ISSN: 2029-4913 / eISSN: 2029-4921

VILNIUS TECH
Vilnius Gediminas Technical University

TECHNOLOGICAL and ECONOMIC DEVELOPMENT of ECONOMY

2024
Volume 30
Issue 3
Pages 527–561
https://doi.org/10.3846/tede.2023.18672

QUANTITATIVE INVESTMENT DECISIONS BASED ON MACHINE LEARNING AND INVESTOR ATTENTION ANALYSIS

Jie GAO[1], Yunshu MAO[1], Zeshui XU[2]✉, Qianlin LUO[1]

[1] *School of Business Administration, Southwestern University of Finance and Economics, 610207 Chengdu, China*
[2] *Business School, Sichuan University, 610064 Chengdu, China*

Article History:
- received 31 May 2022
- accepted 02 February 2023
- first published online 24 August 2023

Abstract. According to the trading rules and financial data structure of the stock index futures market, and considering the impact of major emergencies, we intend to build a quantitative investment decision-making model based on machine learning. We first adopt the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) signal decomposition technology to separate the short-term noise, cycle transformation and long-term trend from the original series, and use the CSI 500 Baidu index series to reflect the investors' attention, which provides data support for establishing a more effective forecasting model. Then, the CEEMDAN-BP neural network model is designed based on the obtained effective information of low-frequency trend series, investor attention index and CSI 500 stock index futures market transaction data. Finally, an Attention-based Dual Thrust quantitative trading strategy is proposed and optimized. The optimized Attention-based Dual Thrust strategy solves the core problem of breakout interval determination, effectively avoids the risk of subjective selection, and can meet investors' different risk preferences. The quantitative investment decision-making model based on CEEMDAN-BP neural network utilizes the advantages of different algorithms, avoids some defects of a single algorithm, and can make corresponding adjustments according to changes in investors' attention and the occurrence of emergencies. The results show that considering investor attention can not only improve the predictive ability of the model, but also reduce the cognitive bias of the market, effectively control risks and obtain higher returns.

Keywords: behavioral economics, decision making, signal decomposition, investor attention.

JEL Classification: C61, D91, E37, G41.

✉Corresponding author. E-mail: ***xuzeshui@263.net***

Introduction

With the in-depth application of emerging technologies such as big data, artificial intelligence, and blockchain in the financial industry, the role of technology in finance has been continuously strengthened. Among them, computer technology represented by machine learning is also expanding the depth and breadth of its application in the field of quantitative investment (De Prado, 2018; Krollner et al., 2010). Unlike previous investment methods that were mainly based on human subjective decisions, quantitative investment uses advanced mathematical models and computer programming strategies to automatically buy and sell (Chan, 2021; Schumaker & Chen, 2009). It can effectively prevent investors from making wrong judgments due to emotional fluctuations, especially when they are affected by market sentiment and uncertain events. Therefore, the application of quantitative investment based on machine learning in the financial market has received more and more attention, and it is necessary to propose corresponding quantitative investment strategies to cope with the rapid changes in the financial market.

In addition, with the development of financial markets, many abnormal phenomena cannot be explained by traditional economic theory (Daniel et al., 2002), which assumes that investors are homogeneous, completely rational, and have complete information in the market. Therefore, methods such as behavioral economics have emerged (Mullainathan & Thaler, 2000; Camerer & Loewenstein, 2004), which mainly study the impact of investor psychology and cognitive biases on asset pricing. Hirshleifer and Teoh (2003) introduced the findings of psychological "limited attention" into the behavioral finance framework, assuming that investors' attention and processing capacity are limited, and believed that the consequence of limited attention is that the disclosure of informationally equivalence can have different effects on investor perceptions, which in turn affects the market prices. Based on this, scholars have used indicators such as the number of news reports, trading volume (Peng & Xiong, 2006), and search volume index (Da et al., 2011) to construct an investor attention index and determine its impact on the stock market performance. Studies have shown that the Internet search index can more intuitively and accurately measure the degree of investor attention (Vozlyublennaia, 2014; Smales, 2021).

At the same time, the occurrence of global uncertain events has become the norm. The financial market often fluctuates due to various risk events, and spreads rapidly through the network and affects the attention distribution and behavioral decisions of individual investors. COVID-19 is currently the most destructive and influential uncertain event, and investors' emotions and psychology have been seriously affected (Zhang et al., 2020; Sansa, 2020). However, existing research on investment decision-making does not consider the impact of these uncertain events on the market. On the other hand, there are still a large number of individual investors in China's stock index futures market, but these individual investors generally lack basic knowledge and easily affected by market uncertain events (Kou et al., 2019). The occurrence of uncertain events and changes in investor attention will lead to changes in market conditions (Mbanga et al., 2019). In this new market environment, this paper takes COVID-19 as an example to explore the impact of considering risk events on quantitative investment decision-making models, and deepen the understanding of the relationship between investor attention and the stock market. Better formulate quantitative investment strategies and make dynamic adjustments to risk events. Specifically, the innovations of this study are as follows:

First, consider the potential impact of investor attention on market trends. In this paper, Baidu Index is used to represent investors' attention to stock index futures. Compared with text data (financial news, investor reviews, etc.), the Internet search data is real-time data automatically aggregated and weighted by search engines, especially in emergencies, which has a faster response speed. At the same time, due to the anonymity, investors are more inclined to express their true emotions, making Internet search data more reflective of investors' psychological and behavioral trends. Therefore, the Baidu Index can more effectively reflect changes in investors' attention to the market. Introducing the investor attention index into the forecasting model can make corresponding adjustments according to changes in market sentiment and the occurrence of uncertain events, thereby improving the forecasting accuracy of stock index futures prices.

Secondly, construct the CEEMDAN-BP neural network prediction model and apply it to the prediction of the stock index futures closing price sequence trend. In this paper, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) signal decomposition technique is applied to the quantitative investment decision-making of stock index futures. It reduces the interference of noise and periodic transformations to market fluctuations, and obtains signal combinations of different frequencies, thereby providing more effective market trend fluctuation information for quantitative investment decisions. Then, the obtained low-frequency trend sequence is used as one of the input variables of the BP neural network to construct the CEEMDAN-BP neural network prediction model. Through experimental comparison, the prediction accuracy of the CEEMDAN-BP model established in this paper is better than that of EMD-BP and EEMD-BP neural network models.

Finally, two Attention-based Dual Thrust strategies are proposed. This paper considers the influence of investors' attention on the stock price trend, and takes investors' attention as one of the input variables of the CEEMDAN-BP neural network model to predict the closing price sequence. Based on this, an Attention-based Dual Thrust strategy is proposed, and then we further optimize it. The optimized Attention-based Dual Thrust strategy uses the investor attention index to determine the parameters K_1 and K_2 in the Dual Thrust strategy trading algorithm, thus solving the core problem of how to determine the breakout interval in the dual-thrust strategy, effectively avoiding possible risk due to subjective selection. In the optimized Attention-based Dual Thrust strategy, K_1 and K_2 change with the investor attention, and the coefficients can also be set according to the different risk preferences of investors, effectively controlling risks and obtaining higher returns.

The remainder of this paper is organized as follows: The following Section 1 introduced the literature review. Section 2 describes the research methodology for the construction of the investor attention index and the extraction of financial signals. Section 3 reports the quantitative analysis based on CEEMDAN-BP neural network. Section 4 constructs and implements the Attention-based Dual Thrust strategy, and conducts optimization and comparative experiments on the strategy. The paper ends up with some conclusions.

1. Literature review

1.1. Measurement of investor attention

The existing research on investor attention is mainly derived from the finite attention theory (Kahneman, 1973), which explains that attention is a scarce resource, that is, attention to one thing will inevitably reduce attention to another thing. Furthermore, "investor attention" (our focus) is defined as awareness as to whether a piece of information exists, and greater investor attention can exacerbate the effect of investor behavioral biases on asset prices (Smales, 2021). Compared with traditional financial theory, limited attention theory can be used to explain a large number of anomalies in financial markets, and the way investors focus and analyze information can have a significant impact on stock prices (Da et al., 2011). Some researchers have considered stock returns (Audrino et al., 2020), asset prices (Andrei & Hasler, 2015), and advertising spending (Lou, 2014) as indicators for investor attention.

With the development of the Internet, more and more studies use the Internet search index to measure investor attention. When users use search engines to search for stock index futures, users' attention degree can be displayed directly. Da et al. first proposed the use of Google search data, namely Search Volume Index (SVI) as a measure of investor attention to study the relationship between SVI and stock price volatility (Da et al., 2011). Audrino et al. (2020) have studied the influence of attention factors on stock market volatility using a novel large dataset that integrates social media, news articles, information consumption, and search engine data. Sampath et al. (2022) studied stock price volatility by analyzing investor attention to tweets and messages from specific companies or public figures. Compared with traditional approaches that use indirect indicators as proxies for investor attention, Internet search data is less susceptible to other factors and is the result of investors' active inquiries, especially in emergencies, which respond faster.

