ISSN: 2029-4913 / eISSN: 2029-4921

TECHNOLOGICAL and ECONOMIC DEVELOPMENT of ECONOMY

2025
Volume 31
Issue 2
Pages 447–479
https://doi.org/10.3846/tede.2024.21961

ADAPTING TO UNCERTAINTY: A QUANTITATIVE INVESTMENT DECISION MODEL WITH INVESTOR SENTIMENT AND ATTENTION ANALYSIS

Jie GAO[1], Xiuran BAI[2], Huimin TAN[3]✉, Chunguo FAN[4]✉✉,
Yunshu MAO[5], Zeshui XU[6]

[1,2,3,4,5]School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China
[6]Business School, Sichuan University, Chengdu, China

Article History:
- received 4 August 2023
- accepted 21 June 2024
- first published online 6 December 2024

Abstract. In the face of global uncertainties, including pandemics, economic fluctuations, disruptions in supply chains, major disasters, wars, and impending economic crises, the financial landscape and the impact of investor sentiment on the return of stock index futures can be significantly altered. Understanding the relationship between investor sentiment, attention, and stock index futures returns in the face of these diverse challenges has become particularly critical. However, existing research does not adequately consider the effect of these unexpected events on the market and the shifts in investor attention. Using the COVID-19 pandemic as a case study, this research proposes a dynamic quantitative investment decision-making model that considers the influence of investors' attention and emotional characteristics, aiming to adapt to the financial market under these global changes and improve the accuracy of quantitative investment forecasting. Initially, the Bidirectional Encoder Representations from Transformers model is employed to analyze investor comment data, extract information on investor attention and emotional characteristics, and construct investor sentiment indicators. Subsequently, a stock index futures forecasting method based on Variational Mode Decomposition algorithm and Support Vector Regression (SVR) model is constructed, and the grey wolf optimization algorithm is introduced to optimize the parameters of the SVR model. Guided by investor sentiment indicators, different market states are further distinguished, and appropriate investment strategies are implemented to effectively enhance the returns of quantitative investment. When compared with models that neglect investor attention and emotional characteristics, the results show that considering investor sentiment indicators not only improves the predictive ability of the model, but also reduces cognitive bias and market risk.

Keywords: investment decision-making, emotional characteristics, investor attention, support vector regression, variational model decomposition.

JEL Classification: C61, D91, E37, G41.

✉Corresponding author. E-mail: ***tanhuimin@swufe.edu.cn***
✉✉Corresponding author. E-mail: ***chunguo_fan@163.com***

1. Introduction

In the current global context, societies and economies are increasingly subjected to a variety of uncertain events (Fan et al., 2024). These events, ranging from pandemics like the Ebola virus and Corona Virus Disease 2019 (COVID-19), economic fluctuations, disruptions in supply chains, major disasters, wars, to impending economic crises, have direct repercussions on economic and social development worldwide. These events are spread through the Internet, which can easily cause information infection (Xie et al., 2023). In the continuous dissemination of information, investors' attention to the futures market is attracted. At the same time, the stock market al.o generates greater volatility due to major public events (Zaremba et al., 2020), which causes psychological anxiety and panic among investors (Selmi et al., 2021), which in turn affects the attention allocation and behavioral decisions of individual investors, and ultimately leads to changes in investor sentiment (Xie et al., 2023). This paper, taking the COVID-19 pandemic as an example, explores the impact of such uncertainties on investor behavior. Despite the substantial population of investors in China, a high proportion of these are individual investors who, typically deficient in fundamental financial knowledge, are particularly vulnerable to fluctuations in market sentiment (Lan et al., 2018). Especially in the post-epidemic era, investor sentiment has taken on a heightened importance (Gao et al., 2022). Changes in investor attention (Mbanga et al., 2019) and the occurrence of uncertain events may cause shifts in how investor sentiment affects stock returns. Thus, within the context of the post-epidemic period, it is of great significance to re-examine the irrational behavior in the market, to investigate the influence of major emergencies on quantitative investment decision-making models, and to construct effective investment strategies.

In order to construct an effective quantitative trading strategy, we first need to explore the price movement law of stock index futures (Zhou et al., 2010), which requires us to start from the internal and external factors that affect its price movement. Currently, many researchers have conducted empirical studies exploring the various factors influencing the price of stock index futures, such as the opening price, trading volume, open interest, and other economic factors (Wen et al., 2019; Zhou et al., 2021b). However, behavioral finance posits that investor sentiment is an important factor affecting investors' decision-making behavior and market stability. That is, the price of stock index futures is impacted not only by economic factors, but also by investor attention and sentiment characteristics (Audrino et al., 2020). Existing research predominantly centers on empirical studies investigating the effect of investor sentiment on financial markets, validating that investor sentiment indeed affects stock returns (Schmeling, 2009). The most classic example is the BW index, which was constructed by Baker and Wurgler (2006) based on observable market variables and aggregates information from six proxies. However, investor sentiment indicators constructed using observable market variables as proxy variables are limited by the frequency of data availability (Li et al., 2020a). Since big data technology was not yet mature at that time, although online comments indeed reflect and can influence the decisions and behaviors of investors, researchers did not incorporate the comments from stock forums into the construction of comprehensive investor sentiment indicators.

With the advancement of text mining techniques, some researchers have started to extract opinions from online text data. Therefore, accurately and effectively mining the emotional information from these texts has become a research hotspot in the field of behavioral finance (Sun et al., 2018). However, most user comments are characterized by colloquialisms and data redundancy, which poses new challenges to the quantitative analysis of investor sentiment (See-To & Yang, 2017). Traditional quantitative statistical analysis methods are mostly ineffective when dealing with such data, and machine learning theory has emerged as the primary scientific analysis method for this purpose. Consequently, many scholars have recently been conducting data mining of online comments with the help of text mining technology.

Among these, the pre-trained language model based on Bidirectional Encoder Representations from Transformers (BERT) effectively addresses the limitation of models that cannot accurately represent complex relationships between different sentences due to insufficient data. Simultaneously, the BERT model can perform bidirectional modeling of text, capturing the relationship characteristics between different sentences, thereby improving the accuracy of text sentiment mining (Devlin et al., 2018). Thus, this paper employs the BERT model to analyze investor review information and introduces investor attention to construct investor sentiment indicators.

Due to the nonlinear, non-stationary and time-varying characteristics of stock index futures data (Hou et al., 2019), machine learning has become the predominant method to predict stock index futures prices. There are two main ways to improve the prediction accuracy: one is the continual optimization and improvement of the prediction model; the other is continual expansion and upgrading of the input data processing methods. Machine learning methods include Artificial Neural Network (ANN), Support Vector Regression (SVR), and others. Although ANN can approximate any function, including nonlinear functions, it is challenging to determine the appropriate network structure and to address the non-convex problem of network training errors (Hong et al., 2011a). SVR has successfully solved prediction problems in many fields, including the stock market (Lu et al., 2009; Yaslan & Bican, 2017), due to its advantages of global optimization, simple structure, strong generalization performance, and being suitable for multi-dimensional and small sample data (Lin et al., 2022). However, the selection of SVR hyperparameters greatly influences the accuracy of the model (Cheng et al., 2022). With the development of intelligent optimization algorithms, more and more researchers integrate these algorithms with machine learning models to optimize the latter (Li et al., 2021). Moreover, to improve model's accuracy, it is necessary to consider not only the model's applicability and the choice of parameters but also the processing of input data. Although the research on Variational Mode Decomposition (VMD) technology in processing financial time series data is still in its infancy, it has the advantage of determining the number of modal decompositions and possesses stronger robustness to sampling and noise (Dragomiretskiy & Zosso, 2014). It is well-suited to solving problems related to the non-stationary and chaotic characteristics of financial time series. Therefore, this paper uses the VMD method to decompose the closing price series, obtains the characteristics of the closing price series under different frequency scales, converts the non-stationary data into stationary one, and substitutes it into the SVR model for prediction. The parameters of the SVR model are further optimized by the Grey Wolf Optimization (GWO) algorithm, which improves the accuracy of the model.

The main contributions and challenges addressed by this study are as follows: To address the significant challenges in achieving accurate and effective sentiment classification, especially given the colloquial nature of user comments and the complexity of sentence relationships, our contribution integrates the BERT model for analyzing investor reviews. This approach not only extracts the nuances of investor sentiment but also innovatively incorporates investor attention as a weighting factor. Unlike existing research that fails to effectively combine investor sentiment with attention levels, our method constructs comprehensive investor sentiment indicators. This combination critically enhances the prediction accuracy of market impacts, offering a novel solution to the prevailing challenges in sentiment analysis.

Then, to tackle the inherent challenges of non-linearity, non-stationarity, and time-variability in financial time series, as well as to overcome the difficulties associated with determining model parameters and avoiding suboptimal outcomes due to non-ideal parameter selection, our study introduces the VMD-GWO-SVR model. This innovative model synergistically combines various algorithms to address the non-stationarity and high volatility characteristic of financial time series. By employing the VMD method, the model effectively decomposes and denoises price series. This is followed by the application of the GWO enhanced SVR machine learning model for the prediction of the decomposed components. The final prediction results are then achieved through aggregation reconstruction. This comprehensive approach not only mitigates the challenges posed by financial time series analysis but also significantly improves prediction accuracy by ensuring the optimization of model parameters.

Lastly, to address the challenge of navigating market movements under the influence of various uncertain events, where the difficulty in discerning between bull and bear markets increases and a singular quantitative trading strategy falls short in a diversified market, our research contributes a novel approach. Utilizing the investor sentiment indicators developed in this study, we categorize the market into short and long markets. A quantitative investment strategy is then formulated based on the closing price sequence predicted by the VMD-GWO-SVR model. This strategy is tailored to accommodate different market conditions, dynamically adjusting to suit the continuously evolving financial market amidst changing epidemic circumstances. By segmenting the market according to different states and adjusting trading strategies accordingly, our approach offers a sophisticated means to adapt to the fluid nature of the market, providing a robust solution to the complexities introduced by uncertainty.

The structure of the paper is as follows: Section 2 presents the literature review. Section 3 establishes the investor sentiment indicators. Section 4 constructs the VMD-GWO-SVR algorithm model, decomposes the stock price using VMD, and ultimately builds the quantitative investment strategy. Section 5 describes the implementation and evaluation of the designed trading strategy. Finally, the conclusions and future outlook of this paper are given.

2. Literature review

2.1. Measurement of investor sentiment

With the development of behavioral finance, many studies have shown that investor sentiment affects investors' decision-making behavior. Conceptually, speculators' bias, either overly optimistic or pessimistic, is what we call investor sentiment (Brown & Cliff, 2004). Similarly, investor sentiment is also correlated with stock price fluctuations. Research by Pagolu et al. (2016) demonstrates a strong correlation between the rise and fall of stock prices and public sentiment expressed in tweets. Positive news and tweets about a company on social media certainly encourage people to invest in the company's stock, leading to an increase in its stock price. This reflects the investment intentions or expectations of market participants. Although online sentiment has been shown to be strongly correlated with the stock market, measuring sentiment remains challenging. The main steps of sentiment analysis include text extraction, refinement, classification, and score aggregation (Gupte et al., 2014). At present, the research on investor sentiment is mainly divided into two categories: one is the research on investor sentiment measurement; the other is the applied research of investor sentiment indicators.

