WEP Warwick Evans Publishing

Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 9
GBMS 2024

Quantitative Statistical Study of Financial Market Sentiment on Economic Cycles: An Analysis Based on the FinBERT Model and TVP-VAR

Weiran Zhang*, #, Xinmeng Zhang#, Yixin Chen#

School of Economics and Management, Yanbian University, Hunchun, China, 133300

* Corresponding Author Email: nikiiiQ@163.com

#These authors contributed equally.

Abstract. Amid global financial market turmoil, the relationship between market sentiment and macroeconomic cycles has garnered significant attention. This study leverages big data from financial markets to quantitatively analyze market sentiment using the FinBERT model and investigates its impact on macroeconomic cycles with the TVP-VAR method. Based on textual data from the Shanghai Stock Exchange Index forums and Baidu Index online engagement metrics, the study employs GIS technology to analyze regional emotional responses to financial market fluctuations and economic activity trends.The research reveals significant regional differences in China's financial sentiment index during 2022-2023, with hotspots in the eastern coastal regions and cold spots in the west. Economically developed areas exhibit higher sensitivity to market fluctuations. TVP-VAR analysis indicates that changes in market sentiment have a minor impact on macroeconomic cycle volatility, typically exerting a mild negative effect at year's end, though the effects are not significant. This study unveils the dynamic relationship between financial market sentiment and macroeconomics, demonstrating the potential of using social media and online data for macroeconomic analysis. It offers practical recommendations for policymakers on leveraging market sentiment data for forecasting and regulating the macroeconomy, fostering interdisciplinary development in economics and financial engineering.

Keywords: Financial Market Sentiment; Macroeconomic Cycles; FinBERT; TVP-VAR; ArcGIS.

1. Introduction

In the digital age, characterized by an explosion of information, the volatility of financial markets is influenced not only by traditional economic indicators but increasingly by changes in public sentiment, which have become key factors affecting market trends. Individual stock investors' emotions are often swayed by the information fluctuations on the East Money Forum. Compared to experienced institutional investors, these individuals are more susceptible to the emotions and opinions propagated on the forum. According to the "2017 Personal Investor Status Survey Report," 99.76% of investors in China are individuals, and their primary sources of information are smartphone-based social platforms. However, these individual investors typically struggle to profit through trading, in stark contrast to the high profit ratios of institutional investors. This phenomenon highlights the significant disparities in information quality between individual and institutional investors [1]. China's stock market is predominantly driven by individual investors, offering a unique perspective for studying the impact of personal investor sentiment on financial markets.

This study explores how to systematically collect and analyze big data from the East Money Forum and search engines and how these data serve as forward-looking indicators for predicting macroeconomic cycles.

Existing research indicates that market sentiment analysis, which studies the emotions and sentiments of market participants, is crucial for predicting market trends and asset price fluctuations. Text mining methods use natural language processing techniques to analyze financial texts for sentiment information. Common tools include sentiment dictionaries, machine learning algorithms, and deep learning models. For instance, Zhu Zhen and Jiang Wenlu (2016) demonstrated a correlation between individual investor sentiment and market returns using tools like the Yu'E Bao Sentiment Index [2]. Ye Luping (2022) applied the Bert + Bi-LSTM model to classify investor sentiment in the East Money Forum's vast text data and used a multiple linear regression model to reveal a relationship between forum sentiment changes and stock market fluctuations [3]. Wang Haina and Wang Zelong (2023) constructed a financial market sentiment index for China and used the SVAR model to empirically test the spillover effects of the Fed's unexpected unconventional monetary policy on this index [4].

With the advancement of text mining and deep learning technologies, extracting sentiment data directly from social media content has become a new trend in studying investor sentiment. These technologies enable researchers to extract sentiment data directly from financial news, stock forums, and social network posts, rather than relying solely on traditional financial indicators. For example, Li Helong et al. (2023) classified extracted text sentiments based on financial text data types, introducing a framework for text sentiment analysis and reviewing related research outcomes on the impact of text sentiment on financial markets [5]. Pan Junyu (2022) used multiple machine learning algorithm models to train and study the sentiment tendencies of financial market sentiment data collected from Tushare, determining SVM (Support Vector Machine) as the final classification algorithm model [6].