At the same time, Baidu Index is a data analysis platform based on massive Internet user behavior data. According to global statistics from gs.statcounter.com, in March 2022, Baidu search engine accounts for approximately 84% market share in China. Therefore, the Baidu Index has become an effective proxy for investor attention (Su & Wang, 2021; Fang et al., 2020). Deng et al. (2022) constructed investor attention indices from the Baidu Index by crawling the code and abbreviation of each stock. Liu et al. (2021) also used Baidu Index as an indicator of investor attention and found that an increase in investor attention may lead to a rise in stock prices. Compared with text data (financial news, investor reviews, etc.), the Internet search data obtained through Baidu Index is real-time data automatically aggregated and weighted by search engines, which can more directly and accurately reflect investors' concerns about the market (Zhang et al., 2021). Therefore, this paper attempts to explore whether the investor attention indices to CSI 500 stock index futures based on the Baidu Index can improve the accuracy of the trend prediction of the closing price sequence, and establish a quantitative trading strategy based on investor attention.

1.2. Time series forecasting methods

Due to the influence of many factors such as economic fundamentals, policies, and herd psychology, financial time series have the characteristics of high volatility, nonlinearity, nonstationarity, and chaos. Therefore, the prediction of financial time series has always been a difficult point and a research hotspot in the financial field. Through the summary of relevant literature, time series forecasting methods can be divided into the two categories: traditional time series methods and machine learning algorithms. Traditional time series methods mostly use econometric statistical models for forecasting calculations. The classic ones are the Moving Average (MA) model, the Auto Regressive (AR) model, the Auto Regressive and Moving Average (ARMA) model, and the Auto Regressive Integrated Moving Average (ARIMA) model (Makridakis & Hibon, 1997; Siami-Namini & Namin, 2018). The traditional methods are usually used to explain linear problems, but they cannot explain complex real-world problems well, and there are large errors in long-term time prediction. Compared with the traditional time series models, machine learning models can capture more fluctuations of time series (Kohzadi et al., 1996). Therefore, many studies have begun to try to use machine learning models to process financial data or investment forecasting (Henrique et al., 2019).

With the continuous development and improvement of artificial intelligence learning methods, machine learning-related technologies and models have been widely used in the financial field (Sezer et al., 2020). The main methods of machine learning for time series forecasting can be briefly classified into three categories: supervised learning, unsupervised learning, and reinforcement learning (Masini et al., 2021). The main methods of supervised learning are support vector machines (SVM), random forests, logistic regression, and neural networks. Among these supervised learning algorithms, more and more researchers have applied neural network machine algorithms to stock price forecasting and found that the prediction results with high accuracy can be achieved (Zhang & Lou, 2021; Garcia et al., 2018; Kuang et al., 2017; Jiang, 2022). For example, Zhang and Lou (2021) applied neural network and BP algorithm to the classification and prediction of stock price patterns with an accuracy rate of 73.29%. Jiang (2022) proposed an optimized particle swarm optimization-neural networks (PSO-BP) neural network model. This model is more accurate in predicting stock prices with less error, and the prediction of future trends is consistent with the actual trend.

Therefore, this paper will study the application of the BP neural network model in the stock index futures market. Compared with other time series models, the BP neural network can approximate any function, including nonlinear functions. At the same time, in order to improve the prediction accuracy, this paper adopts the CEEMDAN method to decompose and denoise the input data, so as to extract effective trend fluctuations in the sequence.

1.3. Quantitative investment strategies

There are different investment strategies in financial markets. According to the flexibility of investment decision-making, they can usually be divided into active strategies and passive strategies. An active investment strategy means that investors strive to maximize investment returns through active selection of securities and timing of purchases. While passive investment strategy refers to the purchase of investment securities from the perspective of long-term income and limited management, not actively seeking to outperform the market, but trying to replicate the performance of the index (Sushko & Turner, 2018). Based on the concept of passive strategies, Olgun and Yetkiner (2011) developed a dynamic hedging strategy applicable to the index futures market. Passive strategies use rules-based investing and track the index by holding the constituent assets or automatically selecting representative samples (Anadu et al., 2020). Huang et al. (2022) constructed an enhanced index investment model based on a new smooth nonparametric kernel (NPK) and lower partial moment (LPM).

Meanwhile, active investment strategies include traditional strategies and quantitative strategies, in which traditional investment relies on the investors' experience and intuition, while quantitative investment is more rational and based on specific data analysis (Chen et al., 2020). It uses computer technology, mathematics and statistical theory to carry out investment transactions, and has been widely used and developed because of its stable income. Modern Portfolio Theory (MPT) was first proposed by Harry Markowitz in 1952 (Markowitz, 1968; Mangram, 2013), which is considered to be the basic theory of quantitative trading. Subsequently, Sharp (1964) proposed the Capital Asset Pricing Model (CAPM), a major innovation in predicting risk and return. With the development of the stock market and internet technology, combined with models and algorithms, quantitative investment is more accurate and efficient, and has been widely used. At present, quantitative investment strategies are mainly divided into four categories: market-neutral strategies (Olgun & Yetkiner, 2011), trend-following strategies (Brooks et al., 2001; Li et al., 2020; Szakmary et al., 2010), arbitrage strategies (Caldeira & Moura, 2013; Montoya-Cruz et al., 2020) and high-frequency strategies (Guilbaud & Pham, 2013), as shown in Table 1.

Table 1. Literature on investment strategies in the futures market

Type	strategy	References
Trend-following strategy	the lead–lag relationship strategy	Brooks et al. (2001)
	intraday-momentum strategy	Li et al. (2020)
	dual moving average crossover and channel strategies	Szakmary et al. (2010)
Market-neutral strategy	dynamic hedging strategy based on binary GARCH estimation	Olgun and Yetkiner (2011)
Arbitrage strategy	statistical arbitrage strategy based on cointegration	Caldeira and Moura (2013)
	statistical arbitrage strategy of paired transactions	Montoya-Cruz et al. (2020)
High-frequency strategy	optimal market-making policies in a limit order book (LOB)	Guilbaud and Pham (2013)

CSI 500 stock index futures are financial derivatives with a wide range of fluctuations and high frequency of changes, so they are more suitable for trend-following strategies. Trend-following strategies perform well in stock index futures trading (Hurst et al., 2017), mainly including Dual Thrust, R-Breaker, Fiari Four Price, Aerial Garden, Channel Breakthrough, etc. (Covel, 2006). Among them, the Dual Thrust trading algorithm is a famous strategy developed by Michael Chalek (Pruitt & Hill, 2012), and was rated as one of the top ten most profitable investment strategies by "Future Thruth" magazine. This strategy is easy to understand, has few parameters, and is widely used in markets such as stocks, currencies, bonds, and futures. In the Dual Thrust trading algorithm, the determination of the breakout interval is very important, and it is also the core of the strategy. Therefore, this paper proposes an Attention-based Dual Thrust strategy, and further optimizes the proposed strategy to solve the core problem.

2. Investor attention index and financial signal extraction

The rapid development of psychology and behavioral economics has prompted scholars to study investors' attention. However, in the existing studies, investor's attention is rarely applied to machine learning models, because machine learning cannot obtain intuitive quantitative models, and most of them are applied to time series forecasting models. This section intends to construct the investor attention index of the CSI 500 index and decompose the signal of the closing price sequence of the CSI 500 stock index futures. Combining the two to forecast and analyze market data can improve the accuracy and help investment decisions.

2.1. Construction of Investor Attention Index

Predicting future trends has always been an important part of stock index futures market research. Researchers generally introduce large amounts of original analysis data into complex forecasting models, but the validity of the forecasts has always been questioned. Because there is often a lag in the collection and publication of traditional data. In contrast, Internet search data obtained through the Baidu Index is different from general market transaction information. It is timely data obtained through automatic aggregation and weighting by search engines, and has higher data quality. At the same time, due to the anonymity of the Internet, investors are more inclined to express their true emotions on the Internet, so the Internet search data can better reflect investors' psychological and behavioral trends. Therefore, the Baidu Index can more accurately reflect the degree of investors' attention to the market.

In this study, we use the Baidu Index to represent the investor attention. Baidu takes up most of the search engine market in China, and across the word, Baidu has more than 1.1 billion users as of 2020 (source: Statcounter GlobalStats, 2020). Being China's top search engine, Baidu Index is a data analysis platform based on the behavior data of massive netizens, and it is also one of the most important statistical analysis platforms in the current Internet era. It counts the number of searches for a certain keyword on Baidu, which is the result of investors' active inquiries and can directly reflect their attention. According to global statistics from gs.statcounter.com, Baidu accounts for approximately 84% market share in the field of searching market in China. When investor inquire about stock index futures information, they generally use three types of search keywords, such as stock abbreviations, corporate names or stock codes. But most investors will use the stock abbreviations as keywords. This is because when investors use the company name as a keyword to search, their purpose may not be for investment, but to obtain the company's product information, so the Baidu Index obtained by using the company name as a keyword is noisy and cannot accurately reflect the investor attention. As for the stock code, it is usually more complicated and difficult to remember, which does not conform to the inquiry habits of investors. In addition, the code 399905 of CSI 500 is not included in the Baidu Index, so it is not used as a query keyword. To sum up, this research uses the stock abbreviations ("CSI 500" and "CSI 500 Index") as keywords to obtain the investor Attention Index. Figure 1 is a sequence diagram of the search times for the keywords "CSI 500" and "CSI 500 Index".