Among them, how to accurately measure investor sentiment is an important issue in stock price forecasting and quantitative investment (Baker & Wurgler, 2007). The existing investor sentiment measurement methods are mainly divided into: the establishment of a single sentiment indicator, the establishment of a composite sentiment indicator, and text mining methods. When choosing a single sentiment indicators, volume (Campbell et al., 1993; Taş & Akdağ, 2012), turnover (Lee & Swaminathan, 2000; Statman et al., 2006) and IPO volume (Benveniste et al., 2003; Lowry & Schwert, 2002) have been used as proxy indicators. In order to reduce the influence of non-emotional factors in a single sentiment indicator, Brown and Cliff (2004) proposed that common factors hidden in multiple sentiment-related variables can be extracted to form a more accurate investor sentiment indicator. Therefore, later scholars used multiple single indicators to construct composite indicators to measure the changes in investor sentiment, such as the composite BW index, the GSI (Global Investor Sentiment) indicator (Baker et al., 2009), and so on. In the current era of rapid development of the Internet, investor sentiment is often reflected on social media such as Twitter, Weibo and forums.

Although online comments can indeed reflect and influence the decisions and behaviors of some investors, due to the immaturity of big data technology at that time, researchers did not incorporate stock forum comments into the construction of comprehensive sentiment indicators for investors. Today, big data technologies are becoming more mature and can better help researchers analyze and extract unstructured data such as comments from forum. Traditional quantitative statistical analysis methods are basically powerless to deal with such data, and machine learning theory is the mainstream scientific analysis method for dealing with this. Therefore, in recent years, many scholars have carried out data mining of online comments with the help of text mining technology. For example, Pak and Paroubek (2010) constructed a language-based sentiment classifier to determine the polarity of documents. Similarly, Zou et al. (2015) developed their approach based on bag-of-words method and syntactic analysis, employing machine learning methods to train a sentiment classifier that effectively enhances classification accuracy. However, these methods have certain limitations. One significant concern is their insufficient effectiveness in classifying Chinese texts in the financial field. Unlike other traditional machine learning algorithms, the BERT model has demonstrated exceptional performance in sentence-level sentiment classification (SST-2), as well as in ten other Natural Language Processing (NLP) tasks (Devlin et al., 2018). Furthermore, the BERT model can achieve better classification of Chinese texts in the financial sector through model fine-tuning.

Existing research primarily focuses on empirical studies investigating the impact of investor sentiment on financial markets, proving evidence that investor sentiment will affect stock returns (Schmeling, 2009). Shu (2010) demonstrated how changes in investor sentiment affect equilibrium asset prices and expected returns through a modified Lucas model, bridging the gap between theoretical and empirical evidence, and showing that sentiment does significantly influence decision-making. In addition, Su et al. (2022) suggested that exploiting market fear may optimize investor decision-making. However, few studies consider investor sentiment and build investment strategies from the perspective of decision-making, especially when it comes to risk and uncertainty.

2.2. Trends and applications of machine learning in quantitative investing

In recent years, the use of machine learning algorithms to predict stock prices has become the focus of research. There are two main ways to improve the prediction accuracy: one is the continuous upgrading and improvement of the prediction model, and the second is the continuous expansion and upgrading of methods for processing the input data.

A large number of machine learning methods have been applied to stock price prediction, such as Multiple Linear Regression (MLR), ANN and SVR. Due to inherent simplicity and the invalidity of linear assumptions, MLR models fail to produce highly accurate results in most cases (Mehtab et al., 2021). Although the ANN model can approximate any function, including nonlinear ones, determining a suitable network structure is challenging. Moreover, any ANN algorithm that minimizes the network training error is non-convex and it is difficult to find the global optimum (Hong et al., 2011b). In recent years, SVR has been widely used to solve nonlinear regression and time series problems, and has successfully solved forecasting problems in many domains, including the stock market (Lu et al., 2009; Yaslan & Bican, 2017). But the selection of SVR hyperparameters has a great influence on the accuracy of the model (Cheng et al., 2022). With the development of intelligent optimization algorithms, more and more researchers combine intelligent optimization algorithms with machine learning models (Li et al., 2021), such as Genetic Algorithm (GA), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), etc. Compared to these well-known meta-heuristics, the GWO algorithm can provide very competitive results (Mirjalili et al., 2014), requires fewer control parameters, and is easy to implement.

To improve the accuracy of the model, it is necessary to consider not only the quality of the model, but also the processing of the input data. Empirical Mode Decomposition (EMD) is a method for handling non-stationary signals. Compared with wavelet decomposition, this method doesn't require pre-setting any basis functions and has adaptive characteristics. The set of subsequences obtained by decomposition is called Intrinsic Mode Function (IMF). Dragomiretskiy and Zosso (2014) proposed VMD, which is different from EMD's method of recursively constructing IMF subsequences through loops. VMD shows non-recursion in the signal decomposition process and is more robust to sampling and noise (Li et al., 2019). Research on VMD in processing financial time series data is still in its infancy.

Furthermore, we find that models combining VMD and machine learning algorithms are less studied in dealing with financial time series problems. Although scholars have confirmed that the combination of VMD and SVR model can improve the prediction accuracy, most of the literature is an empirical study on this combination model for wind speed forecast (Wang et al., 2018), streamflow forecast (Zuo et al., 2020), electricity demand forecast (Niu et al., 2021) and electric load forecast (Zhou et al., 2021a). The VMD-GWO-SVR combined model can retain the effective information from the data while eliminating noise, exhibiting good denoising performance and robustness against sampling rates. Given the high volatility and nonlinearity of stock index futures data, there is a significant need for this combined model, which captures nonlinear data changes and improves prediction accuracy. However, there are few studies on the application of this combined model in stock price prediction, as shown in Table 1.

Table 1. The combination of modal decomposition technology and machine learning algorithm

Models	References	Research field
VMD-BPNN	(Hu et al., 2020)	Streamflow forecast
VMD-CNN	(Zang et al., 2018)	Photovoltaic power forecast
VMD-SVR	(Wang et al., 2018)	Wind speed forecast
	(Su et al., 2019)	Mold-level forecast for continuous casting
	(Zuo et al., 2020)	Streamflow forecast
VMD-GWO-SVR	(Niu et al., 2021)	Electricity demand forecast
	(Zhou et al., 2021a)	Electric load forecast

2.3. Research on stock index futures trading strategies

Adopting scientific and rational investment strategies holds significant importance for investors (Wang, 2020; Zhang et al., 2023). Among these strategies, quantitative investment, which involves analyzing historical data to inform decisions, has emerged as a mainstream choice. The current quantitative strategies mainly include market neutral strategies, trend following strategies, arbitrage strategies and high frequency strategies, as shown in Table 2.

Market-neutral strategies aim to mitigate risk exposure by maintaining multiple positions, including both long and short positions, thus rendering the overall portfolio insensitive to general market movements. Empirical research indicates that market-neutral strategies, particularly dynamic hedging strategies, significantly outperform both static and traditional strategies (Olgun & Yetkiner, 2011). The essence of trend-following strategies lies in trading according to market trends. For instance, investors adopt long positions when a clear upward trend in prices is observed, anticipating further gains as prices continue to rise. Arbitrage strategies seek to profit from price discrepancies across different markets, assets, or within the same market's different instruments. Caldeira and Moura (2013) evaluated the profitability of statistical arbitrage strategies using data from the São Paulo Stock Exchange from January 2005 to October 2012. Their analysis showed that statistical arbitrage strategies yielded an annual excess return of 16.38% with a Sharpe ratio of 1.34, indicating low market correlation. High-frequency trading strategies exploit computer algorithms and rapid trading techniques to execute a large volume of transactions in a very short timeframe. However, these strategies are prone to execution risks, preventing sustained trading at optimal buy/sell prices (Guilbaud & Pham, 2013).

Among them, trend following strategies are widely accepted by investors for their simplicity and efficiency (Hu et al., 2015), and are systematically described in the work of Covel (2004). Three common trend following strategies are moving average strategy, momentum strategy and channel breakout strategy. The simplest and most effective trend following strategy is to use a moving average (James, 2003). The moving average is a curve that draws the average value of a certain price or price index on the coordinate diagram for a certain period of time. It can predict and judge the future trend of the price or price index. Moving averages of different time parameters can be obtained for different time periods: long-term moving average and short-term moving average. The crossover based on these two moving averages is called Dual Moving Average Crossover (DMAC) (Pätäri & Vilska, 2014). Various studies have demonstrated the strong profitability of DMAC rules (Luukka et al., 2016; Szakmary et al., 2010).

To sum up, the use of moving average and its improved strategy in trend tracking can better identify price trends and thus divide market state. Therefore, by combining the impact of investor attention and emotional characteristics on stock returns with the moving average strategy, we can judge stock market trends through investor sentiment indicators.

Table 2. Futures Market Investment Strategies

Type	Strategy	Literature
Trend Following Strategies	Lead-Lag Relationship Strategy	Brooks et al. (2001)
	Intraday Momentum Strategy	Li et al. (2020b)
	Dual Moving Average Crossover and Channel Strategy	Szakmary et al. (2010)
Market Neutral Strategies	Dynamic Hedging Strategy Based on Binary GARCH Estimation	Olgun and Yetkiner (2011)
Arbitrage Strategies	Statistical Arbitrage Strategy Based on Cointegration	Caldeira and Moura (2013)
	Statistical Arbitrage Strategy Using Pair Trading	Montoya-Cruz et al. (2020)
High-Frequency Strategies	Optimal Limit Order Book (LOB) Market Making Strategy	Guilbaud and Pham (2013)

3. Establish investor sentiment indicators

This section mainly introduces the following processes: data collection, text sentiment analysis using the BERT model, and finally, building sentiment indicators.

3.1. Data collection and data preprocessing

This article employs the Selenium and XML package in Python3.8.8 to crawl investor review data from the CSI 300 section of the East Money Network's Guba stock community, accessible at http://guba.eastmoney.com/. The CSI 300 Index, jointly released by the Shanghai and Shenzhen Stock Exchanges on April 8, 2005, reflects the overall trend of the A-share market. The sample of the CSI 300 Index cover about 60% of the market capitalization of the Shanghai and Shenzhen stock exchanges, which offers good market representation and can reflect the general situation of stock price changes in China's securities market. We obtained the CSI 300 Index review text data from April 2019 to December 2021 using web crawlers and screened out more than 14147 comments as the textual source of market sentiment.

The crawled information includes various details such as page information, comment title, author, publication time, number of views, content links, and comment content. Figure 1 reflects the number of investor comments from December 31, 2020 to December 31, 2021, along with changes in the returns of the CSI 300 Index Futures during the same period. It is clear that the yield of CSI 300 Index Futures correlates with the number of comments. For instance, the number of comments at the end of January 2021 was significantly lower

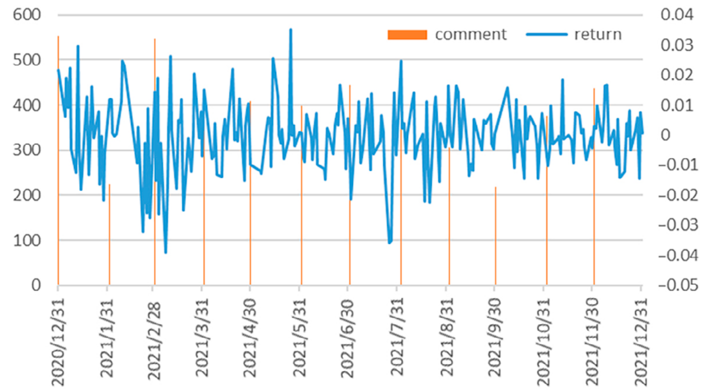

Figure 1. Monthly comments and daily rate of return

compared to the beginning of January, and yields also trended downward. In March, the number of comments related to the CSI 300 Index Futures rose again, which was mirrored by the upward trend of yields. This shift may be related to the rebound of the COVID-19. At the end of January 2021, the epidemic situation rebounded in Harbin, Heilongjiang, Hebei and other places in China. Additionally, due to the large population movement during the Spring Festival, the epidemic resurged to varying degrees in many places. Panic spread in the stock market, and investors' pessimistic expectations were quickly reflected in stock prices, resulting in a sharp drop in index yields at this stage. With the epidemic's improvement in February, investor sentiment gradually turned more optimistic, leading to a steady increase in index yields during this period. Nonetheless, it is apparent that at the end of February, the uncertainty of the COVID-19 epidemic caused the yield of the CSI 300 Index Futures to fluctuate wildly. In conclusion, there appears to be a correlation between investors' attention and the CSI 300 Index. Simultaneously, the occurrence of uncertain events, similar to COVID-19, can also affect investors' attention and emotional characteristics, which in turn influences the trend of the stock market.