Furthermore, social media data's application in economic forecasting has become a research hotspot, including market sentiment prediction, consumer behavior analysis, and public opinion monitoring.Cheng Wanying (2020) studied consumer purchase intentions for green apparel through social media analysis [7], while Wan Yan et al. (2024) explored the spread and management strategies of university network public opinion in the new media era [8].

GIS technology can analyze and visualize geographic spatial data, helping researchers discover spatial distribution patterns and correlations, providing new perspectives for economic research. For instance, Cui Luming et al. (2021) analyzed the spatial distribution characteristics of emotions in Shanghai based on Sina Weibo data, using deep learning methods to develop a spatial emotion perception evaluation method based on social media check-in data [9].

This paper leverages a substantial corpus of textual data extracted from the highly representative Shanghai Stock Exchange Index forum on the East Money website, coupled with the Baidu Index to construct measures of online engagement. Additionally, it utilizes Geographic Information System (GIS) mapping technology to analyze regional responses to financial market sentiments. Employing the Time-Varying Parameter Vector Autoregression (TVP-VAR) model, the study reveals the impact of fluctuations in financial market sentiments on macroeconomic indicators. The findings demonstrate the potential of using social media and online data for macroeconomic analysis and offer practical recommendations for policymakers on utilizing market sentiment data for forecasting and regulating the macroeconomy, fostering interdisciplinary development in economics and financial engineering.

2. The basic funamental of BP neural network

2.1. Sentiment of stock forum comments based on FinBERT

The study utilizes the Chinese FinBERT model optimized specifically for the financial news domain (referred to as FinBERT_ch) to enhance the accuracy and applicability of sentiment analysis. The FinBERT_ch model is adapted from BERT (Bidirectional Encoder Representations from Transformers), with fine-tuning specifically aimed at capturing the semantic and emotional nuances of the financial sector. Developed by hw2942, the model is available on the Hugging Face platform(https://huggingface.co/hw2942/bert-base-chinese-finetuning-financial-news-sentiment-v2). FinBERT_ch leverages the architectural strengths of BERT and, by utilizing a vast amount of domain-specific data for pre-training, effectively captures the subtle emotional variations in financial texts.

Additionally, the FinBERT_ch model employs BERT's deep bidirectional Transformer encoding mechanism, which can comprehensively understand the linguistic details in the context, thus more accurately determining the emotional tendencies of texts. This feature is particularly crucial for handling the complex, multi-level, and subtle emotional expressions found in financial contexts. By using FinBERT_ch for sentiment analysis, this study effectively integrates weekly financial sentiment indices with quarterly GDP forecasts to explore the potential impact of financial sentiments on macroeconomic cycles.

2.1.1. Data preprocessing

The study utilizes PyCharm to scrape comment data from the Shanghai Stock Exchange Index forum on the "East Money Website-Stock Bar," gathering information such as post titles, post view counts, post comment counts, post URLs, comment reply numbers, comment like counts, comment texts, commenter nicknames, comment times, comment IPs, text, time, sentiment, and sentiment scores. The sample data spans from June 1, 2021, to May 14, 2024, for subsequent sentiment analysis. The stock forum comments obtained directly contain a large amount of irrelevant and redundant information, necessitating preprocessing.

The study involves extracting and cleaning text information, standardizing the extracted text by removing duplicate advertising posts, blank posts, invalid characters, and meaningless spaces. It converts full-width characters to half-width, removes special symbols from internet slang (such as "@", "#", "【】"), and converts all characters to lowercase to eliminate data redundancy and inconsistencies. Subsequently, based on standardized scores of comment counts, view counts, and the number of replies and likes for each comment derived from the original text data, a further selection is made to retain 75,222 valuable entries, totaling 5,373 effective comments.