The vertical line in Figure 1 is located on January 20, 2020, when academician Nanshan Zhong announced the human-to-human transmission of COVID-19. Since then, the domestic epidemic has received widespread attention. It can be clearly seen that the search volume of the keywords of "CSI 500" and "CSI 500 Index" dropped sharply after that day, indicating that investors' attention has shifted from the securities market to the epidemic. A few days later, investors' attention turned to the market again. Based on this turning point, we download

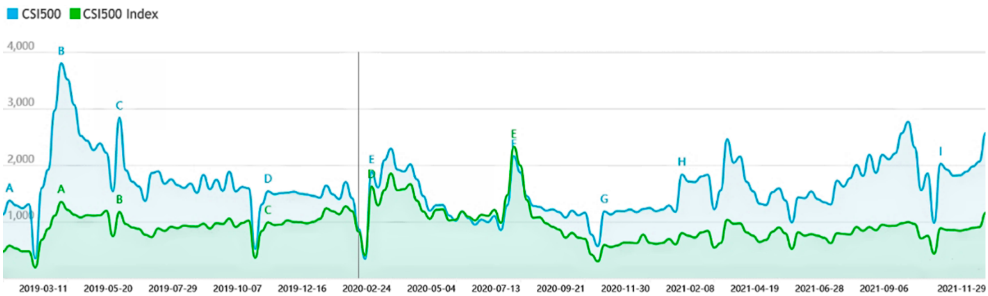

Figure 1. Baidu Index

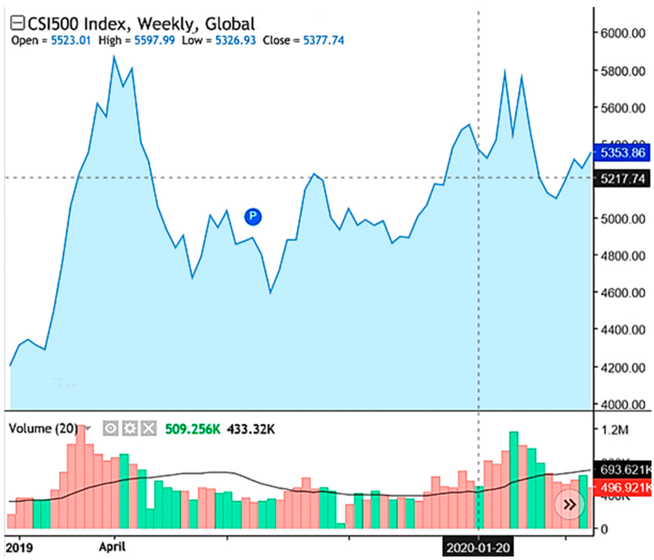

Figure 2. CSI 500 Timing diagrams

the sequence diagram of the CSI 500 Index from Investing.com, as shown in Figure 2. It can be clearly seen that the price of CSI 500 stock index futures also dropped sharply on January 20, and the trading volume also declined. Therefore, it can be speculated that this is caused by the decline of the investor attention index. Some studies have shown that the Baidu Index hides the attention and interest of Chinese investors and can reflect the trend of stock prices (Shen et al., 2017). Therefore, it is reasonable to introduce the Baidu Index with the keywords of “CSI 500” and “CSI 500 Index” in the forecasting model as investor attention index.

Add the daily Baidu Index of the keywords “CSI 500” and “CSI 500 Index” to get the comprehensive daily attention. Let the Baidu Index of “CSI 500” and “CSI 500 Index” on the day t be N_{t1} and N_{t2} respectively, then the investor attention index of CSI 500 index futures on the day t can be expressed as Att_t:

$$Att_t = N_{t1} + N_{t2}. \tag{1}$$

Exploring the potential impact of investor attention index on the market environment will help investors to correctly understand their own behaviors, decision-making process and individual irrational emotions. Thus, a more accurate understanding of the operating rules of the market can guide more rational trading decisions.

2.2. Signal Decomposition Method

Financial data usually has a large amount of accumulation, a high degree of standardization, and great usable value. However, financial time series has the characteristics of non-linearity, non-stationarity, and multi-scale interval. Due to the influence of market noise, it is difficult to extract and utilize effective information. Therefore, this section uses the signal decomposition technology to decompose the closing price sequence of CSI 500 stock index futures to obtain a series of signals with different frequencies, and to separates the noise sequence, periodic transformations and long-term trends, so as to provide data support for building more effective forecasting model. The Empirical Mode Decomposition (EMD) method is considered a breakthrough in Fourier transform-based linear and steady-state spectral analysis (Yu et al., 2008), which allows the analysis of non-stationary time series with high adaptivity (Li, 2006). EMD decomposes the input signal into several intrinsic mode functions(*IMFs*)and a residual, which consists of the following formula:

$$x(t) = \sum_{i=1}^{k} IMF_i(t) + r_k(t), \tag{2}$$

where $x(t)$ is the original signal, $IMF_i(t)$ is the i-th *IMF* with a total of k components, and $r_k(t)$ is the residual term. The steps of the EMD algorithm are shown in Table 2.

Table 2. EMD algorithm

Algorithm 1: EMD algorithm
Input: original sequence signal Output: a series of *IMF* components
Begin While the original signal $x(t)$ has not been decomposed *Flag* = 1 While *Flag*: Step 1: Calculate the local extreme value point of the original data $x(t)$, and calculate the lower envelope $emin(t)$ formed by the minimal value and the upper envelope $emax(t)$ created by the immense value by interpolation Step 2: Calculate the mean value of the upper and lower envelope $m(t) = (emin(t) + emax(t))/2$ Step 3: Calculate $x(t)$ minus $m(t)$ to get $h(t) = x(t) - m(t)$ Step 4: Determine whether the sequence $h(t)$ meets the criteria, and if it does, it will be used as a new IMF_i component, and then make *Flag* = 0. If it does not, the *Flag* will remain unchanged, and then the above steps will be repeated with $h(t)$ as the basis. Calculated $x(t) - IMF_i$, and the decomposition continues based on this sequence. End

However, the EMD algorithm has problems such as mode mixing, aliasing effect and difficulty in judging stopping conditions. The emergence of Ensemble Empirical Mode Decomposition (EEMD) overcomes the shortcomings of mode mixing (Huang et al., 2009). The steps of the EEMD algorithm are similar to the EMD algorithm in Table 2, except that a white noise signal is added to the original sequence for auxiliary decomposition. It is worth noting that although EEMD solves the problem of mode mixing. However, due to the addition of white noise, the integrated *IMFs* will be biased, and the computational complexity will be greater and more time-consuming. Furthermore, the CEEMDAN method (Yeh et al., 2010) makes up for the deficiency of EEMD and also solves the mode mixing problem in EMD. Since CEEMDAN has good adaptive ability, no information and energy are lost during the decomposition process, and the *IMF* components obtained by decomposition can describe the original information from different frequencies. Therefore, in order to better explain market information, this paper chooses the CEEMDAN algorithm to decompose the financial sequence signal. The steps are shown in Table 3.

In this paper, the CEEMDAN method is used to decompose the closing price sequence of CSI 500 index futures to obtain *IMF* components with different frequencies, and combine the resulting low-frequency *IMF* components into the long-term trend of market changes. The main steps are shown in Figure 3.

In this experiment, the time series of CSI 500 stock index futures closing prices are selected as the research object, and the CSI 500 daily trading closing prices from 2019/1/2 to 2021/12/1 are selected as the sample data. The data comes from China Financial Futures Exchange, such as shown in Figure 4. The original data is used as the input signal to obtain

Table 3. CEEMDAN algorithm

Algorithm 2: CEEMDAN algorithm
Input: original sequence signal Output: a series of *IMF* components
Begin While the original signal $x(t)$ has not been decomposed Add k noises to the original signal: $X_i(t) = x(t) + yw_i(t)$, $i = 1, 2, ..., k$ $X_i(t)$ is decomposed by EMD, and a series of *IMF* components obtained by EMD decomposition takes the mean value as an *IMF* obtained by CEEMDAN decomposition: $IMF_q(t) = \frac{1}{k}\sum_{i=1}^{k} IMF_q^i(t)$, $r_q(t) = x(t) - IMF_q(t)$ Repeat the above steps based on $r_q(t)$ End

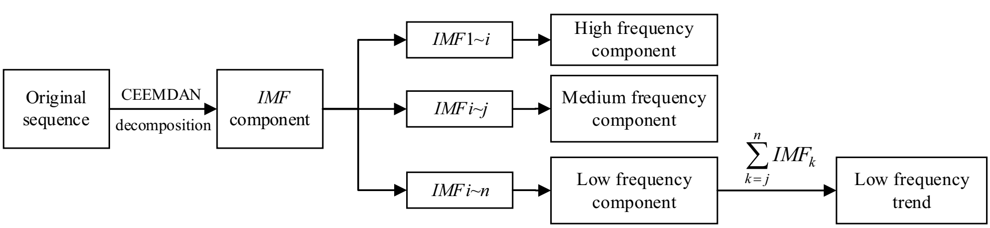

Figure 3. CEEMDAN decomposition experimental design

a series of *IMF* components, and the obtained low-frequency components are linearly combined to obtain the low-frequency trend sequence, which is then used as the input data for BP neural network prediction.