Most of the comments on the Oriental Fortune Internet Stock Bar are real investors' discussions and exchanges about the stock market, but there are also some useless posts, such as some media news reposts and advertising pushes. These posts are repetitive and unrelated to investor sentiment, which may undermine the authenticity of the final analysis. Therefore, before constructing the investor sentiment indicators, the extracted text data is first cleaned, and the reposted news and advertisement pushes are removed from the investor sentiment texts. The specific methods are as follows:

(1) Advertising content with obvious inducement words: such as posts with phrases like "recommended stocks", "exposure" and "teach you" in the title or body text;

(2) The content of news reposted from other websites: such as posts with the words "reposted" and "forwarding news" in the title;

(3) Posts with the same content that appear multiple times consecutively will also be deleted and only counted once.

3.2. Establish investor sentiment indicators based on BERT

We have explained the tasks of data collection in the previous subsection. Data collection includes crawling data from the Internet, analyzing and preprocessing the data. In this subsection, we use the BERT model to analyze investor comments and obtain three categories of investor emotional characteristics. We then comprehensively consider the occurrence of uncertain events and changes in investors' attention to construct investor sentiment indicators.

3.2.1. The BERT model

BERT, a masked language model relying on a multi-head attention mechanism (Devlin et al., 2018), has transformed the way we understand and utilize text data. Its essential building block, the Encoder part of the Transformer, captures bidirectional relationships within sentences, enabling comprehensive, context-aware encoding. The unique characteristic of BERT is its complete reliance on the attention mechanism, a method for understanding and correlating different components of an input sentence. This mechanism enables BERT to capture interdependent information from long sentences, even associating word information at any two positions within a sentence. This capability results in enhanced text classification effects, especially valuable for our sentiment analysis task.

The training of the BERT model is divided into two steps: pre-training and fine-tuning. In this paper, BERT-large is used, and the Softmax model is selected for the sentiment classification task, as shown in Figure 2.

The first stage is Pre-training. During pre-training, the model is trained on unlabeled data through different pre-training tasks. Most of the work on this model has been done by Google, and the pre-trained BERT model weights already encode a substantial amount of information about our language. BERT-large, having more parameters than other models with high accuracy, so we chose it as our pre-training model.

The second stage is Fine-tuning. Here, we initialize the BERT-large model with pre-trained parameters, and fine-tune the labeled data from the downstream sentiment classification task

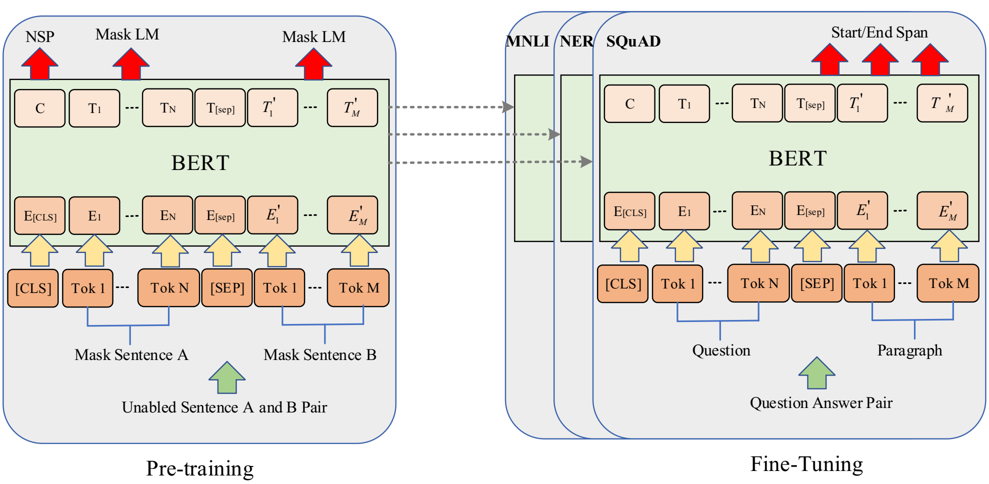

Figure 2. BERT model and Sentiment Classification

using the word embedding layer and Transformer layer in the model. When the BERT model handles the sentiment classification task, the first word of the sequence is identified with a unique token [CLS], and a fully connected layer is attached at the [CLS] position of the last encoder layer. Finally, the Softmax function completes the classification of sentences or sentence pairs. For instance, returning to the investor reviews we collected in Subsection 3.1, Softmax may generate possibilities for reviews belonging to a particular class, as shown in Table 3.

Table 3. Classifications

Classification	Probability
negative	yes/no
neutral	yes/no
positive	yes/no

3.2.2. Use the BERT model for sentiment analysis

According to the requirements of the BERT model, and after observing the distribution of likes and views on the data, we considered that the data may be unevenly distributed, 7500 reviews were randomly selected as machine learning data. Data labeling was done jointly by three experts. The data is labeled for three of the two sentiment polarities: positive, neutral and negative. If there is a disagreement in the data labeling process, it is necessary to discuss and modify each other to obtain consistent values. After further inspection and validation, the expert-labeled data is divided into training set and test set. The former accounts for 80% of the data volume, and the latter accounts for 20%. The details of the labeled data are shown in the Table 4 below:

Table 4. Number of training set, test set, and validation set

Classification	The amount of data	Percentage	Percentage of the total data set: 7500 Training data set: 6000 Test data set: 1500
negative	2500	35.71%	
neutral	2000	28.57%	
positive	2500	35.71%	

We classify online reviews into positive, neutral and negative categories using the BERT model. Additionally, we noticed that every comment on Oriental Fortune includes a “click” column for the post, indicating the attention each Oriental Fortune user pays to the post, and this value can be defined as “investor following”. Therefore, we assert that the sentiment values of different stock themes cannot be directly weighted and summed, as each text message signifies a varying level of attention. Investors’ attention is defined here as shown in Formula (1).

$$att_t = \sum_{i=1}^{i=n} clicks_i, \tag{1}$$

Among them, n is the number of reviews on day t, *clicks* is the number of click per review. Attention represents the total number of clicks for the current trading day. There are many factors that affect investor sentiment, among which the impact of major uncertainties is the most significant. For example, since the outbreak of COVID-19, information related to the outbreak has been widely disseminated on social media, and many investors have reacted strongly to the outbreak. It has not only had a significant impact on people's daily routines and lives, but the global economy has also continued to decline, and unemployment has increased as the outbreak has occurred in different regions for different periods of time. The volatility of the economy has severely affected the psychological and attentional behavior of investors. According to the attention-driven purchase theory, increased investor attention to a particular stock signals broader interest. Given their limited cognitive resources, investors typically screen and invest in familiar stocks. Consequently, heightened attention not only elevates potential investment demand but also boosts trading volumes and stock prices (Barber & Odean, 2008).

In this process, while investors' psychology is seriously affected, investors will also actively adjust their attention behaviors to adapt to the new market environment. The outbreak of the epidemic has increased investors' uncertainty about the development of the market economy, and at the same time, investors' consumption desires and consumption scenarios have been restricted. The growing demand for investment and wealth management of individuals and the transformation of offline consumption ability to online investment ability have promoted investors' enthusiasm for investment transactions and also had an impact on investor sentiment (Zhao, 2020). Therefore, the design idea of the investor sentiment indicator in this paper can be represented in Figure 3, and the investor sentiment index is finally obtained by using investor attention to weight the sentiment classification results.

We utilize the BERT model to analyze text sentiment directly, obtaining the sentiment polarity of reviews. On day *t*, we use M_t^{buy} to represent the number of posts with positive sentiment and M_t^{sell} for the number of posts with negative sentiment. $sum\left(RN_t^+\right)$ represents the total readership of all positive sentiment posts M_t^{buy} on day t, and $sum\left(RN_t^-\right)$ represents the total readership of all negative sentiment posts M_t^{sell} on day t. Accordingly, we can divide investor attention into positive attention and negative attention, with positive attention on day *t* being $sum\left(RN_t^+\right)$ and negative attention being $sum\left(RN_t^-\right)$. In previous research, Antweiler and Frank (2001) demonstrated a method for constructing sentiment indicators from messages posted on Yahoo's stock message boards, as shown in Formula (2), where E_t represents the investor sentiment value on day *t*.

$$E_t = \ln\left(\frac{1+M_t^{buy}}{1+M_t^{sell}}\right). \tag{2}$$

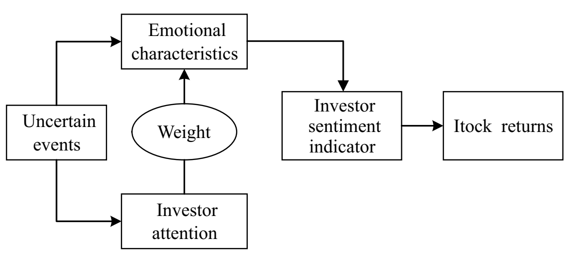

Figure 3. The impact of investor attention and investor sentiment on stock returns

This paper improves upon the aforementioned sentiment indicators by incorporating positive and negative attention, as shown in Formula 3:

$$E_t = \ln\left(\frac{1+sum(RN_t^{+})}{1+sum(RN_t^{-})}\right). \quad (3)$$

In this improved indicator, we use readership to construct positive and negative attention as weights for the number of sentiment posts, meaning that posts with higher readership will have a greater influence when calculating the sentiment indicators. Such an improvement can more accurately reflect changes in investor sentiment, as it associates posts with higher attention levels with the sentiment indicators, while disregarding those with less impact. We correlate the calculated investor sentiment values with the closing price of the CSI 300 Index Futures on the same day, and the results can be seen in Figure 4.

Through Figure 4, we can clearly see the correlation between the investor sentiment indicators and the closing price series. For example, at the end of December 2020, investor sentiment gradually strengthened, and the closing price series also showed an increasing trend in early January 2021. Before the closing price series showed a significant downward trend at the end of February 2021, the investor sentiment indicators had already shown a declining trend. Therefore, the investor sentiment indicators constructed in this paper can effectively express investor sentiment, which is beneficial for the construction of prediction models and quantitative strategies.

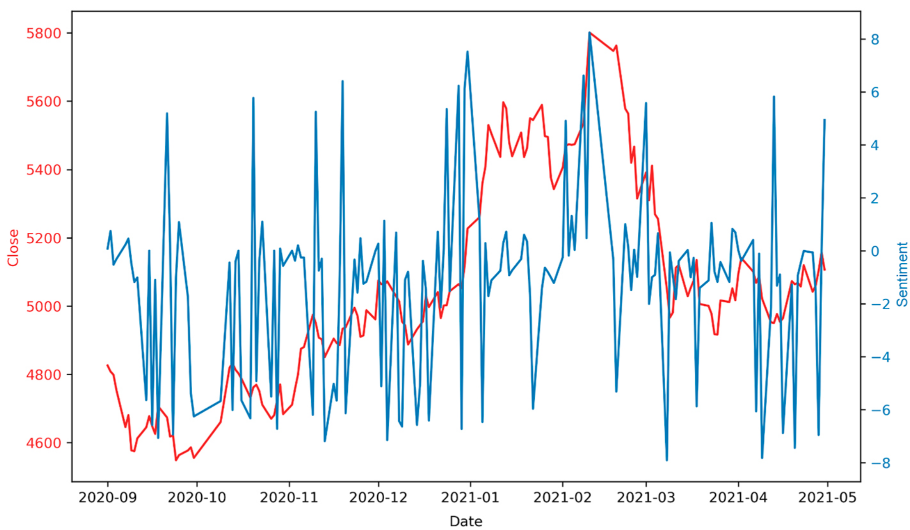

Figure 4. Closing price and sentiment indicators of CSI 300 Index Futures

4. Trading strategy design based on VMD-GWO-SVR and investor sentiment indicators

This section will build a stock index futures price prediction model combining VMD-GWO-SVR, and formulate investment strategies using investor sentiment indicators in order to reduce risks and achieve higher returns compared to the market.