The mathematical expression for text cleaning is as follows, where f represents a series of text cleaning functions applied consecutively:

Cleaned Text

= f(Original Text)

= lower(convert(strip(symbols(Original Text))))

Text Extraction: Extract the "comment text" column from the Excel file, which serves as the basis for subsequent sentiment analysis.

2.1.2. Sentiment analysis

The core of the study involves utilizing the FinBERT_ch model, which is an optimization of the BERT model specifically for Chinese text in the financial domain. The steps are as follows:

①Model Loading

Load the pre-trained FinBERT_ch model, which has been fine-tuned on financial news text to cater to the sentiment analysis needs of the financial sector.

②Sentiment Determination

Input the cleaned text into the FinBERT_ch model for sentiment classification. The model will output the emotional tendency of the text, such as positive, negative, or neutral.

The functional expression for the final sentiment classification is as following:

Sentiment Score,Label= FinBERT(Cleaned Text)

2.1.3. Results summary and analysis

After the analysis, aggregate the sentiment scores by time dimension to explore the potential impact of sentiment changes on economic indicators. Specifically, calculate the average sentiment score at each time point:

$$S_d = \frac{1}{N_d} \sum_{i=1}^{N_d} \text{Score}_i \quad (1)$$

Where S_d is the average sentiment score for a specific date, and N_d is the total number of comments on that date.

2.2. Regional economic sentiment variations revealed by GIS mapping

To delve into the regional differences in economic sentiments across China's provinces, this study utilized the ArcGIS tool for visualizing the spatial distribution of financial sentiment indices. Geographic Information System (GIS) mapping allowed for an intuitive display of emotional responses in economic activities across various regions, thereby identifying potential economically active areas and their trends. The research incorporated the Baidu Index as a crucial tool for measuring public attention, selecting seven keywords directly related to the financial markets: stocks, stock market, A-shares, stock software, stock account opening, stock trading, and investment management. Using these keywords, we collected Baidu Index data from July 1, 2022, to December 31, 2023, covering six quarters. Given that stock forums opened their IP addresses on June 1, 2022, the study focused on the fourth quarter of 2022, the second quarter of 2023, and the fourth" quarter of 2023 as the periods of interest to obtain the average financial sentiment index and the average online attention levels of each region for these quarters. These data not only reflect the public's interest and attention to the financial markets but also reveal the dynamics of regional financial market focus at different time points.

Utilizing ArcGIS for visualization, the study not only delineates the spatial distribution of financial sentiments across provinces but also analyzes how these sentiments evolve over time. By comparing GIS maps across different quarters, it is possible to observe the provinces' emotional states, both positive and negative, in response to economic expectations or significant events in the financial markets. This comprehensive spatial and temporal analysis offers new perspectives and deep insights into regional economic development differences. Such insights are crucial for formulating regional economic policies and financial market strategies.

2.3. Impact of Social Media Sentiment on Macroeconomic Cycles

In the empirical section of the article, the Time-Varying Parameter Vector Autoregression (TVP-VAR) model analysis was conducted using OxMetrics6 software [10]. All models were set with a lag order of one, employing the Markov Chain Monte Carlo (MCMC) algorithm to perform 1000 sampling iterations on the model data. After conducting unit root and cointegration tests, the data were log-transformed for measurement purposes. The study utilized GDP as a representative variable for economic cycles and applied the TVP-VAR model to analyze the relationship between weekly financial sentiment indices and GDP. The timeframe for the study spanned from January 1, 2022, to March 31, 2024, exploring the applicability and effectiveness of this high-frequency indicator in predicting changes in low-frequency economic cycles. The GDP data were sourced from the National Bureau of Statistics.