Then, the application of different signal decomposition methods in the CSI 500 stock index futures data is verified by experiments as follows:

(1) EMD Decomposition

The time series data of CSI 500 stock index futures closing price is used as the input signal, and EMD decomposition is performed.

As shown in Figure 5, EMD decomposes the CSI 500 stock index futures closing price series into 6 *IMF* components and 1 residual term. The EMD decomposition does not require artificial selection of wavelet bases, and can decompose the original series signal according to high and low frequencies. The *IMF* components decrease sequentially, but there are some mode mixing problems.

Figure 4. CSI 500 stock index futures closing price series

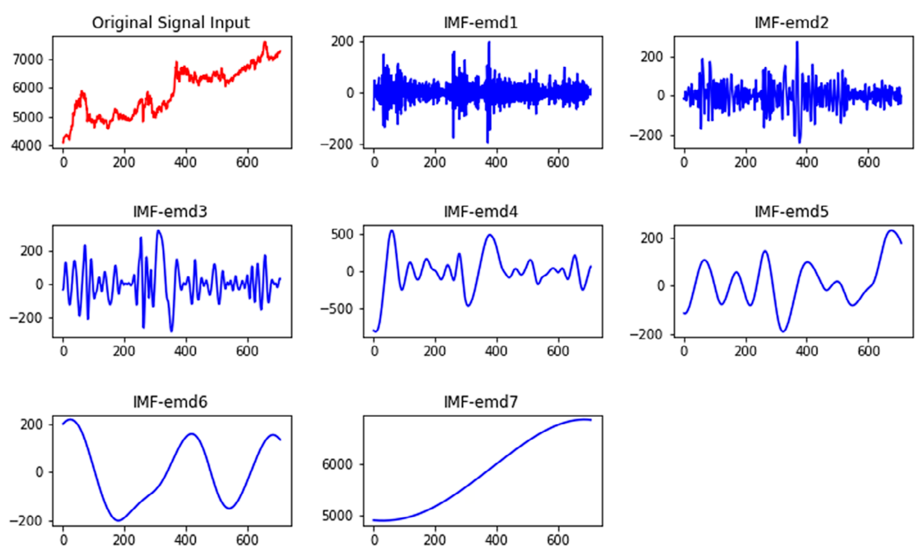

Figure 5. EMD decomposition simulation experimental results

(2) EEMD decomposition

Taking the time series data of the CSI 500 stock index futures closing price as the input signal, the EEMD decomposition is performed. White noise is introduced to solve the mode mixing problem that may occur with the EMD decomposition.

The above Figure 6 shows the *IMF* variables decomposed by using the EEMD method, the original signal is decomposed into 7 *IMF* components and 1 residual term. EEMD has good self-adaptivity and can solve the modal mixing problem of EMD decomposition. The method incorporates white noise, which is canceled out after several integrations so that the integration result can be taken as the final result. However, because the additional noise of the whole process belongs to the mixture, the test results are noisy, and the introduction of white noise may produce a certain deviation to the *IMF* components.

(3) CEEMDAN decomposition

The CSI 500 stock index futures closing price time series data is taken as the input signal and subjected to CEEMDAN decomposition. CEEMDAN decomposition deals with the noise interference problem generated by EEMD decomposition by adding a set of noise signals to the original series signal.

As shown above Figure 7, the CEEMDAN method decomposes the CSI 500 series signal to obtain 6 *IMF* components and 1 residual term. The technique adds a set of white noise signals to the original signal to resolve the bias arising from the introduction of noise in EEMD. Therefore, for non-smooth signals like stock index futures, it is more advantageous to use CEEMDAN for decomposition to extract the effective low and high frequency information fluctuations among them. For the seven *IMF* components obtained by CEEMDAN decomposition, the signal frequencies are reduced in order. Let the first two *IMF* components be high-frequency components, IMF_3 and IMF_4 be medium-frequency components, IMF_5 and IMF_6 be low-frequency components, and the last component IMF_7 is the residual term of CEEMDAN decomposition, which does not contain vibration modes and represents the trend of the original signal. In this paper, the components IMF_5, IMF_6 and IMF_7 are summed to obtain the combination of low-frequency components, which are incorporated into the forecasting model as trend signal variables.

Many existing prediction methods based on signal decomposition technology use machine learning algorithms to predict each signal component, and then add the predicted values of each component to obtain the final prediction result (Cao et al., 2019; Ji et al., 2022). However, building a prediction model for each component is complex and computationally intensive. The low-frequency components obtained by signal decomposition can be used as the long-term trend of the original signal (Han et al., 2019). At the same time, 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*. The low-frequency *IMF* usually consists of large values representing the macro trend of the sequence, which can better fit the long-term trend of market changes (Zhu et al., 2015). Therefore, the long-term trend variable composed of the sum of low-frequency components can play a greater role as the input variable of the prediction model.

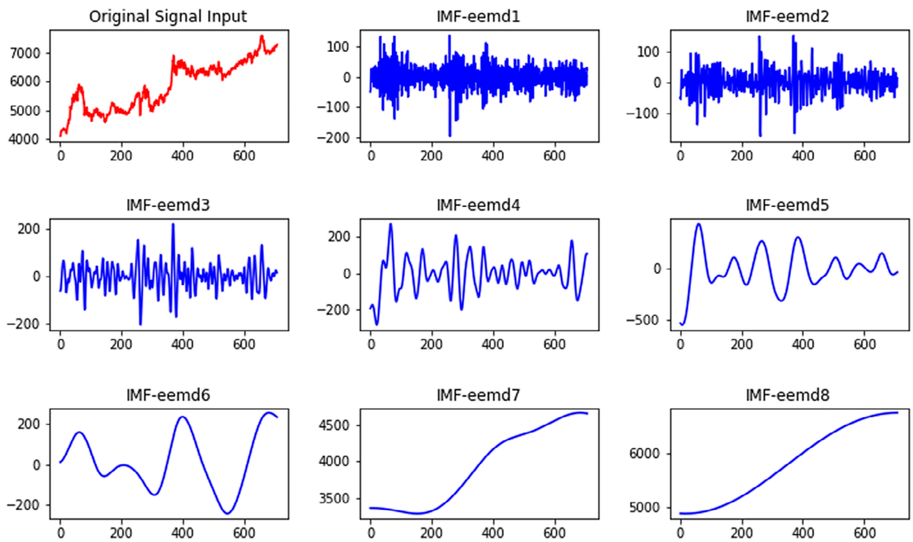

Figure 6. Experimental results of EEMD decomposition simulation

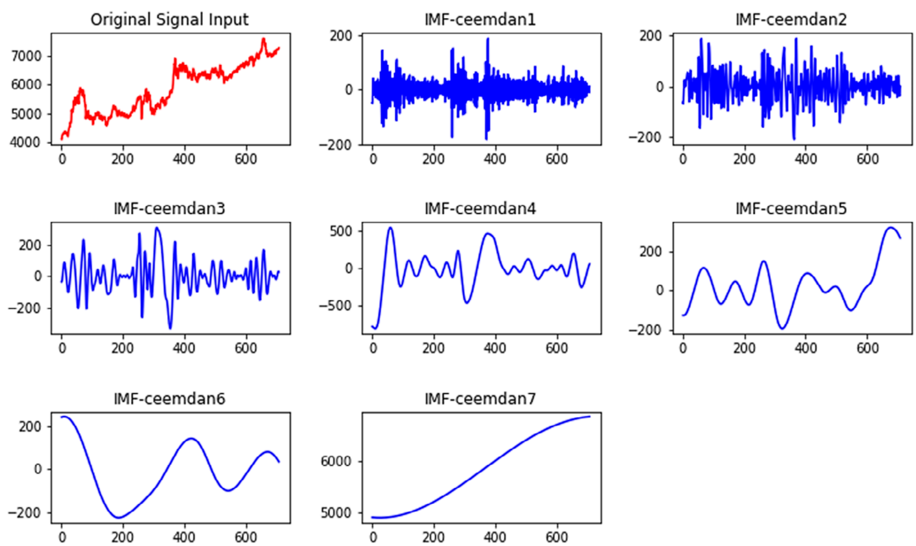

Figure 7. Experimental results of CEEMDAN decomposition simulation

3. Quantitative analysis based on CEEMDAN-BP neural network

In this section, the BP neural network algorithm is applied to financial time series prediction, and the low-frequency trend sequence obtained by CEEMDAN signal decomposition technology is combined with the investor attention index to improve the prediction accuracy of the model. Then, compare the prediction effect of CEEMDAN-BP neural network with EMD-BP and EEMD-BP models, and the impact of investors' attention on the prediction accuracy of the model is also compared.