4.1. Theoretical framework of trading strategy

4.1.1. GWO-SVR algorithm

GWO is a meta-heuristic algorithm introduced by Mirjalili et al. (2014), which simulates the predatory behavior of grey wolves in nature to achieve optimization based on cooperative mechanisms among wolves. SVR is an important application branch of the support vector machine (SVM) and is a commonly used supervised learning method. However, the fitting capability of SVR is affected by three factors: the penalty factor c, the parameter gamma (g) that determines the kernel function, and the epsilon parameter (e) defining the loss function. In this paper, the radial basis kernel function (RBF) is selected as the kernel function due to its relatively mature and ability to approximate any continuous function within a compact set. Therefore, we utilize GWO to optimize the parameters (c, g, e) of SVR model, and then predict the components of the intrinsic mode function (*IMF*) decomposed by VMD. The pseudo-code for the GWO-SVR algorithm is provided in Table below.

Table 5. GWO-SVR algorithm

Algorithm 1: GWO-SVR algorithm

Input: Number of grey wolf in the population N, maximum number of iterations allowed T, range of the penalty parameter c, range of the RBF kernel function parameter g, range of the loss function parameter e, and VMD decomposition dataset D
Output: Optimal parameters *Best_c*, *Best_g* and *Best_e*, the component IMF_i predicted values and error

```
Begin
    1. Initialize the grey wolf population positions X_i(i = 1, 2, ..., N) based on c, g, and e values.
    2. Call the SVR, calculate and update the fitness of each search agent.
       while each iteration 0 < t < T :
    a. while each wolf 0 < i < N:
                  For wolf i, the fitness is calculated using the current parameters c, g, and e
   b. The first three wolves are selected by calculation:
                 X_α, X_β, X_δ.
                  Each of the three agents represents the top three in the performance ranking.
   c. while each wolf 0 < i < N:
                  Updates the position of the wolf i affected by X_α, X_β, X_δ.
   d. Re-evaluate all current wolves and update X_α, X_β, X_δ if there is a better position.
            e. Enter the next iteration of searching:
                 t = t + 1
       end while
       3. The output gets: Best_c, Best_g and Best_e.
End
```

In this paper, we set N = 15, N = 50, then set the parameter range of g and e to [0, 5], and the setting range of c to [0, 15].

4.1.2. VMD-GWO-SVR algorithm model

The actual CSI 300 Index Futures data is volatile and non-linear complex data. In view of the advantages of VMD, GWO and SVR (Drucker et al., 1996; Mensi et al., 2017; Mirjalili et al., 2014): VMD can solve the problems of non-stationary and chaotic nature of financial time series. SVR has the advantages of global optimum, simple structure, strong generalization performance, and is suitable for multi-dimensional and small sample data. At the same time, GWO algorithm with strong convergence, few parameters and easy implementation can assist SVR model to determine the optimal parameters and improve the prediction performance of the model. Therefore, we combine them to build a hybrid model VMD-GWO-SVR. The following Figure 5 describes the implementation process:

The specific operation process of VMD-GWO-SVR is as follows:

Step 1: VMD technology is used to decompose the original closing price sequence into K mutually independent sub sequences, denoted by IMF_1, IMF_2, ..., IMF_K, which represent different local price fluctuations from high frequency to low frequency. The initial sequence is reconstructed in the form of *IMF*: $x(t)=\sum_{k=1}^{K} IMF_k(t)$. Then, each *IMF* series becomes a new prediction sample of the GWO-SVR model.

Step 2: Each component *IMF* obtained by VMD decomposition is divided into training data set and test data set, and the two sets are completely separated to ensure generalization ability. The segmentation interval of specific training set and test set will be shown in Section 5.1. The GWO-SVR is used to train and establish the prediction model based on the training data set.

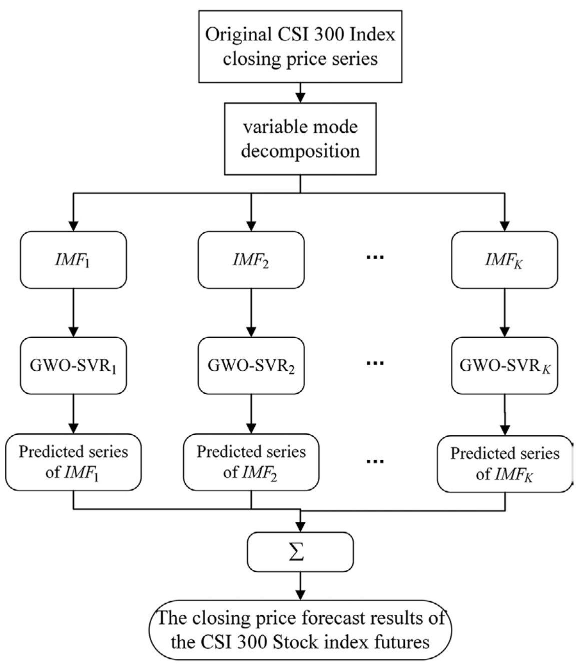

Figure 5. Flow chart of the VMD-GWO-SVR model

Before optimizing SVR parameters: penalty parameter c, RBF kernel function parameter g, loss function parameter e, the parameters that are important to ensure the prediction accuracy of the model, such as the initial population size M, the maximum number of iterations N, need to be set in advance. Then, according to the constructed GWO-SVR model, the prediction results of the *IMF* sub series on the stock price are obtained.

Step 3: Each *IMF* component is independently predicted to output $f_k(t)$, $k = 1, 2, \ldots K$, and the final prediction result of the original stock price time series is obtained by summarizing, that is $\sum_{k=1}^{K} f_k(t)$. The *IMF* prediction errors of different dimensions have different effects on the final prediction results; Among them, the low-frequency *IMF* usually consists of large values representing the macro trend of the sequence. From the perspective of prediction accuracy, the prediction error of low-frequency *IMF* has a more significant impact on the final prediction result than that of high-frequency *IMF*.

Step 4: Introduce mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), goodness of fit (R^2) evaluation indicators to evaluate the prediction ability of VMD-GWO-SVR from different perspectives.

4.2. Stationarity analysis and signal decomposition process

In order to solve the problem that financial time series are unstable and noisy, this paper uses VMD method to solve the prediction lag caused by the non-stationary data. This section will test the stationarity of the closing price sequence of stock index futures. When the verification sequence is non-stationary, VMD decomposition will be carried out on the sequence.

4.2.1. Stability test of stock index futures price

This paper selects CSI 300 Index Futures (IF) as the research object for the following two reasons: First, the CSI 300 Index was jointly developed by the Shanghai and Shenzhen exchanges under the coordination of the China Securities Regulatory Commission. It is one of the most authoritative cross-market indexes in China and offers good representation. Second, CSI 300 Index Futures is the earliest listed stock index futures in China, which can provide more trading data for the prediction model. At the same time, the stock index futures market adopts $T + 0$ trading mode, which can construct a more flexible trading strategies and enrich the investment strategies.

Because stock index futures contracts have a certain term, this paper needs to splice discontinuous and term futures contracts before time series analysis. In this paper, the main contracts of stock index futures are spliced to form the main continuous contracts of CSI 300 Index Futures, referred to as IF9999 for short. In consideration of investor emotional characteristics and investor attention, the corresponding sample data is selected from 658 trading days from April 4, 2019 to December 31, 2021. The opening and closing prices are as shown in following Figure 6.

In order to observe the data characteristics of the research object, this paper makes a statistical test on the closing price series by using EViews software, and the results are shown in Figure 7.

As shown in the above figure, the skewness of the closing price sequence of the main continuous contracts of CSI 300 Index Futures is less than 0, which is slightly to the right; Kurtosis is less than 3, and the data is flatter than the normal distribution.

Then, in order to verify the stability of IF9999 opening price and closing price series, EViews is used for ADF test. The results are shown in Figure 8.

Figure 6. Closing and opening price series of IF9999

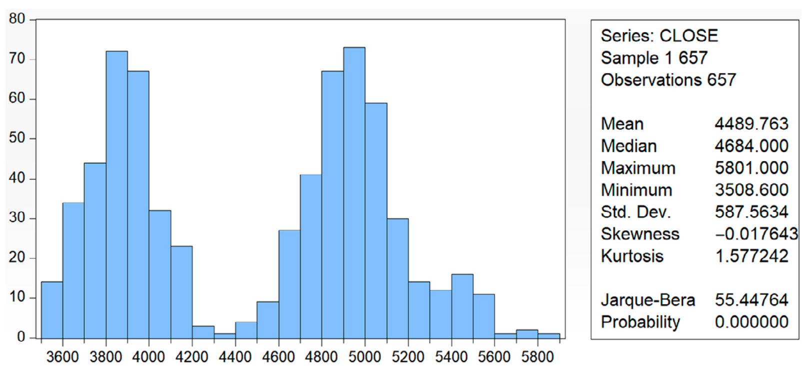

Figure 7. Description and statistical analysis chart of closing price of IF9999

Null Hypothesis: CLOSE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic – based on SIC, maxlag = 19)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		−2.495610	0.3303
Test critical values:	1% level	−3.972104	
	5% level	−3.416683	
	10% level	−3.130681	

*MacKinnon (1996) one-sided p-values.

Figure 8. ADF test of IF9999 closing price series

As shown by the ADF test results of the closing price series, the ADF unit root test of the closing price series is greater than the critical value at the 5% significance level. It shows that the closing price series of IF main continuous contracts are non-stationary time series at the 95% confidence level.

4.2.2. VMD decomposition of stock index futures price time series

Before VMD decomposition, the number of modes K, penalty factor α, fidelity coefficient τ and convergence stop condition ε need to be determined in advance. Among them, the penalty factor α is taken according to experience, which is usually 1.5 to 2 times the sample size. According to the sample size of this paper, the penalty factor α is 727. The fidelity factor τ is set to 0, and the convergence stop condition takes the default value $\varepsilon = 10^{-7}$. Last, the choice of the number of modes K determines whether there is mode aliasing or under decomposition in VMD decomposition, which plays an important role in the decomposition effect. Therefore, this paper studies the selection of K value of VMD decomposition method, and analyzes the residual and correlation of the number of modes k after VMD decomposition.

For K value, K = 3, 4, ..., 13, decomposing the opening price and closing price series of CSI 300 Index Futures. Reconstruct the decomposed data, and then solve the RMSE value of each group of reconstructed sequence and the original input sequence to measure the deviation between the reconstructed sequence and the original sequence. The calculation formula of RMSE is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \tag{4}$$

After decomposing the original closing price series of CSI 300 Index Futures, the RMSE calculation results under different K values are shown in the following Figure 9.

From the calculation results of RMSE, the residual decreases gradually with the increase of K value, but the slowing down trend is not obvious. In other words, it is not easy to distinguish whether there is over decomposition in VMD decomposition only from the size of the residual. Therefore, from the perspective of the influence of K value on the data stationarity of *IMF* component series, this paper will test the stationarity of *IMF* component series obtained under different K values.