In the dynamic state-space model framework, y_t is a $k \times 1$ vector of observed variables, c_t is a $k \times 1$ vector of state variables, $B_{i,t}, i = 1 \cdots s$ are $k \times k$ matrices of time-varying coefficients, and u_t is a $k \times 1$ vector of innovations, assumed to follow a normal distribution Ω_t.

$$y_t = c_t + B_{1,t} y_{t-1} + \cdots + B_{s,t} y_{t-s} + u_t \\ t = s + 1, \cdots n. \tag{2}$$

Transform the covariance matrix Ω_t to $A_t \Omega_t A_t' = \Sigma_t \Sigma_t'$.

Assuming A_t is a lower triangular matrix.

$$A_t = \left(\begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1,t} & \cdots & a_{kk-1,t} & 1 \end{pmatrix}\right) \tag{3}$$

Σ_t is a diagonal matrix.

$$\Sigma_t = \left(\begin{pmatrix} \sigma_{1t} & 0 & \cdots & 0 \\ 0 & \sigma_{2t} & 0 & \vdots \\ \vdots & 0 & \vdots & 0 \\ 0 & \cdots & 0 & \sigma_{kt} \end{pmatrix}\right) \quad (4)$$

Transform equation (1) into

$$yt = ct + B1, tyt - 1 + \cdots + Bs, tyt - e + A_t' \textstyle\sum_t \varepsilon_t, t = s + 1, \cdots n \quad (5)$$

$$V(\varepsilon_t) = I_n, \ \text{Let } X_t' = I_n \otimes [1, y_{t-1}', \cdots y_{t-s}'] \quad (6)$$

Transform equation (2) into

$$y_t = X_t' B_t + A_t' \textstyle\sum_t \varepsilon_t \qquad t = s + 1, \cdots n \quad (7)$$

B_t is a time-varying coefficient,and both parameters A_t and Σ_t are time-varying parameters. The TVP-VAR model mainly refers to the methods of Primiceri (2005) and Nakajima Jouchi (2011), making the following assumptions:

Let α_t represent the vector in A_t where elements are neither 0 nor 1, $\alpha_t = (\alpha_{21}, \alpha_{31}, \alpha_{41}, \cdots \alpha_{k,k-1})'$.For the estimation of A_t, define$h_t = (h_{1t}, \cdots h_{kt})'$,where $h_{jt} = \log\sigma_{jt}^2$,$j = 1, \cdots k$,$t = s + 1, \cdots n$,and it is assumed that the parameters follow a random walk.

$$\beta_{t+1} = \beta_t + \mu_{\beta t} \quad (8)$$

$$\alpha_{t+1} = \alpha_t + \mu_{\alpha t} \quad (9)$$

$$h_{t+1} = h_t + \mu_{ht} \quad (10)$$

$$\begin{pmatrix} \varepsilon_t \\ \mu_{\beta t} \\ \mu_{\alpha t} \\ \mu_{ht} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & \sum_\beta & 0 & 0 \\ 0 & 0 & \sum_\alpha & 0 \\ 0 & 0 & 0 & \sum_h \end{pmatrix}\right), t = s + 1, \cdots n \quad (11)$$

$$\beta_{t+1} \sim N(\mu_{\beta_0}, \textstyle\sum_{\beta_0}), \alpha_{t+1} \sim N(\mu_{\alpha_0}, \sum_{\alpha_0}), h_{t+1} \sim N(\mu_{h_0}, \sum_{h_0})$$

To simplify the model estimation,assumng $\sum\beta$、$\sum\alpha$and$\sum h$ are Diagonal matrixes.Set the initial values of the model, let $\mu_{\beta_0} = \mu_{\alpha_0} = \mu_{h_0} = 0$, $\sum_{\beta_0} = \sum_{\alpha_0} = \sum_{h_0} = 10 \times I$, and assume the prior distribution of the model as:

$(\sum_\beta)_i^2 = \text{Gammma}(20{,}0.01)$,$(\sum_\alpha)_i^2 = \text{Gammma}(2{,}0.01)$,$(\sum_h)_i^2 = \text{Gammma}(20{,}0.01)$.