3.1. Construction of CEEMDAN-BP neural network model

We first design the CEEMDAN-BP neural network, and then introduce four measurement indexes to evaluate the experimental effect.

3.1.1. Training of BP neural network

The structure of BP neural network consists of input layer, hidden layer and output layer. Given a training set $D=\{(x_i, y_i) \mid i=1,2,\dots,n\}$, where $x_i \in R^d$, $y_i \in R^l$ denote that the input sample consists of d attributes and the output value is an l-dimensional variable, a simple three-layer BP neural network structure is shown in Figure 8. In the model d, N, l respectively represent the number of neurons in input layer, hidden layer and output layer; w_{ih} represents the connection weight between the input layer and the hidden layer, and w_{hk} represents the connection weight between the hidden layer and the output layer; i represents the i-th neuron node of the input layer, h represents the h-th neuron node of the hidden layer, and k represents the k-th neuron node of the output layer.

Before the BP neural network is used for prediction, the network needs to be trained first, and the network has associative memory and prediction ability through training. Taking the three-layer neural network training in Figure 8 as an example, the detailed process of each step is as follows:

Step 1: Neural network parameter initialization: According to the input and output sequence $D=\{(x_i, y_i) \mid i=1,2,\dots,n\}$ to determine the number of input layer nodes (d), the number of nodes in the hidden layer (N), and the number of nodes in the output layer (l); Then we initialize the connection weight (w) between the neurons of each layer connection.

Step 2: Compute the output of the hidden layer. As shown in Figure 8, based on the input sequence $\{x_1, x_2, \dots, x_d\}$, the connection weight w_{ih} between input layer and hidden layer, and the hidden layer threshold a_h, the hidden layer output H is calculated:

$$H_h = f\left(\sum_{i=1}^{d} w_{ih} x_i - a_h\right), \; h=1,2,\dots,N, \tag{3}$$

where N is the number of hidden layer nodes, f is the activation function of the hidden

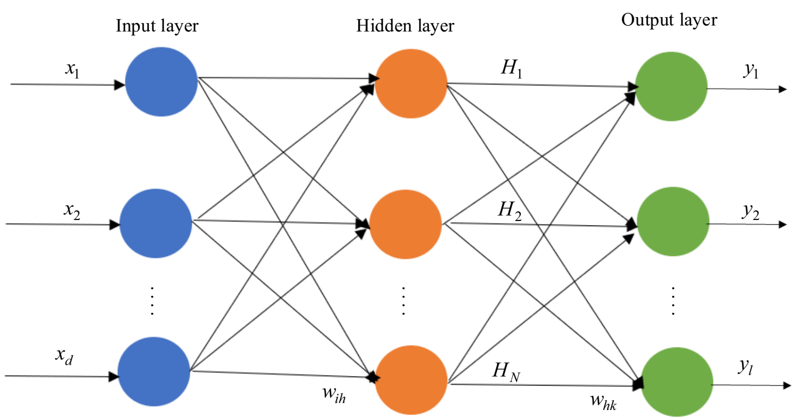

Figure 8. BP neural network model

layer, and *Reul* (Rectified Linear Unit) is selected as the activation function. For *Reul* activation function, the output is 0 when the input signal is less than 0, and the result is equal to the input when the input signal is greater than 0. *Reul* function is partially linear, and the obtained stochastic gradient descent method converges faster than both *Sigmoid* and *Tanh* activation function.

Step 3: The output layer calculates the output. Based on the hidden layer output H, the predicted output y_k is obtained by BP neural network by connecting the weights w_{hk} and the threshold b_k,

$$y_k = \sum\nolimits_{i=1}^{N} H_h w_{hk} - b_k,\ k = 1,2,\ldots,l. \tag{4}$$

Step 4: Calculate the error. The network prediction error E_k is calculated according to the network prediction output y_k and the expected output o_k. The error formula is as follows:

$$E_k = \frac{1}{2}\sum_{j=1}^{l}(y_k - o_k)^2,\ k = 1,2,\ldots,l. \tag{5}$$

Step 5: Update the weights: w_{ih} and w_{hk}. The weight update formula is as follows:

$$w_{ih} = w_{ih} + \eta H_h(1 - H_h x_i)\sum\nolimits_{k=1}^{l} w_{hk}E_k,\ i = 1,2,\ldots,d,\ h = 1,2,\ldots,N; \tag{6}$$

$$w_{hk} = w_{hk} + \eta H_h E_k,\ h = 1,2,\ldots,M,\ k = 1,2,\ldots,l, \tag{7}$$

where η is the learning efficiency, which is set to 0.01.

Step 6: Update the thresholds. The thresholds a_h and b_k of the network nodes are updated by the training error.

$$a_h = a_h + \eta H_h(1 - H_h)\sum\nolimits_{k=1}^{l} w_{hk}\mathrm{E}_k,\ h = 1,2,\ldots,N; \tag{8}$$

$$b_k = b_k + E_k,\ k = 1,2,\ldots,l. \tag{9}$$

Step 7: Determine whether the iteration is completed according to the threshold of training error or the maximum number of iterations, and if not, return to Step 2.

3.1.2. CEEMDAN-BP neural network design

Then, the *IMF* component obtained by CEEMDAN decomposition technology is incorporated into the BP neural network model to build the CEEMDAN-BP neural network model. For this model, the initial parameters are set as follows: the maximum number of iterations is set to 5000, the learning rate is 0.001, the activation function defaults to *Reul*, and the random gradient descent method selects *Adam*. These parameters will remain unchanged. This paper does not carry out special parameter optimization for these parameters, but uses grid search method to optimize the number of hidden layers and the number of neurons in each layer. Note that the initial input data needs to be normalized to avoid the influence of variable dimensions, which is also required for solving. Specifically, the algorithmic steps of CEEMDAN-BP neural network training and optimization search in this paper are shown in the following Table 4.

Table 4. CEEMDAN-BP neural network algorithm

Algorithm 3: CEEMDAN-BP neural network algorithm
Input: CSI 500 market transaction data, attention index, and trend variables decomposed by CEEMDAN (all standardized). Output: Predicted closing price
Begin Step 1: Pre-set the parameters that need to be searched for and the searched grid set by the grid search method. Step 2: Parameter initialization: maximum number of iterations Max_*n*, learning efficiency, number of neural network layers, initial weights, threshold value, learning accuracy Step 3: Calculate the input and output values of each layer Step 4: Calculate the output layer error E_k Step 5: While $E_k > \varepsilon$ And number of iterations < Max_*n* Correction of weights and thresholds Calculate the input and output values of each layer Calculate the output layer error E_k Step 6: Output predicted closing price Step 7: Update the goodness of fit on the test set if higher integrity of fit is obtained on the test set Step 8: Update the number of layers of the neural network and the number of neurons in each layer, return to Step2, and loop several times until the set grid search method parameter optimization grid is traversed. End

This algorithm takes the maximization of the goodness of fit on the test set as the training goal, and carries out the parameter optimization process. After all parameters are determined, the model is formally trained, and the training process is shown in Table 4. We calculate the input and output values of each layer of the neural network, then calculate the error of the output layer, and correct the weight and threshold according to the error. Multiple iterations are performed until the error of the output layer is less than the threshold. Finally, the CSI 500 closing price can be predicted based on the trained model. It should be noted that, compared with the BP neural network model, the CEEMDAN-BP neural network model here considers the non-stationary and strong volatility of financial time series, and combines the low-frequency *IMF* components obtained by the CEEMDAN decomposition technology. The model fully utilizes the advantages of different algorithms and avoids the problem of certain defects in a single algorithm.

3.1.3. Evaluation indicators

In order to evaluate the prediction ability of the proposed machine learning model, this paper selects four evaluation indicators as follows:

(1) MSE: mean square error, the average of the squares of the difference between each predicted value and the actual value. The larger the value is, the more the predicted value deviates from the actual value, the worse the prediction ability of the model is. The specific formula is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^* - y_i\right)^2, \tag{10}$$

where y_i^* represents the predicted value, y_i is the actual value.

(2) MAE: mean absolute error, similar to MSE, is also used to evaluate the difference between each predicted value and the actual value of the price. The larger the value, the greater the prediction error. The specific formula is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i^* - y_i |. \tag{11}$$

(3) MAPE: average absolute percentage error. Compared with the first two, it eliminates the influence of dimension. The larger the value, the greater the prediction error. The specific calculation formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}| \frac{y_i^* - y_i}{y_i} |. \tag{12}$$

(4) R^2: goodness of fit. The larger the value, the closer the predicted value is to the actual value, and the higher the prediction ability of the model. In machine learning, a decision coefficient of more than 40% indicates a good prediction ability of the model. The specific calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^* - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2}, \tag{13}$$

where $\bar{y}_i$ represents the average of the actual values.

3.2. Model training of CEEMDAN-BP neural network used in stock index futures

This section will train CEEMDAN-BP neural network model to obtain the optimal parameters. In addition, it will test whether the trained CEEMDAN-BP neural network model can adapt to the stock price fluctuation after the epidemic, and further establish a prediction model that can adapt to the impact of the epidemic.