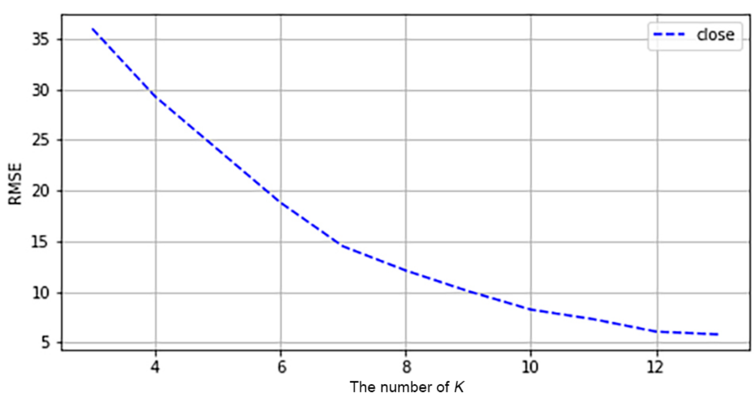

Figure 9. Root mean square error under different K values

Table 6. ADF test value under different K value

	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8	K = 9	K = 10
IMF_1	–1.30	–1.41	–1.74	–1.89	–1.74	–2.11	–2.23	–2.02
IMF_2	–7.18	–4.72	–3.88	–3.00	–3.86	–3.77	–3.32	–3.61
IMF_3	–12.83	–12.40	–10.30	–9.36	–9.52	–8.01	–10.43	–10.16
IMF_4		–13.26	–16.18	–11.50	–11.44	–11.22	–11.59	–12.19
IMF_5			–19.60	–11.94	–10.98	–11.63	–11.57	–9.03
IMF_6				–6.48	–7.16	–11.94	–12.47	–7.96
IMF_7					–6.25	–6.88	–6.45	–7.28
IMF_8						–6.32	–5.51	–7.74
IMF_9							–5.31	–6.68
IMF_{10}								–6.52

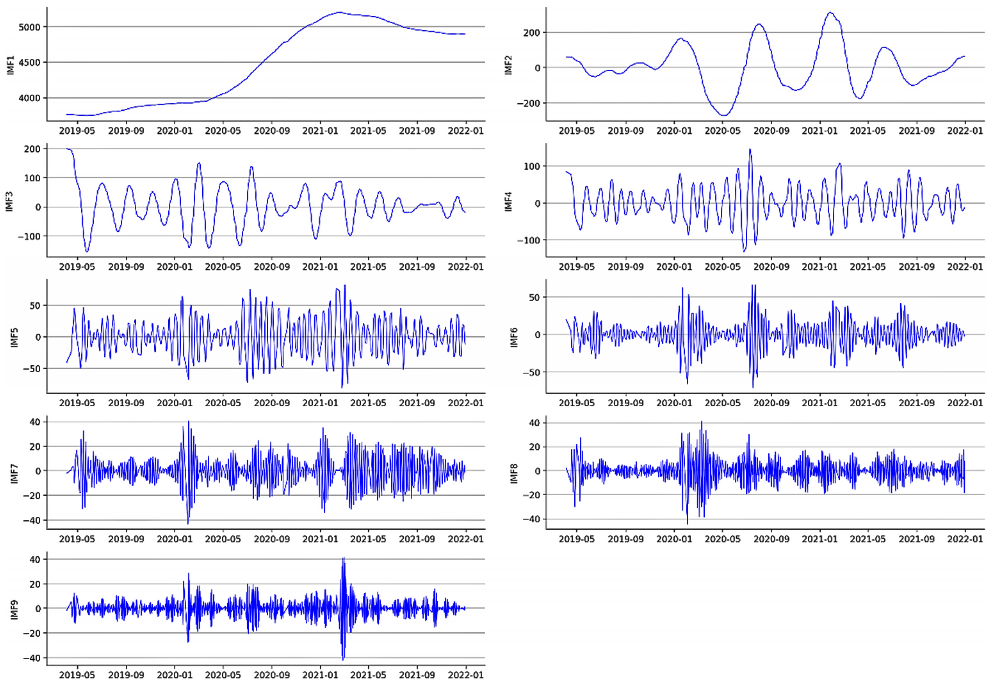

Figure 10. VMD decomposition results of closing price series

In this paper, the ADF unit root test is still used to test the stationarity of the decomposed sequence *IMF*. The smaller the estimated value of the ADF test statistics, the more the original hypothesis is rejected, and the more likely the sequence is to be stable. This paper will select the K value that is most likely to make the sequence stable. For closing price series, the results of ADF test under different K values are shown in the Table 6 above.

It can be found from the above table that when K = 10, the ADF test value of *IMF* component series increases significantly, so it can be considered that when K increases to 10, VMD has over decomposition. To sum up, the number of modes of VMD decomposition in this paper is K = 9.

According to the VMD parameters selected in this paper, i.e., K = 9, penalty factor = 727, fidelity coefficient = 0 and convergence stop condition = 10, the VMD decomposition is carried out for the closing price distribution of CSI 300 Index Futures. As shown in the Figure 10 above, 9 groups of component sequences are obtained by VMD decomposition.

4.3. Quantitative investment strategy design of stock index futures based on VMD-GWO-SVR model

Behavioral finance believes that the price of securities market is not determined by the intrinsic value of securities (Hirshleifer, 2015). The price of securities often deviates from its intrinsic value. To a large extent, the price of stock index futures index is also affected by the behavior of investors. That is to say, investor sentiment will have an impact on the securities market and will affect the return rate of stocks. We assume that the price of stock index futures index is affected by market sentiment in the process of change (Dash & Maitra, 2018; Seok et al., 2019). When the market sentiment is high, people's enthusiasm for buying and selling stocks will increase, and the market liquidity will become higher, which will promote the rise of stock prices; On the contrary, when the market mood is low, people are not willing to participate in the market too much or even not willing to participate in the market, and the market liquidity decreases, which will make the stock price subject to downward pressure and depress the stock price.

Because investor sentiment will affect the returns of stock index futures, and then affect the return rate of investment, it is necessary to formulate investment strategies through investor sentiment to reduce risks and obtain more than market returns. At the same time, considering the risk of market investment, this paper combines the prediction results of VMD-GWO-SVR model on the closing price of stock index futures to build an investment strategy. First, we can judge the trend of the market in the future by judging the investor sentiment indicators. According to the market trend, we can divide it into long market and short market. Then, under different market conditions, the trading signal is determined according to the prediction result of the closing price.

Commonly used indicators to judge the market state include moving average, comprehensive moving average, brin line, etc. In this paper, the traditional comprehensive moving average, that is, the intersection of multiple moving averages, is selected as the indicator to judge the market state. Based on the investor sentiment indicators series, this paper selects two moving averages: the 10-day moving average and the 30-day moving average, which are respectively recorded as SMA10 and SMA30. Generally speaking, the trend reflected by the short-term moving average is a short-term trend, with high sensitivity and small hysteresis. However, it will also be affected by the frequent fluctuations of the stock index futures and produce a certain error signal; The long-term moving average can well reflect the operating track of the price. Although there is a certain lag, it is very suitable for band operation. This paper chooses the idea of dual moving average crossing strategy, which makes use of the relative relationship of double moving average to better identify the price trend and thus divide the market state.

If SMA10 is above SMA30, it is judged that the price of stock index futures has an upward trend, and the current market state is a long market; If SMA10 is below SMA30, it is judged that the price of stock index futures index has a downward trend and the current market state is short market. The determination methods of trading signals under different market conditions are as follows:

(1) Construction of trading signals in the long market. If the current market is in a long position, the underlying asset shows a short-term long position trend, so the signal of building a long position can be determined. The closing signal under the current market state will be obtained by the VMD-GWO-SVR model predicting the closing price data of the next trading day. If the forecast value is less than today's closing price, it means that the market will conduct a short-term correction on the next day. At this time, a position closing signal will be constructed and the position will be closed at the opening of the next day.

(2) Construction of trading signals in short market. If it is currently in the short market state, the underlying asset shows a short-term short trend, so the signal of short position can be determined. Under the condition of holding short positions, if the forecast signal of VMD-GWO-SVR model shows that the closing price of the next trading day will rebound, the position clearing operation will be carried out at the opening of the next day, and then wait for the next buy signal. To sum up, the specific trading strategy is shown in the above Figure 11.

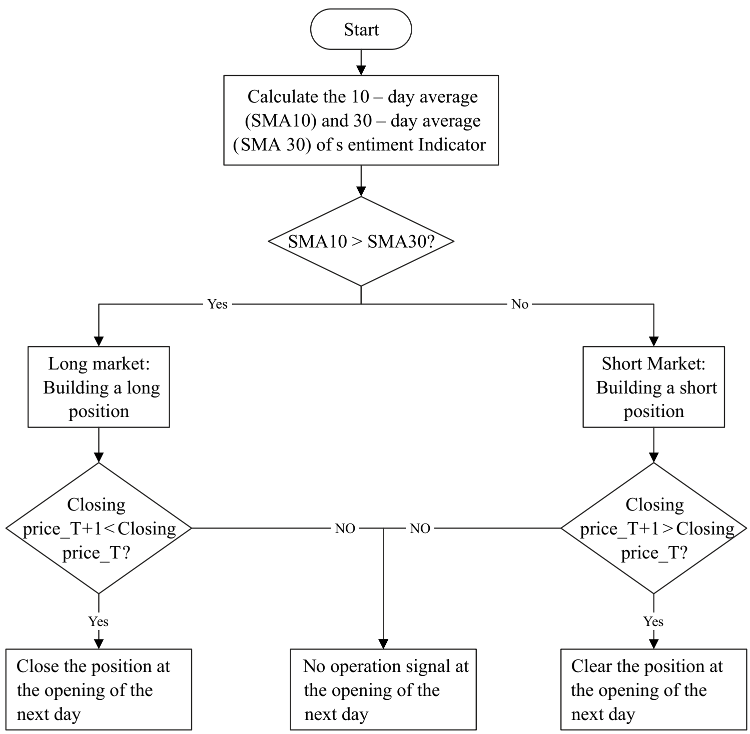

Figure 11. Trading strategy

5. Implementation and analysis of quantitative investment strategy of stock index futures

This section will use VMD-GWO-SVR model to forecast the price of CSI 300 Index Futures, and build a quantitative trading strategy based on the forecast results. This section not only implements the quantitative investment strategy proposed in this paper, but also compares it with other strategies for further effectiveness analysis.

5.1. Prediction based on VMD-GWO-SVR model

The research object of this paper is the CSI 300 Index Futures, and the analysis period is from April 4, 2019, to December 31, 2021, with a total of 657 sets of data. Use 400 sets of data from April 4, 2019, to December 11, 2020, as the training set to train the model and perform parameter optimization, and use 256 sets of data from December 14, 2020, to December 31, 2021, as the test set to measure the model's accuracy. In Section 4.2, the original closing price sequence is decomposed by VMD to obtain 9 sets of *IMF* components with different physical properties and a residual sequence.

This paper will use GWO-SVR model to forecast and analyze each group of *IMF* components. When forecasting the *i*th and component of IMF_i, the selected characteristic indicators are: sentiment indicators, IMF_i, today's opening, highest price, lowest price, trading volume, transaction amount, position, position change, today's closing, today's settlement, previous settlement, up and down 1, up and down 2, a total of 14 indicators. Taking IMF_1 as an example, the model is trained on the training set and the parameters are optimized. When building the GWO-SVR model, the radial basis kernel function is selected as the kernel function, the parameters c, e, and g are optimized using the GWO algorithm, and the number of iterations is set to 50.