3. Results

3.1. Time series of investor sentiment index

The study, as shown in Figure 1, presents the daily frequency financial sentiment index series within the sample period, constructed using Stata 17.

From Figure 1, the daily financial sentiment index time series exhibits significant volatility. The index fluctuates around the zero axis without a clear long-term trend, indicating no sustained improvement or deterioration in market sentiment. The fluctuation range is between -1 and 1, with some periods showing larger amplitudes, reflecting market instability potentially influenced by major events or economic data. Frequent short-term fluctuations demonstrate market sensitivity to daily news, indicating quick investor reactions. From early to mid-2022, larger fluctuations were observed, slightly converging from mid-2022 to early 2023, but increasing again by mid-2023, indicating greater uncertainty. Monitoring this index is valuable for understanding market dynamics and risk management.

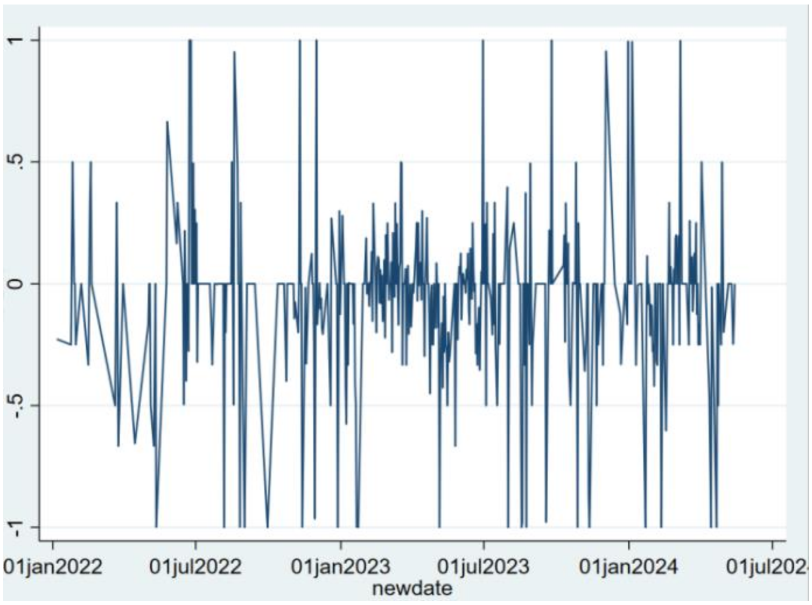

Figure 1. Time Series of the Financial Sentiment Index

3.2. GIS spatial analysis results display

Processed using ArcGIS 10.8.1, the map boundaries reflect the most recent administrative divisions as of January 2024.The GIS spatial analysis results are presented in Figure 2, where the left side shows the GIS maps of online attention for the fourth quarter of 2022, the second quarter of 2023, and the fourth quarter of 2023. The right side depicts the corresponding financial sentiment indices, reflecting spatial correlations and the distribution of sentiments across different regions. White layers indicate areas with missing data.

During 2022 and 2023, the "Heihe-Tengchong Line" (also known as the "Hu Line") was considered the demarcation line for the financial sentiment index. During this period, hotspots and sub-hotspots were primarily distributed in Liaoning and the eastern coastal regions. Cold and sub-cold areas were mainly concentrated in the western and northwestern regions. Notably, the areas of hotspots and sub-hotspots showed a decreasing trend, while cold and sub-cold areas remained relatively stable.