3.2.1. CEEMDAN-BP neural network model training

This paper attempts to study whether major events will affect the prediction effect of the established neural network model, downloading the main contract information of Shanghai and Shenzhen 500 stock index futures from January 2, 2019 to December 1, 2021 from the China Financial Futures Exchange, and taking January 20, 2020 as the turning point. Because since Dr. Nanshan Zhong announced the existence of human-to-human transmission on January 20, 2020, the epidemic has attracted widespread attention.

A total of 244 sets of samples from January 2, 2019 to December 31, 2019 are used as the training set, and the sample data from January 2, 2020 to January 20, 2020 are used as the test set, with the objective function of maximizing the R^2 on the test set. To train the

CEEMDAN-BP neural network model, today's open, high price, low price, volume, turnover, position, position change, today's close, today's settlement, previous settlement, up/down 1, up/down 2, attention index, and trend signal of day T are used as independent variables, and today's close of day $T + 1$ is used as dependent variable to train the model. The specific process is shown in the Figure 9.

First, we consider a single-layer neural network with an upper limit of 100 neurons, and the following training results can be obtained by the grid-seeking method, as shown in Table 5.

It can be seen from Table 5 that the fit of the single-layer neural network is not good, and the maximum R^2 is only 50.4133%. Therefore, the next step is to consider a two-layer neural network, with the upper limit of the number of neurons in each layer is set to 100, and the following results are obtained, as shown in Table 6.

Table 5. Single layer neural network

Number of neurons in a single layer	Goodness of fit R^2
1	–223.068258
2	–184.587728
3	–68.593608
4	–26.659548
6	–12.130418
11	–9.765761
20	–5.323241
25	–3.820805
27	–1.128698
32	–0.883848
43	–0.511080
46	–0.294646
57	0.504133

Table 6. Two-layer neural network

Number of neurons in the first layer	Number of neurons in the second layer	Goodness of fit R^2
1	1	–209.822945
1	2	–34.196272
1	6	–4.858401
1	14	–0.736318
2	1	–0.148540
3	87	0.239097
6	67	0.444762
7	72	0.589397
13	59	0.651275
24	27	0.729542
32	63	0.803933
32	87	0.804151
78	86	0.823646

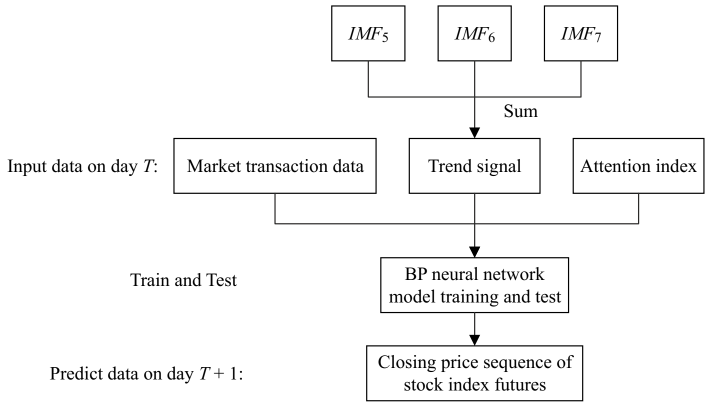

Figure 9. BP Neural Network Prediction Process

It can be seen from Table 6 that compared with the single-layer neural network, the prediction effect of two-layer neural network is better, but the highest goodness of fit is only 82.3646%. Therefore, in order to improve the prediction accuracy, we reconsider the three-layer neural network, set the upper limit of the number of neurons in the first and second hidden layers to 100, and the upper limit of the third layer to 10, and obtain the following training results, as shown in Table 7.

Table 7. Three-layer neural network

Number of neurons in the first layer	Number of neurons in the second layer	Number of neurons in the third layer	Goodness of fit R^2
1	1	1	–266.003812
1	1	2	–203.586717
1	1	3	–203.311155
1	1	4	–48.188658
1	1	7	–27.061298
1	2	2	–0.345499
1	14	1	–0.115468
1	21	1	–0.015476
1	41	2	–0.000030
3	86	8	0.038443
3	96	10	0.089340
6	6	4	0.465598
6	54	10	0.500369
6	88	6	0.660002
7	52	8	0.711148
13	69	7	0.723578
18	76	3	0.774343
24	72	10	0.799029
35	89	8	0.837472
37	79	7	0.854654

It can be seen from Table 7 that, when the number of neurons in the three layers is 37, 79 and 7 respectively, the maximum R^2 of the three-layer neural network on the test set is optimal, which is 85.46%. At this time, the CEEMDAN-BP neural network has the best fitting effect on the test set, and the optimal R^2, MSE, MAE and MAPE of the model are 85.64%, 687.33, 21.71 and 0.015, respectively.

3.2.2. CEEMDAN-BP neural network model validity test

Now, the neural network model before the epidemic has been established, but the training and parameter optimization of the model are based on the data before the epidemic (before January 20, 2020). Next, the 487 sets of samples from January 2, 2019 to December 31, 2020 are used as the training set, and the 118 sets of data from January 4, 2021 to June 30, 2021 are used as the test set. To test whether the data fluctuations before and after the epidemic have significant changes, and whether the optimal parameters found can adapt to the price fluctuations after the epidemic.

As shown in Figure 10, the optimal parameters found by training on the pre-epidemic are not adapted to the data changes after the epidemic, and the R^2 is only 26.67%. The MSE and MAE on the test set from January 4, 2021 to June 30, 2021 are 20951.6 and 118.6 respectively, indicating that the error is large. Therefore, it can be judged that the model trained before the epidemic and the optimal parameters found cannot adapt to the price fluctuations after the epidemic. We need to retrain and incorporate the post-epidemic data into the training model and parameter optimization to obtain higher prediction accuracy.

3.2.3. CEEMDAN-BP neural network model optimization

A total of 708 sets of samples from 2019/1/2 to 2021/12/1 are now divided into three parts: the first part, a total of 487 sets of samples from January 2, 2019 to December 31, 2020, is used as the training set, and this part needs to incorporate the post-epidemic data to make the model adapt to the post-epidemic data transformation; the second part, a total of 118 sets of samples from January 4, 2021 to June 30, 2021, is used as the test set. In the third part, 103 samples from July 1, 2021 to December 1, 2021 are used as the backtest set, and the previously established neural network model is used as the basis for predicting the stock price in the period from July 1, 2021 to December 1, 2021. The optimization search process is similar to the previous one and will not be repeated here. The following Figure 11 shows the optimization of single-layer neural network parameters, where the horizontal coordinate is the number of neurons and the vertical coordinate is the goodness-of-fit, which shows that the R^2 of the single-layer neural network is not very good, and is mostly in negative values.

The optimal number of neurons cannot be obviously derived due to the influence of the magnitude, and it also looks very small and close to 0 when the R^2 is greater than 0. Figure 12 shows a two-layer neural network parameter finding graph, where the x-axis indicates the number of neurons in the first layer, the y-axis indicates the number of neurons in the second layer, and the z-axis indicates the goodness-of-fit. Again, due to the effect of the magnitude, the goodness of fit greater than 0 is not apparent in the image, but it can still be seen that the image has many small protrusions that appear to be relatively close to 1. It is obvious that the fit of the two-layer neural network is better than that of the single-layer neural network.

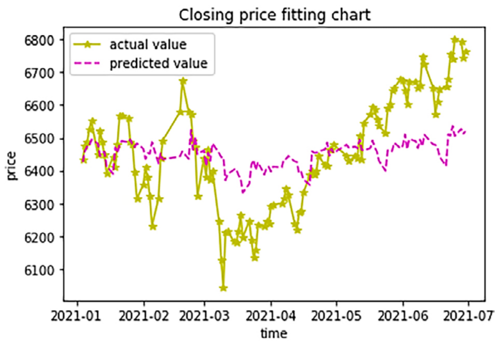

Figure 10. Fitted plots after the outbreak

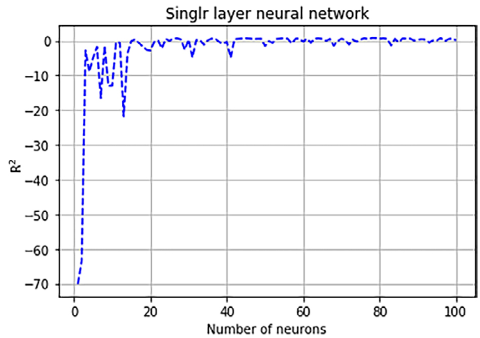

Figure 11. Single layer neural network parameter seeking

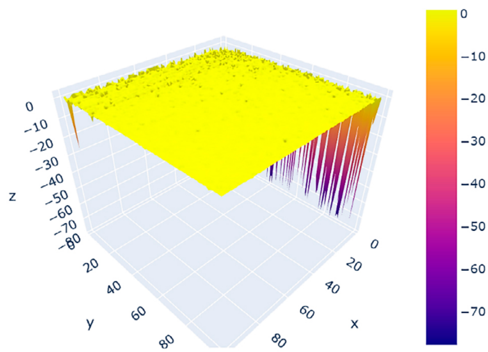

Figure 12. Parameter optimization of double layer neural network

Similarly, the parameter optimization results of the three-layer neural network can be obtained. The optimal parameter results of the single-layer, two-layer, and three-layer neural networks and the corresponding evaluation index values on the test set are shown in Table 8.