When using GWO to optimize the SVR parameters for the IMF_1 component, the change of the goodness of fit with the number of iterations is shown in the following Figure 12.

The parameter optimization results are: c = 9.92562094, e = 6.10528099e − 5, g = 0.123332077. Then use the trained model to predict IMF_1 on the test set, and the prediction results are shown in the following Figure 13.

It can be seen from the above figure that the prediction effect of IMF_1 is very accurate, and the goodness of fit R^2 reaches 99.76%. Similarly, GWO-SVR is used to predict other *IMF* components. The following Figure 14 shows the prediction results of all *IMF* components.

The empirical results show that most *IMF* components have good prediction effects, and only the prediction effect of IMF_6, IMF_7 and residual series is poor. The poor prediction performance of high-frequency components and residual sequences is due to the frequent and random fluctuations of high-frequency components, making it difficult to predict. However, due to its relatively small value, its impact on the final predicted value is minimal. Now, sum all the predicted *IMF* components and the predicted residual series to calculate the predicted value of the closing price of the CSI 300 Index Futures. The following Figure 15 shows the closing price forecast results of the CSI 300 Index Futures.

The predictive performance of the GWO-SVR model was evaluated using four key metrics: R^2, MSE, MAE, and MAPE. The model exhibits a high R^2 of 97.80%, indicating it can explain

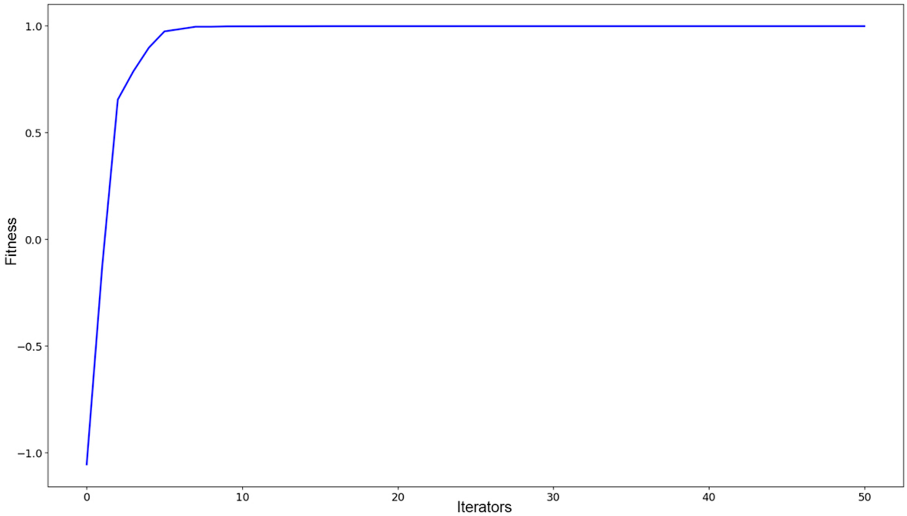

Figure 12. Variation of goodness of fit with number of iterations

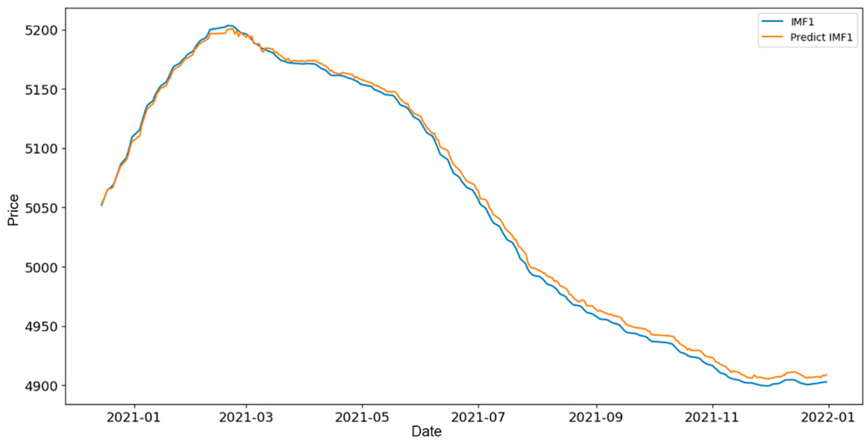

Figure 13. Forecast results of IMF_1

97.80% of the variance in the dataset, which signifies strong predictive accuracy. Although the MSE and MAE are relatively high at 1055.01and 24.16 respectively, these values are typical for models dealing with stock index futures, which often have high closing prices and significant volatility. In contrast, the MAPE is exceptionally low at 0.0047, almost nearing zero, which suggests that the model's predictions are very close to the actual values. Given the low MAPE and high R^2, the model demonstrates robust forecasting capabilities.

To further illustrate the effectiveness of the VMD-GWO-SVR algorithm and the sentiment indicators proposed in this paper, we also conducted two comparative experiments. Experiment 1 involves predicting closing prices using sentiment indicators without performing VMD decomposition, while Experiment 2 involves predicting closing prices without either VMD decomposition or the use of sentiment indicators. The prediction results are shown in Table7.

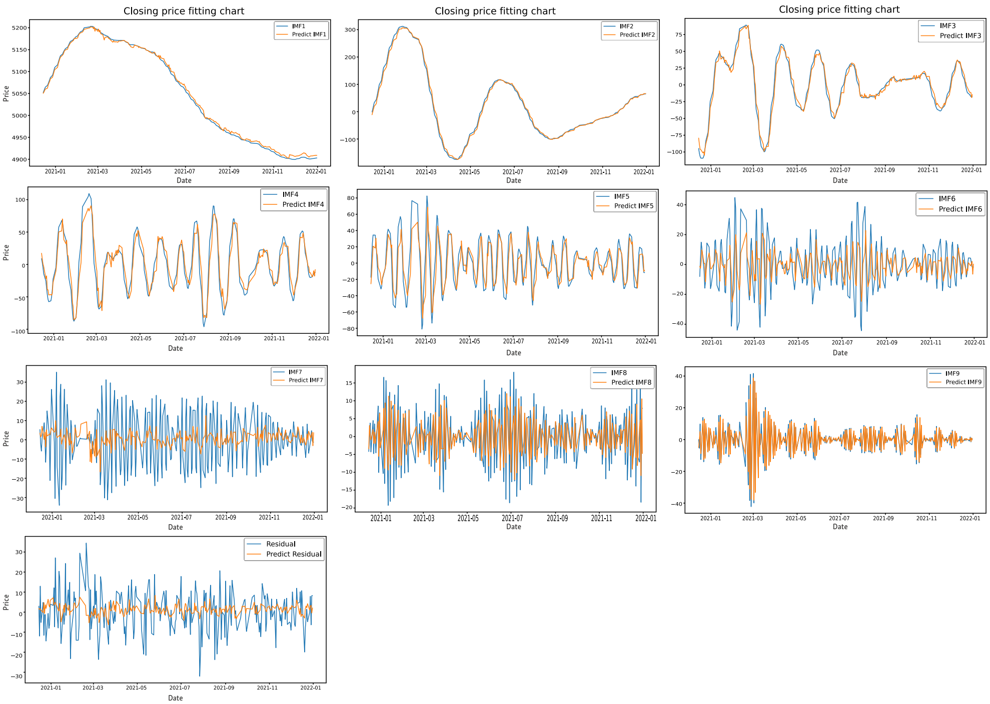

Figure 14. Prediction results of each component series based on GWO-SVR model

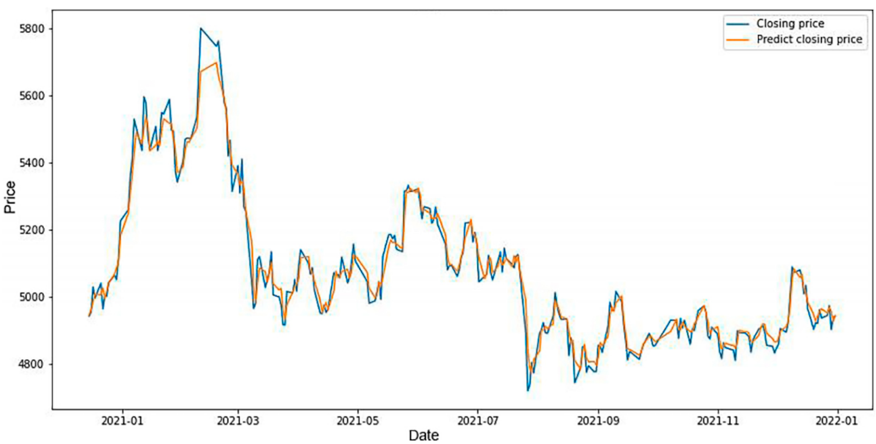

Figure 15. Prediction Results of the Closing Price of the CSI 300 Index Futures

Table 7. Comparative Experiment Results

	R2	MSE	MAE	MAPE
GWO-SVR (no sentiment)	0.9289	3408.14	43.96	0.0086
VMD-GWO-SVR (no sentiment)	0.9763	1132.97	25.07	0.0049
VMD-GWO-SVR (having sentiment)	0.9780	1055.01	24.13	0.0047

From the results in Table 7, we can clearly see that our proposed VMD-GWO-SVR algorithm demonstrates a significant improvement in prediction outcomes compared to the GWO-SVR algorithm without VMD decomposition. The VMD-GWO-SVR (no sentiment) has R^2 = 0.9763 MSE = 1132.97, MAE = 25.07, and MAPE = 0.0049, all of which are significantly better than the results obtained by the GWO-SVR (no sentiment) algorithm. When we introduce the investor sentiment indicators, the prediction effect is further enhanced. In the predictions of VMD-GWO-SVR (having sentiment), MSE, MAE, and MAPE are all lower than the results without investor sentiment indicators. Therefore, it can be confirmed that the VMD-GWO-SVR algorithm proposed in this paper significantly outperforms the traditional GWO-SVR algorithm, and the investor sentiment indicators proposed herein can further enhance the prediction effect.

5.2. Implementation of quantitative investment strategy based on VMD-GWO-SVR model

In Section 4.3, the idea of quantitative investment strategy in this paper is introduced. The first step is to divide the market state according to the size of the two moving averages SMA10 and SMA30 of the sentiment indicators. When SMA10 is higher than SMA30, it is divided into a long market and a long position is established; when SMA10 is lower than SMA30, it is divided into a short market and a short position is established. The specific division is shown in Figure 16 below.

In Figure 16, a status tag of 1 indicates a long market and a status tag of –1 indicates a short market. The back testing period is from December 14, 2020 to December 31, 2021, including 132 days in the long market and 124 days in the short market. Then, implement an investment strategy based on the segmented market and the closing price forecast obtained in Section 5.1. The back testing platform of this paper selects BackTrader third party library of Python for back testing, and sets the volume of each transaction is 10 lots, the initial capital is 10,000,000, the sliding point is 0.20%, the stop loss line is 20% of the principal and the handling fee is 0.23%.