In 2022, hotspot and sub-hotspot areas grew rapidly. Liaoning emerged as a hotspot, while Heilongjiang, Jilin, Hubei, Anhui, and Sichuan became sub-hotspots. In the second quarter of 2023, there was an alternation of cold regions and hotspots, with increases in sub-cold areas, general regions, and cold areas, and a reduction in hotspot areas. Liaoning transitioned from a hotspot to a general area, Henan, Hebei, Sichuan, and Fujian shifted from sub-hotspots to general areas, and Hubei moved from a sub-hotspot to a sub-cold area. Heilongjiang, Jilin, Guangdong, Hunan, Jiangxi, Shaanxi, Shanxi, and Anhui shifted from general to sub-cold areas, while Inner Mongolia, Gansu, and Guizhou moved from sub-cold to cold areas.

By the fourth quarter of 2023, cold areas had contracted but remained concentrated, tropical areas decreased, and sub-hotspot, general, and sub-cold areas increased. There were frequent alternations between different regions, with more pronounced concentrations of cold and hotspot areas, and hotspots shifting eastward. During this period, Inner Mongolia, Shanxi, and Guangdong transitioned from cold areas to sub-cold areas; Henan and Anhui moved from sub-cold to general areas; Shandong shifted from a general to a hotspot area; and Zhejiang was newly classified as a general area.

Overall, the financial sentiment index exhibited distinct spatial differentiation characteristics. During the fourth quarter of 2022 and the second and fourth quarters of 2023, with the Hu Huanyong Line as a boundary, high attention and financial sentiment indices were concentrated eastward, while cold and sub-cold areas concentrated westward, reflecting a transition from hot to cold from east to west.

