologies can be used to integrate information from

different sources to form a more robust market sentiment index and fully grasp the changes in investor sentiment.

(2) Strengthen text semantic understanding: In the face of the particularity of Chinese financial texts, sentiment analysis needs to break through the limitations of traditional dictionary methods. Pre-training models more suitable for the financial field, such as optimizing FinBER, can be developed to improve the understanding of professional expressions such as "bull market" and "thunder burst" through self-supervised learning. At the same time, new technologies such as cue learning can be introduced to improve the recognition ability of complex linguistic phenomena such as irony and metaphor on the premise of reducing the dependence on annotated data.

(3) Refine the division of emotional dimensions: emotional indicators should not be limited to the binary distinction of "optimise-pessimism" but should distinguish the overall market sentiment from individual stock sentiment, short-term sentiment from long-term confidence, and identify different emotional states such as fear, greed and uncertainty. This refined measurement is helpful to more accurately analyze the impact mechanism of different emotions on market volatility.

4.2 Innovation and improvement of models and methods

In terms of research methods, we should focus on solving problems such as causal identification, dynamic relationship captures and model interpretability.

(1) Pay attention to causal inference and mechanism analysis: in order to overcome endogenous interference, natural experiments can be constructed with the help of exogenous events (such as sudden policy adjustment and international major events), and causal inference methods such as DID and regression discontinuity can be used to identify the real impact of emotions on fluctuations. Can also be through the intermediary effect analysis to verify "flock effect" and "overconfidence" behavior whether mechanisms play a role in emotional influence, enhance the theory persuasive.

(2) Develop more flexible dynamic models: although existing models, such as TVP-VAR and MS models, can capture certain time-varying characteristics, they can still further integrate their advantages, such as constructing a model combining time-varying parameters with regime transition (TVP-MS-VAR) to capture both long-term structural transformation and short-term smooth change. In addition, mixed frequency data model (MIDAS) can be tried to more effectively combine high frequency sentiment data with low frequency fluctuation data to improve the timeliness of prediction.

(3) Improve model interpretability: In the face of the "black box" problem of machine learning models, interpretable artificial intelligence technologies, such as SHAP and LIME, should be promoted to clarify which words or indicators play a key role in sentiment judgment and fluctuation prediction, so as to make the decision-making process of the model more transparent, and effectively combine with behavioral finance theory.

4.3 Promote practical application and cross-field cooperation

It is an important goal of future research to make research results truly serve market practice and financial regulation.

(1) Building a systemic risk early warning system: It is suggested that regulators cooperate with academia and technology enterprises to develop a real-time emotion monitoring and early warning platform based on big data and artificial intelligence. By scientifically setting the sentiment threshold, an early warning can be issued when extreme optimism or pessimism occurs in the market, providing a basis for macro-prudential

management and counter-cyclical regulation, to realize early risk identification and early intervention.

(2) Develop sentiment-driven investment and risk control strategies: in the field of asset management, the differentiated performance of sentiment factors in different industries, sectors and market cycles can be explored, and more adaptive industry rotation or Smart Beta strategies can be designed. At the same time, sentiment indicators can be incorporated into VaR, CVaR and other risk control models to more accurately evaluate the potential losses under extreme market emotions.

(3) Strengthen cross-market and international research: the existing research mainly focuses on A-share market, and can be expanded to bond market, foreign exchange and commodity market in the future to explore the contagion and linkage effect of sentiment among different assets. In addition, studying the spillover impact of investor sentiment between China and the US from a global perspective is of great value to understand the transmission of international financial risks and improve the efficiency of asset allocation.

In short, the research on investor sentiment and market volatility is still in constant development. In the future, deepening data fusion, innovating research methods and expanding practical applications will not only improve the scientific and practical nature of research in this field, but also provide solid support for financial market stability, effective supervision and rational investment, and fully reflect its academic value and practical significance.

5 Summary

This paper aims to analyze the relationship between investor sentiment and stock market volatility in China. Studies have shown that in China, the proportion of retail investors is high, and information dissemination is fast, so the impact of sentiment on the market is particularly significant. Although existing research has confirmed the important role of sentiment, there are still limitations such as single data source and weak causal inference. The innovation of this paper is to advocate the integration of multi-source data to construct a comprehensive sentiment index, and use long-term data combined with frontier econometric model to reveal the behavioral mechanism of emotion-driven fluctuations more comprehensively and deeply. Finally, this study not only enriches the empirical research of behavioral finance in the Chinese market at the theoretical level, but also provides practical application value and reference basis for regulatory early warning and investment decision-making.

Further research could incorporate alternative data sources, develop cross-market sentiment contagion models, and explore regulatory applications in real-time monitoring systems.

References

1. M. Baker, J. Wurgler, Investor sentiment in the stock market. J. Econ. Perspect. **21**, 129-151 (2007) https://doi.org/10.1257/jep.21.2.129
2. Y. Zhou, Z. Chen, Z. Dong, Social media sentiment and stock market anomalies: Evidence from Chinese microblogs. J. Behav. Finance. **18**, 130-145 (2017) https://doi.org/10.1080/15427560.2017.1308362
3. L. Wang, Y. Li, C. Lin, Text-based sentiment analysis and volatility forecasting in emerging markets. J. Empir. Finance. **61**, 1-19 (2021) https://doi.org/10.1016/j.jempfin.2020.101919.

4. X. Chen, Y. Xu, J. Dai, Investor sentiment and stock market volatility: Evidence from China. J. Financ. Planet Econ. **135**, 212-229 (2020) https://doi.org/10.1016/j.jfineco.2019.06.003
5. Q. Zhang, J. Wang, Y. Chen, Investor sentiment and market liquidity: A TVP-VAR analysis using deep learning. J. Bank. Finance. **136**, 106-118 (2022) https://doi.org/10.1016/j.jbankfin.2021.106118.
6. J. Yang, K. Zhang, L. Sun, The role of investor sentiment in the cross-section of stock returns: Evidence from Chinese social media. Pacific-Basin Finance J. **62**, 101386 (2020) https://doi.org/10.1016/j.pacfin.2020.101386
7. P. Li, R. Wang, Q. Feng, Investor sentiment and stock market anomalies: A machine learning approach. Expert Syst. Appl. **185**, 115595 (2021) https://doi.org/10.1016/j.eswa.2021.115595
8. D. Wu, H. Zhang, L. Mei, Investor sentiment and stock market liquidity: Evidence from high-frequency data. J. Int. Financ. Mark. Inst. Money. **73**, 101339 (2021) https://doi.org/10.1016/j.intfin.2021.101339
9. S. Guo, T. Huang, N. Wu, Market sentiment and extreme stock returns: A nonlinear analysis. Econ. Lett. **195**, 109450 (2020) https://doi.org/10.1016/j.econlet.2020.109450
10. L. Zhang, W. Zhao, Y. Liu, Investor sentiment and stock market anomalies: A comparative study of developed and emerging markets. J. Comp. Econ. **50**, 487–502 (2022) https://doi.org/10.1016/j.jce.2021.12.005