Figure 16. Market State Segmentation Chart

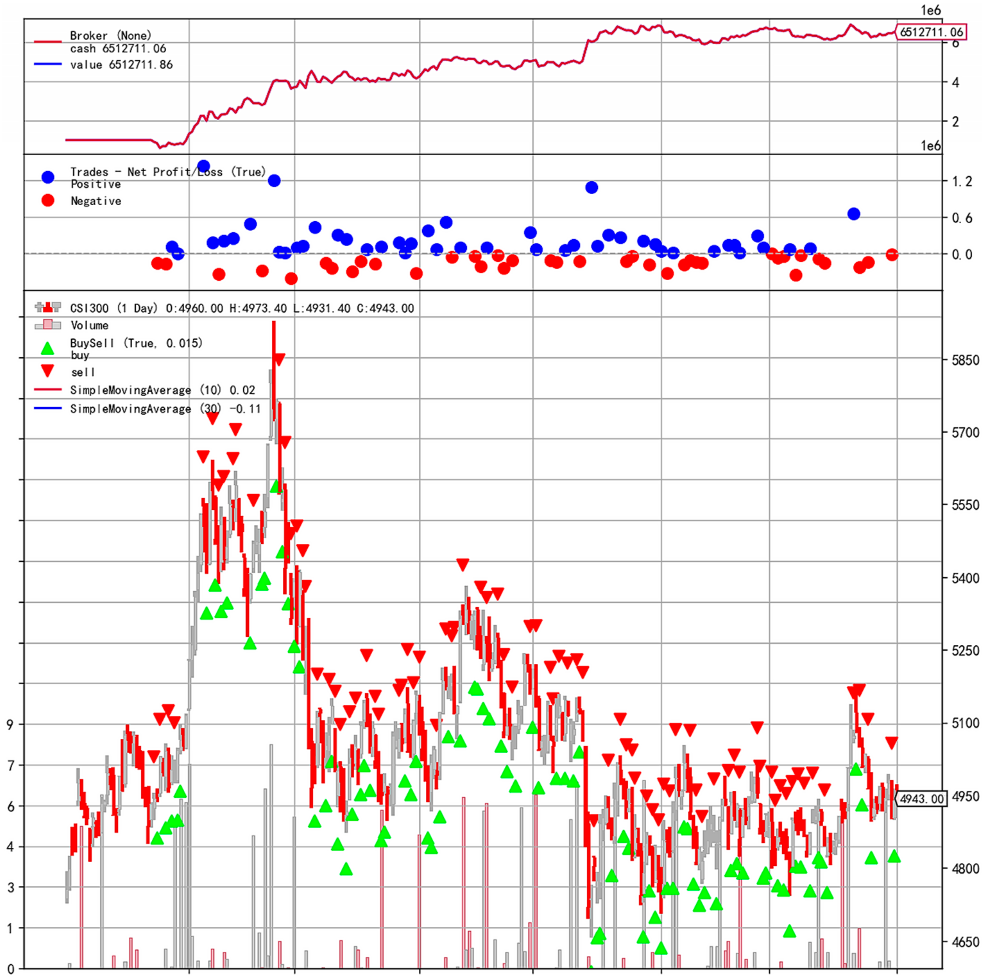

Figure 17. Back test result graph

Back testing results from December 14, 2020 to December 31, 2021 are shown in Figure 17. As shown in the first part of the figure, the initial capital is 1000000. After quantitative investment, the total capital reaches 6312009, and the net income is 5312009. As shown in the second part of the chart above, there are 45 profitable trades and 39 losing trades, for a total of 84 closings. The longest winning streak is 6 times, and the longest losing streak is 4 times. The third part of the above chart gives the time of entering and exiting the market, the green triangle means entering the market, and the red triangle means exiting the market. At the same time, the 10-day SMA10 (red line) and 30-day SMA30 (blue line) of the sentiment indicators are also drawn.

The following Table 8 gives the evaluation metrics for this strategy.

The strategy has an annualized return of 504.52%, a cumulative return of 531.20%, a Sharpe ratio of 2.30, a maximum drawdown of 40.16% and a winning probability of 53.57%. Its returns far exceed the returns of the CSI 300 Index Futures. The return of this strategy in the stock index futures market is very considerable.

Table 8. Back testing revenue evaluation indicators

Evaluation indicator	Indicator value
Annualized rate of return	504.52%
Cumulative rate of return	531.20%
Sharpe ratio	2.30
Maximum drawdown	40.16%
Winning probability	53.57%

5.3. Comparative analysis of quantitative strategies and further analysis

In order to evaluate the effect of the proposed quantitative investment strategy that uses sentiment indicators to divide the market, this paper will use a quantitative strategy that does not use sentiment to divide the market for a comparative analysis. The comparative strategy adopts a double moving average strategy, which also belongs to the trend tracking strategy (Liu et al., 2017).

The double moving average strategy is a trading strategy that uses the relative position between the moving average and the predicted closing price to make trend judgments. It requires setting the stride of two moving averages. This paper sets the comparison strategy as follows: First, calculate the 5-day and 30-day moving averages of the closing price sequence, which are recorded as MA5 and MA30 respectively. Then, the buying and selling conditions are judged by comparing the predicted closing price with the size of the two moving averages: when the predicted closing price rises above MA30, the market is considered to be in an upward trend and a long position is established. When in the open position, if it is predicted that the closing price will drop and break through MA5, it is considered that the market price will drop in the short term, and the position will be closed. Choosing a moving average step size of 5 and 30 can to some extent balance the capture of short-term and long-term trends, provide cross signals, and adapt to the needs of medium - and short-term traders. Carry out this comparison strategy, and the backtesting result is shown in the following Figure 18.

After the implementation of the comparison strategy, the total capital rose from 1000000 to 3570886, and the net income was 2570886. During the retracement period, there are 11 profitable trades and 17 losing trades, resulting in a total of 28 closings. The longest winning streak is 3 and the longest losing streak is 5. It can be seen that the risk of this comparison strategy is extremely high, and the number of losing trades is higher than the number of profitable trades.

The following Table 9 gives the evaluation indicators for this strategy, and compares them with the strategy proposed in this paper:

Table 9. Comparison of evaluation indicators

Evaluation indicator	Comparison strategy	Strategy in this paper
Annualized rate of return	246.6%	504.52%
Cumulative rate of return	257.1%	531.20%
Sharpe ratio	2.02	2.30
Maximum drawdown	28.67%	40.16%
Winning probability	31.03%	53.57%

Figure 18. Back test result graph of the comparison strategy

The strategy has an annualized return of 246.6%, a cumulative return of 257.1%, a Sharpe ratio of 2.02, and a maximum drawdown of 28.67%, and a wining probability of 31.03%. It can be found that the annualized rate of return and cumulative rate of return of the proposed strategy are greatly increased compared with the comparison strategy, with the annualized rate of return increasing by 104.59% and the cumulative rate of return increasing by 106.61%. The high return also has a higher probability of winning, which increases by 72.64% compared to the comparison strategy. It can be found through comparative analysis that the quantitative investment strategy for sentiment-segmented markets presented in this paper not only achieves significantly higher annualized and cumulative returns compared to the comparison strategy but also boasts a higher Sharpe ratio and win rate, indicating that it is not only more profitable but also more stable and lower in risk, demonstrating considerable practicality.

6. Conclusions

This study introduces a quantitative investment decision-making model based on investor attention and emotional characteristics. Utilizing the BERT model, this paper analyzes investor review data and incorporates investor attention as a weighting factor to construct sentiment indicators. These indicators are crucial not only as input variables for the prediction model but also as a basis for market segmentation. The proposed prediction model and quantitative trading strategy consider the uncertainty brought about by the development and evolution of unexpected events, as well as the impact of the resultant sentiment fluctuations on the trends of the stock index futures market. Additionally, we develop a VMD-GWO-SVR prediction model, combining multiple algorithms to achieve high accuracy in forecasting. Given the non-stationarity and high volatility of financial time series, the VMD signal decomposition method is used to decompose the closing price, yielding a component series of closing price with different volatility. Then, SVR is used to predict each component. To enhance prediction accuracy, the GWO optimization algorithm is also applied to optimize the parameters. The aggregated predictions of these components provide a comprehensive forecast of closing price series. Finally, the market is divided into bull and bear markets based on investor sentiment indicators. According to different market conditions, different investment strategies are adopted to maximize returns and reduce risks associated with uncertainty. Despite the method's effectiveness in addressing some existing challenges, it might not fully account for certain aspects. For instance, the VMD's decomposition level is fixed based on historical data, which may not adapt well to the evolving futures market, potentially affecting forecast precision. Future research will focus on enhancing stock price prediction accuracy without adding computational complexity.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 72301219), China's National Social Science Foundation (Grant No. 20XSH012), Sichuan Provincial Philosophy and Social Science Foundation (Grant No. SCJJ23ND196), and also supported by the Fundamental Research Funds for the Central Universities (Grant No. JBK2406045).

References

Antweiler, W., & Frank, M. Z. (2001). *Is all that talk just noise? The information content of internet stock message boards* (Sauder School of Business Working Paper). https://doi.org/10.2139/ssrn.282320

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. https://doi.org/10.1016/j.ijforecast.2019.05.010

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–152. https://doi.org/10.1257/jep.21.2.129

Baker, M., Wurgler, J., & Yuan, Y. (2009). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, *104*(2), 272–287. https://doi.org/10.1016/j.jfineco.2011.11.002

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Benveniste, L. M., Ljungqvist, A., Wilhelm Jr., W. J., & Yu, X. (2003). Evidence of information spillovers in the production of investment banking services. *The Journal of Finance*, *58*(2), 577–608. https://doi.org/10.1111/1540-6261.00538

Brooks, C., Rew, A. G., & Ritson, S. (2001). A trading strategy based on the lead–lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting*, *17*(1), 31–44. https://doi.org/10.1016/S0169-2070(00)00062-5

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001

Caldeira, J., & Moura, G. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Revista Brasileira de Finanças*, *11*, 49–80. https://doi.org/10.2139/ssrn.2196391

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, *108*(4), 905–939. https://doi.org/10.2307/2118454

Cheng, R., Yu, J., Zhang, M., Feng, C., & Zhang, W. (2022). Short-term hybrid forecasting model of ice storage air-conditioning based on improved SVR. *Journal of Building Engineering*, *50*, Article 104194. https://doi.org/10.1016/j.jobe.2022.104194

Covel, M. W. (2004). Trend following: How great traders make millions in up or down markets. *Choice Reviews Online*, *42*(10).

Dash, S. R., & Maitra, D. (2018). Does sentiment matter for stock returns? Evidence from Indian stock market using wavelet approach. *Finance Research Letters*, *26*, 32–39. https://doi.org/10.1016/j.frl.2017.11.008

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. Arxiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, *62*(3), 531–544. https://doi.org/10.1109/TSP.2013.2288675

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in Neural Information Processing Systems*, *9*, 155–161.

Fan, F., Weng, Z., & Tian, J. (2024). Impact of government support policies on regional economic resilience under the COVID-19 outbreak. *Technological and Economic Development of Economy*, *30*(1), 74–106. https://doi.org/10.3846/tede.2024.20397

Gao, Y., Zhao, C., Sun, B., & Zhao, W. (2022). Effects of investor sentiment on stock volatility: New evidences from multi-source data in China's green stock markets. *Financial Innovation*, *8*(77). https://doi.org/10.1186/s40854-022-00381-2

Guilbaud, F., & Pham, H. (2013). Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, *13*(1), 79–94. https://doi.org/10.1080/14697688.2012.708779

Gupte, A., Joshi, S., Gadgul, P., & Kadam, A. (2014). Comparative study of classification algorithms used in sentiment analysis. *International Journal of Computer Science and Information Technologies*, *5*(5), 6261–6264.