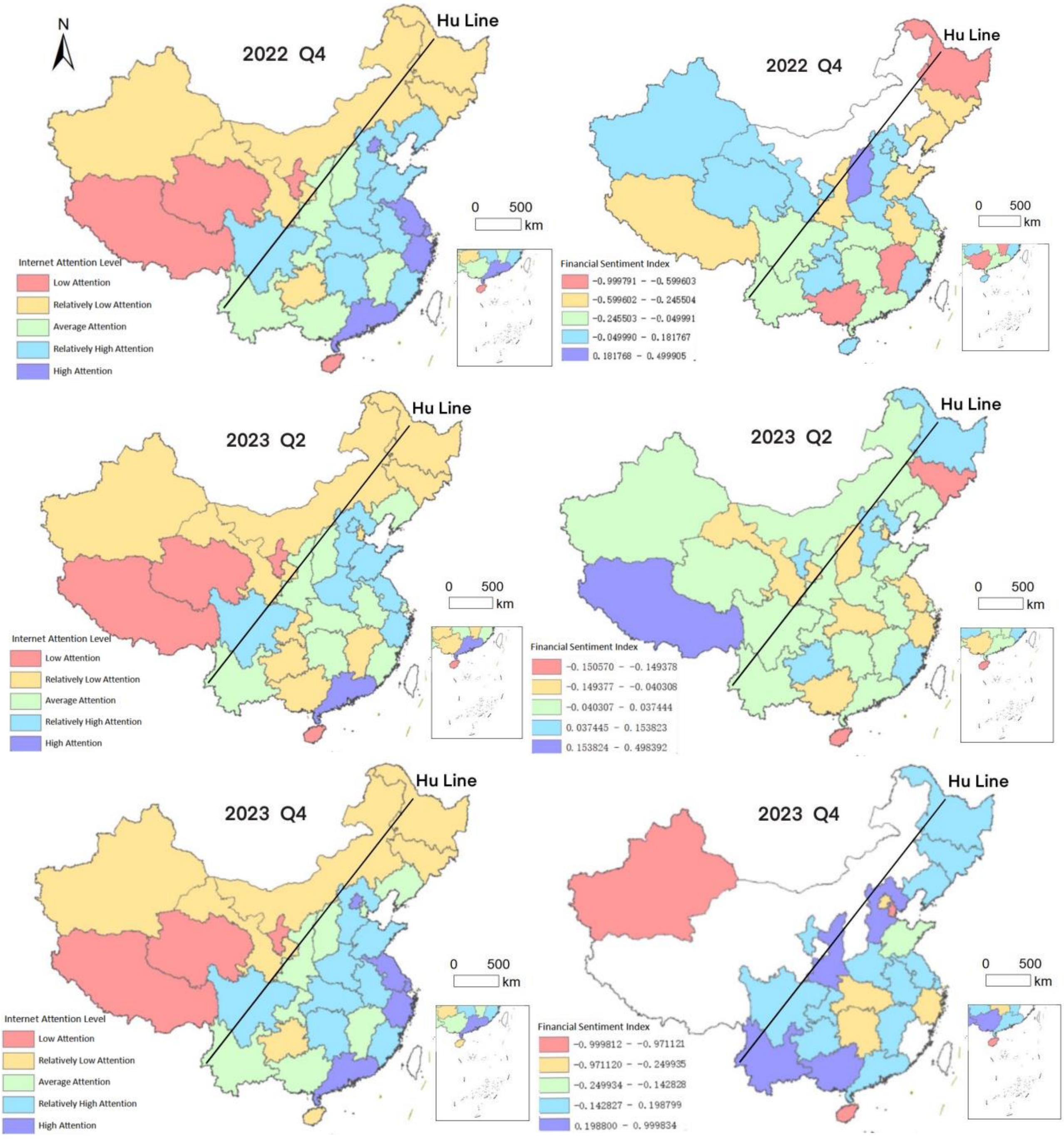

Figure 2. GIS Map of National Q3 Online Attention and Financial Sentiment Index

3.3. Measuring the impact of financial sentiment cycles on economic cycles

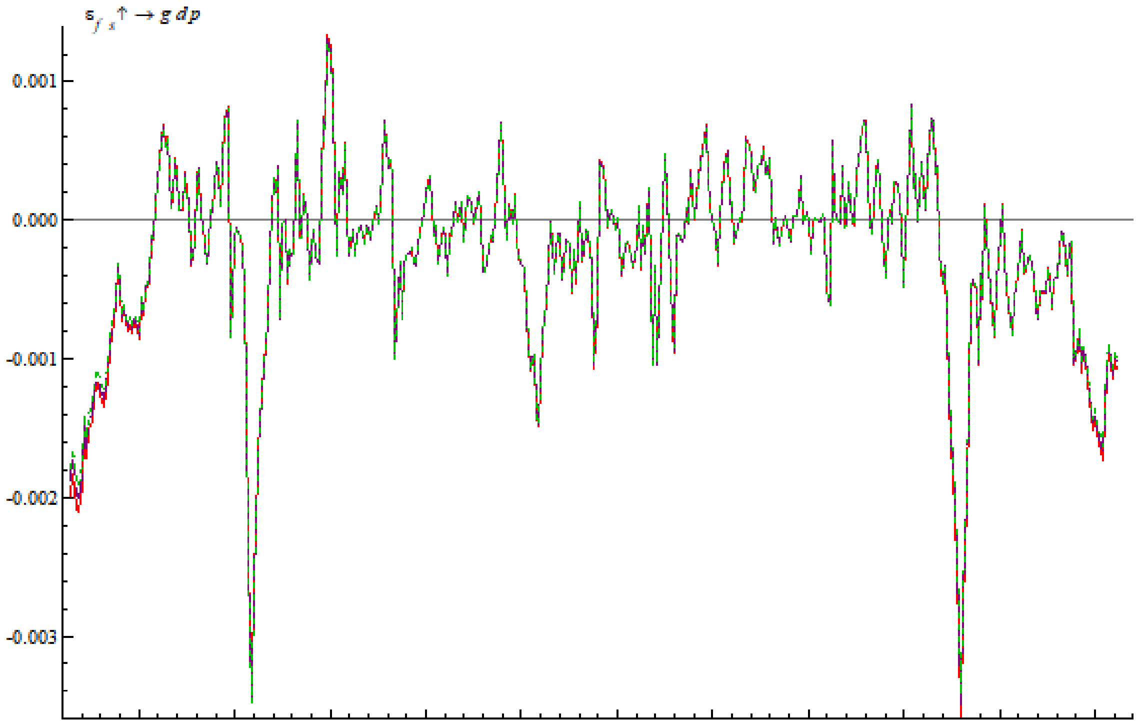

Figure 3. Impulse Response Function of the Financial Sentiment Index on China's GDP Over the Full Sample Period