The maximum R^2 achieved by a three-layer neural network is 88.92%, so a three-layer neural network is used, with each layer containing 84, 95, and 4 neurons, respectively. the following Figure 13 shows the fit of this neural network on the test set.

Then, the trained CEEMDAN-BP neural network model is used to predict the closing price sequence from 2021/7/1 to 2021/12/1 of the backtest set. The prediction results are shown in the Figure 14.

It can be seen from the above Figure 14 that the CEEMDAN-BP neural network prediction model established in this paper has a good prediction effect on the back test set. MSE, MAE, MAPE and R^2 are 5295.52, 56.92, 0.008 and 88.38% respectively. MSE and MAE are a little high, but MAPE is very small, approaching 0, and R^2 is about 88.38%, indicating that the predicted value is close to the actual value. Considering the high closing price of stock index futures and large fluctuations, it is a normal phenomenon that MSE and MAE values are large. If we look only at MAPE indicators, we think CEEMDA-BP model has good forecasting ability. And R^2 is about 88.38%, indicating that the model can explain the change of 88.38%, so the model established in this paper has a high prediction accuracy.

Table 8. Parameter optimization

Network Structure	Number of neurons	R^2	MSE	MAE	MAPE
single-layer	(78)	86.50%	3857.3	49.1	0.008
two-layer	(61, 82)	88.16%	3381.5	45.7	0.007
three-layer	(84, 95,4)	88.92%	3166.4	43.4	0.0068

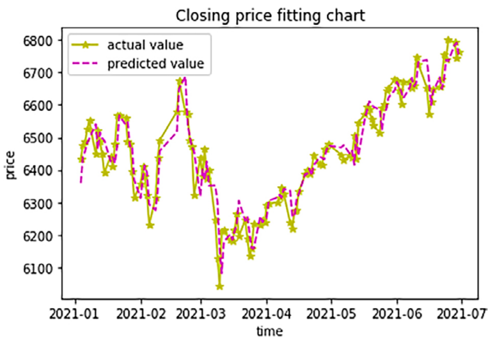

Figure 13. The fit of neural network on the test set

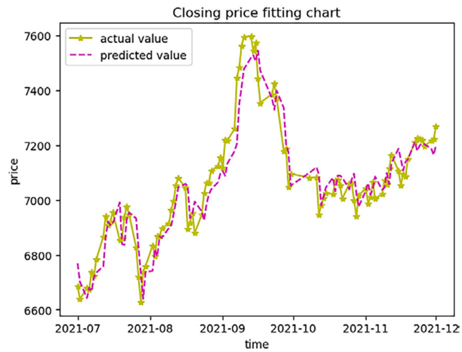

Figure 14. The fit of neural network on the backtest set

3.3. Comparison of model accuracy

This section will verify that the prediction effect of CEEMDAN-BP neural network proposed in this paper is better than that of EMD-BP and EEMD-BP models. And compare the impact of investors' attention on the prediction accuracy of the model.

3.3.1. Comparison of three signal decomposition methods

The training and optimization process of EMD-BP and EEMD-BP neural network models is similar to the prediction process of CEEMDAN-BP neural network in Section 3.2, so the detailed training process is omitted here. The following Figure 15 shows the prediction fitting diagram of EMD-BP and EEMD-BP neural network.

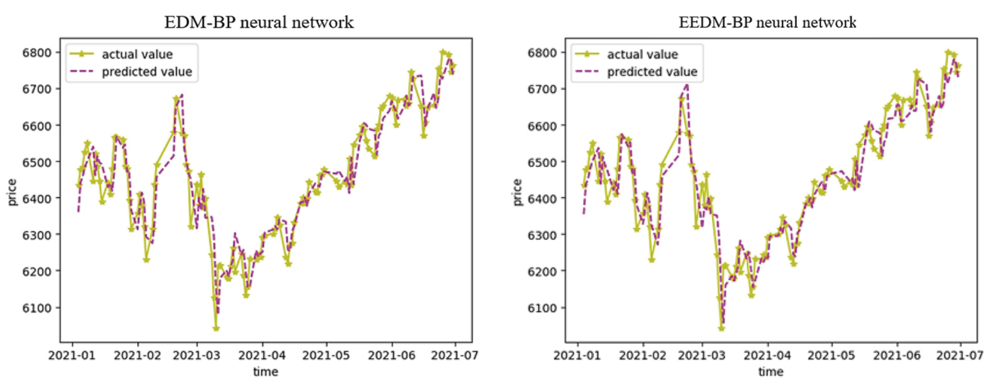

Figure 15. Fitting of EDM-BP and EEDM-BP neural network model in test set

The highest goodness of fit of EMD-BP neural network on the test set is 88.91, and the goodness of fit of EEMD-BP neural network model on the test set is 88.42%. Similarly, it can be obtained that the goodness of fit of EMD-BP and EEMD-BP neural network models on the backtesting set are 87.97% and 86.45% respectively.

In order to more intuitively compare the prediction effects of CEEMDAN-BP, EEMD-BP and EMD-BP models, the following Table 9 shows the prediction effects of the three models on the test set and the back test set.

Table 9. Comparison of prediction results of three models

Model	Test Set				Backtest set			
	R^2	MSE	MAE	MAPE	R^2	MSE	MAE	MAPE
CEEMDAN-BP	88.91%	3166.44	43.40	0.007	88.37%	5295.52	56.92	0.008
EEMD-BP	88.42%	3308.37	44.58	0.007	86.45%	6172.47	64.43	0.009
EMD-BP	88.91%	3167.29	43.67	0.007	87.97%	5479.45	58.25	0.08

It can be seen from the above Table 9 that the three models have good prediction effects on test sets and back test sets, and have strong generalization ability. In general, CEEMDAN-BP neural network has the best prediction effect, and the MSE and MAE of this model in test set and back test set are the smallest of the three models, and R^2 is the highest. This indicates that the predicted value of CEEMDAN-BP model is the least deviated from the actual value. Therefore, CEEMDAN is selected as the signal decomposition method in this paper, and the CEEMDAN-BP neural network model constructed has high prediction accuracy and strong generalization ability.

3.3.2. Comparison of methods without consideration of attention

In this section, we compare the predictive effects of models that consider investor attention and models that do not, as shown in Figure 16.

It can be clearly found that the prediction effect of considering investor attention is better than that of not considering investor attention. The following table shows the prediction effects of these two methods on the test set and the back test set respectively.

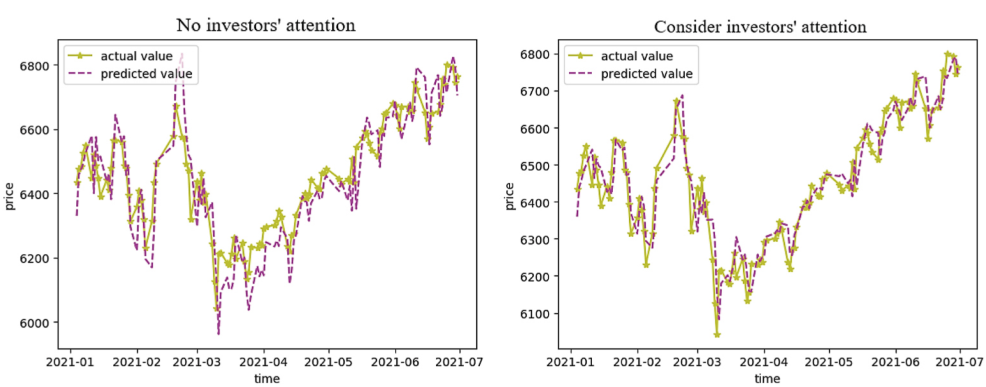

Figure 16. Fit on test set

Table 10. Comparison of prediction results of two methods

	Regardless of the investor's attention				Consider the investor's attention			
	R^2	MSE	MAE	MAPE	R^2	MSE	MAE	MAPE
Test set	74.63%	7247.47	69.12	0.010	88.91%	3166.44	43.40	0.007
Backtest set	74.08%	11809.94	81.21	0.011	88.37%	5295.52	56.92	0.008

It can be seen from Table 10 above that the prediction accuracy of CEEMDAN-BP neural network model considering investor concern is indeed higher than that of the model not considering investor attention. Therefore, in establishing CEEMDAN-BP model, it is necessary to consider the investor's attention factor, which will effectively improve the accuracy of model prediction.

4. Design and implementation of trading decision-making strategies

Finally, we should note that although financial time series forecasting is a very attractive target, many studies have not built trading strategies based on their forecasting models. In fact, it is very important when building a trading strategy to prove that the proposed forecasting method can be profitable in the actual market, after all forecasting does not equal profit. Therefore, this section is based on the CSI 500 stock index futures data and uses the neural network model established in Section 4.2 to predict the 103 days closing price of the CSI 500 stock index futures from July 1 to December 1, 2021. Then according to the Dual Thrust strategy, a quantitative investment strategy based on the investor attention index and CEEMDAN-BP neural network model is constructed.