Hirshleifer, D. (2015). Behavioral finance. *Annual Review of Financial Economics*, *7*(1), 133–159. https://doi.org/10.1146/annurev-financial-092214-043752

Hong, W.-C., Dong, Y., Chen, L.-Y., & Wei, S.-Y. (2011a). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. *Applied Soft Computing*, *11*(2), 1881–1890. https://doi.org/10.1016/j.asoc.2010.06.003

Hong, W.-C., Dong, Y., Zheng, F., & Wei, S. Y. (2011b). Hybrid evolutionary algorithms in a SVR traffic flow forecasting model. *Applied Mathematics and Computation*, *217*(15), 6733–6747. https://doi.org/10.1016/j.amc.2011.01.073

Hou, Y., Li, S., & Wen, F. (2019). Time-varying volatility spillover between Chinese fuel oil and stock index futures markets based on a DCC-GARCH model with a semi-nonparametric approach. *Energy Economics*, *83*, 119–143. https://doi.org/10.1016/j.eneco.2019.06.020

Hu, H., Zhang, J., & Li, T. (2020). A comparative study of VMD-based hybrid forecasting model for nonstationary daily streamflow time series. *Complexity*, *2020*, 1–21. https://doi.org/10.1155/2020/4064851

Hu, Y., Feng, B., Zhang, X., Ngai, E. W. T., & Liu, M. (2015). Stock trading rule discovery with an evolutionary trend following model. *Expert Systems with Applications*, *42*(1), 212–222. https://doi.org/10.1016/j.eswa.2014.07.059

James, J. (2003). Simple trend-following strategies in currency trading. *Quantitative Finance*, *3*(4), C75–C77. https://doi.org/10.1088/1469-7688/3/4/604

Lan, Q., Xiong, Q., He, L., & Ma, C. (2018). Individual investment decision behaviors based on demographic characteristics: Case from China. *PLOS ONE*, *13*(8), 1–16. https://doi.org/10.1371/journal.pone.0201916

Lee, C. M. C., & Swaminathan, B. (2000). Price momentum and trading volume. *The Journal of Finance*, *55*(5), 2017–2069. https://doi.org/10.1111/0022-1082.00280

Li, F., Li, R., Tian, L., Chen, L., & Liu, J. (2019). Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions. *Mechanical Systems and Signal Processing*, *116*, 462–479. https://doi.org/10.1016/j.ymssp.2018.06.055

Li, Y., Bu, H., Li, J., & Wu, J. (2020a). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting*, *36*(4), 1541–1562. https://doi.org/10.1016/j.ijforecast.2020.05.001

Li, Y., Shen, D., Wang, P., & Zhang, W. (2020b). Does intraday time-series momentum exist in Chinese stock index futures market? *Finance Research Letters*, *35*, Article 101292. https://doi.org/10.1016/j.frl.2019.09.007

Li, W., Wang, G.-G., & Gandomi, A. H. (2021). A survey of learning-based intelligent optimization algorithms. *Archives of Computational Methods in Engineering*, *28*(5), 3781–3799. https://doi.org/10.1007/s11831-021-09562-1

Lin, G., Lin, A., & Gu, D. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences*, *608*, 517–531. https://doi.org/10.1016/j.ins.2022.06.090

Lowry, M., & Schwert, G. W. (2002). IPO market cycles: Bubbles or sequential learning? *The Journal of Finance*, *57*(3), 1171–1200. https://doi.org/10.1111/1540-6261.00458

Liu, X., An, H., Wang, L., & Guan, Q. (2017). Quantified moving average strategy of crude oil futures market based on fuzzy logic rules and genetic algorithms. *Physica A: Statistical Mechanics and Its Applications*, *482*, 444–457. https://doi.org/10.1016/j.physa.2017.04.082

Lu, C.-J., Lee, T.-S., & Chiu, C.-C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, *47*(2), 115–125. https://doi.org/10.1016/j.dss.2009.02.001

Luukka, P., Pätäri, E., Fedorova, E., & Garanina, T. (2016). Performance of moving average trading rules in a volatile stock market: The Russian evidence. *Emerging Markets Finance and Trade*, *52*(10), 2434–2450. https://doi.org/10.1080/1540496X.2015.1087785

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, *11*(6), 601–618. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<601::AID-JAE417>3.0.CO;2-T

Mbanga, C., Darrat, A. F., & Park, J. C. (2019). Investor sentiment and aggregate stock returns: The role of investor attention. *Review of Quantitative Finance and Accounting*, *53*(2), 397–428. https://doi.org/10.1007/s11156-018-0753-2

Mehtab, S., Sen, J., & Dutta, A. (2021). Stock price prediction using machine learning and LSTM-based deep learning models. In S. M. Thampi, S. Piramuthu, K.-C. Li, S. Berretti, M. Wozniak, & D. Singh (Eds.), *Machine learning and metaheuristics algorithms, and applications: Vol. 1366. Communications in computer and information science* (pp. 88–106). Springer. https://doi.org/10.1007/978-981-16-0419-5_8

Mensi, W., Hammoudeh, S., Shahzad, S. J. H., Al-Yahyaee, K. H., & Shahbaz, M. (2017). Oil and foreign exchange market tail dependence and risk spillovers for MENA, emerging and developed countries: VMD decomposition based copulas. *Energy Economics*, *67*, 476–495. https://doi.org/10.1016/j.eneco.2017.08.036

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, *69*, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Montoya-Cruz, E., Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. Á. (2020). Exploring arbitrage strategies in corporate social responsibility companies. *Sustainability*, *12*(16), Article 6293. https://doi.org/10.3390/su12166293

Niu, D., Ji, Z., Li, W., Xu, X., & Liu, D. (2021). Research and application of a hybrid model for mid-term power demand forecasting based on secondary decomposition and interval optimization. *Energy*, *234*, Article 121145. https://doi.org/10.1016/j.energy.2021.121145

Olgun, O., & Yetkiner, I. H. (2011). Determination of optimal hedging strategy for index futures: Evidence from Turkey. *Emerging Markets Finance and Trade*, *47*(6), 68–79. https://doi.org/10.2753/REE1540-496X470604

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016, October 3–5). Sentiment analysis of Twitter data for predicting stock market movements. In *Proceedings of the 2016 International Conference on Signal Processing*, *Communication*, *Power and Embedded System* (*SCOPES*) (pp. 1345–1350). Paralakhemundi, India. IEEE. https://doi.org/10.1109/SCOPES.2016.7955659

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (*LREC'10*). LREC 2010. Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf

Pätäri, E., & Vilska, M. (2014). Performance of moving average trading strategies over varying stock market conditions: The Finnish evidence. *Applied Economics*, *46*(24), 2851–2872. https://doi.org/10.1080/00036846.2014.914145

Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, *16*(3), 394–408. https://doi.org/10.1016/j.jempfin.2009.01.002

See-To, Eric. W. K., & Yang, Y. (2017). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets*, *27*(3), 283–296. https://doi.org/10.1007/s12525-017-0254-5

Selmi, R., Hammoudeh, S., Errami, Y., & Wohar, M. E. (2021). Is COVID-19 related anxiety an accelerator for responsible and sustainable investing? A sentiment analysis. *Applied Economics*, *53*(13), 1528–1539. https://doi.org/10.1080/00036846.2020.1834501

Seok, S. I., Cho, H., & Ryu, D. (2019). Firm-specific investor sentiment and daily stock returns. *The North American Journal of Economics and Finance*, *50*, Article 100857. https://doi.org/10.1016/j.najef.2018.10.005

Shu, H. C. (2010). Investor mood and financial markets. *Journal of Economic Behavior* & *Organization*, *76*(2), 267–282. https://doi.org/10.1016/j.jebo.2010.06.004

Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. *The Review of Financial Studies*, *19*(4), 1531–1565. https://doi.org/10.1093/rfs/hhj032

Su, C. W., Xi, Y., Tao, R., & Umar, M. (2022). Can Bitcoin be a safe haven in fear sentiment? *Technological and Economic Development of Economy*, *28*(2), 268–289. https://doi.org/10.3846/tede.2022.15502

Su, W., Lei, Z., Yang, L., & Hu, Q. (2019). Mold-level prediction for continuous casting using VMD–SVR. *Metals*, *9*(4), Article 458. https://doi.org/10.3390/met9040458

Sun, Y., Fang, M., & Wang, X. (2018). A novel stock recommendation system using Guba sentiment analysis. *Personal and Ubiquitous Computing*, *22*(3), 575–587. https://doi.org/10.1007/s00779-018-1121-x

Szakmary, A. C., Shen, Q., & Sharma, S. C. (2010). Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking & Finance*, *34*(2), 409–426. https://doi.org/10.1016/j.jbankfin.2009.08.004

Taş, O., & Akdağ, Ö. (2012). Trading volume trend as the investor's sentiment indicator in Istanbul stock exchange. *Doğuş Üniversitesi Dergisi*, *13*(2), 290–300.

Wang, D. (2020). Quantitative investment model based on data mining. *Revista Internacional de Métodos Numéricos Para Cálculo y Diseño En Ingeniería*, *36*(1), 1–7. https://doi.org/10.23967/j.rimni.2020.03.006

Wang, X., Yu, Q., & Yang, Y. (2018). Short-term wind speed forecasting using variational mode decomposition and support vector regression. *Journal of Intelligent & Fuzzy Systems*, *34*(6), 3811–3820. https://doi.org/10.3233/JIFS-169553

Wen, F., Xu, L., Ouyang, G., & Kou, G. (2019). Retail investor attention and stock price crash risk: Evidence from China. *International Review of Financial Analysis*, *65*, Article 101376. https://doi.org/10.1016/j.irfa.2019.101376

Xie, W., Tang, Y., Xu, Z., Zhang, X., & Lai, D. (2023). The impact of the infodemic on the stock market under the COVID-19: Taking the investors' information infection index as the intermediary variable. *Technological and Economic Development of Economy*, *29*(2), 653–676. https://doi.org/10.3846/tede.2023.18571

Yaslan, Y., & Bican, B. (2017). Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. *Measurement*, *103*, 52–61. https://doi.org/10.1016/j.measurement.2017.02.007

Zang, H., Cheng, L., Ding, T., Cheung, K. W., Liang, Z., Wei, Z., & Sun, G. (2018). Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Generation, Transmission & Distribution*, *12*(20), 4557–4567. https://doi.org/10.1049/iet-gtd.2018.5847

Zaremba, A., Kizys, R., Aharon, D. Y., & Demir, E. (2020). Infected markets: Novel coronavirus, government interventions, and stock return volatility around the globe. *Finance Research Letters*, *35*, Article 101597. https://doi.org/10.1016/j.frl.2020.101597

Zhang, H., Wang, H., Wei, G., & Chen, X. (2023). An integrated decision support system for stock investment based on spherical fuzzy PT-EDAS method and MEREC. *Technological and Economic Development of Economy*, *29*(4), 1353–1381. https://doi.org/10.3846/tede.2023.19123

Zhao, B. (2020). COVID-19 pandemic, health risks, and economic consequences: Evidence from China. *China Economic Review*, *64*, Article 101561. https://doi.org/10.1016/j.chieco.2020.101561

Zhou, H., Geppert, J., & Kong, D. (2010). An anatomy of trading strategies: Evidence from China. *Emerging Markets Finance and Trade*, *46*(2), 66–79. https://doi.org/10.2753/REE1540-496X460205

Zhou, M., Hu, T., Bian, K., Lai, W., Hu, F., Hamrani, O., & Zhu, Z. (2021a). Short-term electric load forecasting based on variational mode decomposition and grey wolf optimization. *Energies*, *14*(16). https://doi.org/10.3390/en14164890

Zhou, Z., Gao, M., Xiao, H., Wang, R., & Liu, W. (2021b). Big data and portfolio optimization: A novel approach integrating DEA with multiple data sources. *Omega*, *104*, Article 102479. https://doi.org/10.1016/j.omega.2021.102479

Zou, H., Tang, X., Xie, B., & Liu, B. (2015, December 7–9). Sentiment classification using machine learning techniques with syntax features. In *Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence* (*CSCI*) (pp. 175–179). Las Vegas, NV, USA. IEEE. https://doi.org/10.1109/CSCI.2015.44

Zuo, G., Luo, J., Wang, N., Lian, Y., & He, X. (2020). Two-stage variational mode decomposition and support vector regression for streamflow forecasting. *Hydrology and Earth System Sciences*, *24*(11), 5491–5518. https://doi.org/10.5194/hess-24-5491-2020