Figure 3 illustrates the impulse response function of the financial sentiment index on China's total output over the full sample period from January 2022 to March 2024. The graph depicts that the total output is influenced by the financial sentiment index, generally exhibiting an initial negative impact followed by persistent fluctuations. The negative shocks peaked at the end of 2022 and 2023; however, the overall positive and negative impacts were not significant, with weak influence. The lag periods align with the volatility patterns, indicating that the impacts were not assimilated. The transmission of shocks from financial market sentiment typically manifests short-lived effects that are difficult to digest within a brief period.

4. Conclusions

This study explores the relationship between financial market sentiment and macroeconomic cycles in China from 2022 to 2024, using big data collection via PyCharm and analysis with FinBERT_ch deep learning technology. Spatial analysis through ArcGIS and the Financial Sentiment Index revealed distinct spatial characteristics of financial market sentiment, impacting economic cycles. From 2022 to 2023, changes in the Financial Sentiment Index influenced regional financial market activities and overall economic output, though the overall impact on total output was weak and slow to digest.This research addresses the gap in understanding the relationship between financial market sentiment and macroeconomic dynamics, offering insights into their interplay. It recommends further exploration of other influencing factors and refining prediction models for better accuracy. The findings provide valuable references for academic research and practical applications in economics and financial engineering, supporting the healthy development and long-term stability of China's economy.

References

[1] Li Jiaqi, Ahn Hee-Joon. Sensitivity of Chinese stock markets to individual investor sentiment: An analysis of Sina Weibo mood related to COVID-19[J]. Journal of Behavioral and Experimental Finance, 2024, 41: 100860.

[2] Zhu Zhen, Jiang Wenlu. Empirical Analysis of the Mutual Influence between Yu'E Bao Sentiment Index and China's Stock Market[J]. China Price, 2016, (05): 36-39.

[3] Ye Luping. Study on Investor Sentiment and Stock Market Volatility Based on Stock Forum Comments: A Case Study of Jiuzhou Medical[J]. Advances in Applied Mathematics, 2022, (11): 8222-8230.

[4] Wang Haina, Wang Zelong. Study on the Spillover Effects of the Unexpected Unconventional Monetary Policy of the Federal Reserve on China's Financial Market Sentiment[J]. North China Finance, 2023, (06): 1-14.

[5] Li Helong, Ren Changsong, Liu Xinru, Wang Cunhua. A Review of Financial Market Text Sentiment Research[J]. Data Analysis and Knowledge Discovery, 2023, (12): 22-39.

[6] Pan Junyu. Construction and Verification of Financial Market Sentiment Index Based on Text Mining (Master's thesis, Hangzhou Dianzi University), 2021.

[7] Cheng Wanying, Liang Jianfang, Peng Xintong. Analysis of Consumer Purchase Intentions for Green Apparel in Social Media[J]. Journal of Basic Science of Textile Universities, 2020, (01): 45-50. doi:10.13338/j.issn.1006-8341.2020.01.008.

[8] Wan Yan, He Huiling. Analysis of the Characteristics and Management Strategies of University Network Public Opinion in the New Media Era[J]. Journalism Research Guide, 2024, (04): 82-85.

[9] Cui Luming, Qu Lingyan, He Dan. Evaluation of Urban Hotspot Spatial Emotion Perception Based on Deep Learning: A Case Study of Shanghai [J]. Human Geography, 2021, 36(5): 121-130.

[10] Nakajima J. Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications [J]. Monetary and Economic Studies, 2011, 29: 107-142.