4.1. Quantitative investment strategy modeling

The trend following strategy is an investment strategy developed based on the characteristics of price trends. We choose the Dual Thrust strategy, which is an intraday trading strategy, and applied it to our daily quantitative investment decisions. This strategy defines a price breakout interval, and regards the breakout of the range as a sign of "changes". This breakout interval will frame most of the market fluctuations. The breakout interval is not constant, but gradually narrows or expands as the market fluctuates. and enlarges when the market moves up or down. This breakout interval is constructed using the strategy indicators shown in Figure 17.

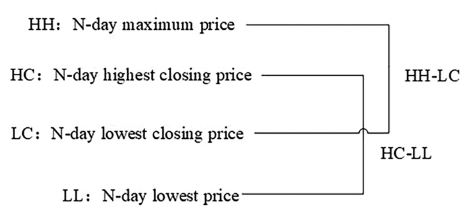

Figure 17. The strategy indicators

The breakout interval is [*Selline*, *Buyline*], which is calculated based on the highest and lowest prices and closing prices in the last N days, plus a certain adjustment ratio. The specific formulas are as follows:

$$\begin{cases} Range = Max(HH - LC, HC - LL) \\ Buyline = Open + K_1 \times Range \\ Sellline = Open - K_2 \times Range \end{cases}, \tag{14}$$

where *HH* is the highest price in n days, *HC* is the highest closing price in *n* days, *LC* is the lowest closing price in *n* days, *LL* is the lowest price in *n* days and Open is the opening price. *Range* is the breakout interval, *Buyline* is the upper rail line, when the price breaks through the line, the buy strategy is executed, and *Sellline* is the lower rail line, when the price breaks below the line to execute the sell strategy. Specifically:

(1) When the closing price of $T + 1$ is predicted to be higher than the opening price of T-day plus K_1 times the *Range*, it indicates that the market may have an upward trend, and a buy operation is performed, marked as 1;

(2) When the closing price of $T + 1$ is predicted to be lower than the opening price on T-day minus K_2 times the *Range*, it indicates that the market may have a downward trend and a sell operation is performed, marked as –1;

(3) When no action is taken and the stock index futures are held, marked as 0.

The parameters K_1 and K_2 are mostly used to adjust the triggering difficulty of long and short positions. When $K_1 > K_2$, short positions are relatively easier to trigger, and when $K_1 < K_2$, long positions are relatively easier to trigger. In summary, the quantitative decision-making model of this paper is shown in Figure 18.

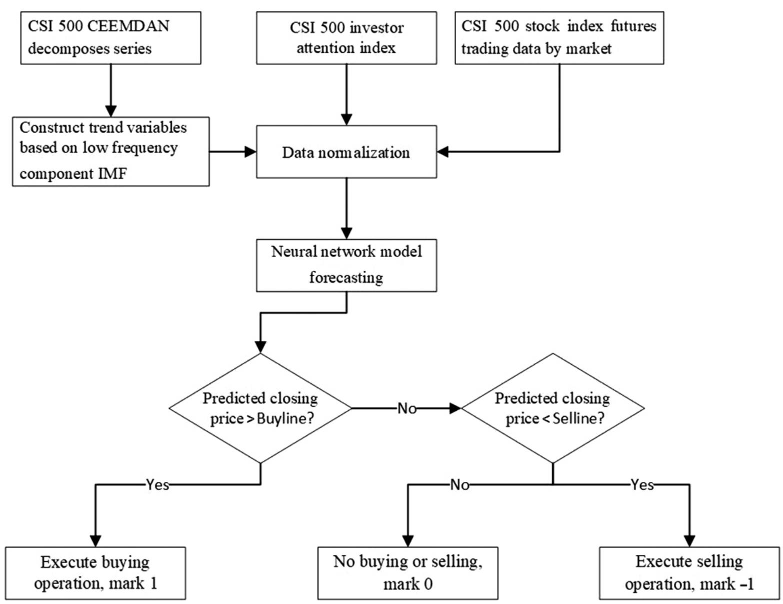

Figure 18. Design of Attention-based Dual Thrust strategy

The basic process of the Attention-based Dual Thrust strategy constructed in this paper based on the CEEMDAN-BP neural network is: firstly, the CEEMDAN decomposition technique is used to decompose the CSI 500 stock index futures series to obtain the low, medium and high frequency components, and the trend variables are constructed based on the low frequency component *IMFs*. The obtained trend variables are then normalized with the CSI 500 attention factor and market trading data and fed into a BP neural network model, which is formally used for closing price prediction after the merit search process in Section 3.2. The next step is to invest in the Dual Thrust strategy, which compares the forecasted closing price with the upper and lower lines. If the forecast closing price is above the upper line, a buy operation is performed, marked as 1. If the forecast closing price is less than the lower line, a sell operation is performed, marked as –1. If the value of the forecast closing price is between the upper and lower lines, no operation is performed, marked as 0. The specific algorithm steps are shown in Table 11.

Table 11. Quantitative Decision Model Algorithm

Algorithm 4: Quantitative Decision Modeling Algorithm
Input: CSI 500 trading data, attention index and trend variables by market Output: Quantitative investment strategy

Begin
Step 1: Normalize the input data separately
Step 2: Neural network model prediction
Step 3: Construct the upper and lower trajectory lines
Step 4: Use Dual Thrust strategy
While not all the predicted closing prices are judged.
if Predicted closing price > *Buyline*:
xecute a buy operation, marked it as 1, and included it in the quantitative investment strategy
else if Predicted closing price < *Sellline*:
Execute a sell operation, marked it as -1, and include it in the quantitative investment strategy
else:
Not buying or selling, marked as 0, included in the quantitative investment strategy
Step 5: Output quantitative investment strategy
End

4.2. Strategy implementation

This paper uses the neural network model developed in Section 3.2.3 to predict the closing price for a total of 103 days from July 1 to December 1, 2021. In order to illustrate the feasibility of the model established in this paper, the values of MSE, MAE, MAPE and R^2 on the back testing set are given in the following Table 12.

Table 12. Indicators' value on the backtest set

Indicators	MSE	MAE	MAPE	R^2
Value	5295.5	56.9	0.008	88.4%

As can be seen from the above table, the model established in this paper has a high prediction accuracy and good generalization ability. Therefore, the model established in this paper can accurately predict the change trend of the closing price of CSI 500 stock index futures. On this basis, the implementation of dual thrust strategy may have a high return on investment. Based on the predicted closing price, Dual Thrust strategy is adopted, and both K_1 and K_2 are set to 0.2 in this paper. In order to ensure that the empirical analysis results of the strategy are more realistic, this study selects the BackTrader quantitative trading framework based on Python for backtesting (Rodriguez, 2020), and sets the volume of each transaction is one hand, the initial capital is 1,000,000, the sliding point is 0.20%, and the stop loss line is 20% of the principal.

Then execute the strategy, and carry out the clearing operation on the last day of the backtesting period. The daily trading operation is shown in the Figure 19.

Figure 20 below shows the cumulative rate of return change during the backtesting period. The cumulative yield of the quantitative trading strategy proposed in this paper reached 88.4% during the backtesting period.

For more objective evaluation, five evaluation indicators are selected to evaluate the return and risks of the constructed Attention-based Dual Thrust investment decision model, as shown in Table 13.

Figure 19. Daily trading operation

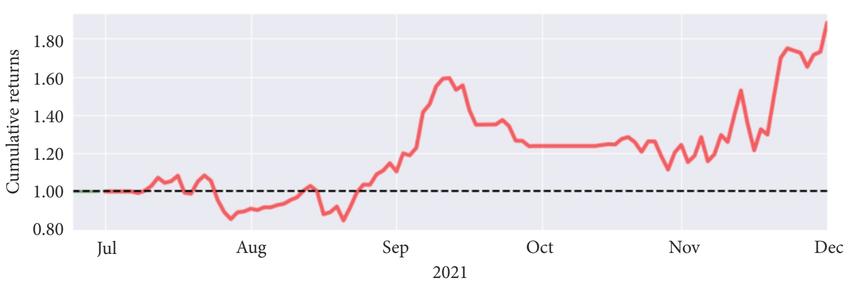

Figure 20. Cumulative rate of return

Table 13. Back testing revenue evaluation indicators

Evaluation indicator	Indicator value
Annualized rate of return	370.9%
Cumulative rate of return	88.4%
Sharpe ratio	2.26
Calmar ratio	12.34
Maximum drawdown	30.1%

According to the backtesting of the CSI 500 stock index futures quantitative investment decision-making model established in this section, the Sharpe ratio is 2.26, the Calmar ratio is 12.34, the maximum drawdown is 30.1%, the cumulative return is 88.4%, and the cumulative return is 883831. It is found that the proposed Dual Thrust strategy based on the investor attention index and CEEMDAN-BP neural network model makes use of the advantages of different algorithms and can achieve higher investment return with lower risk. This model avoids some defects of a single algorithm and can make corresponding adjustments according to the changes of market sentiment, so as to obtain more excess returns under low-risk control.

4.3. The optimized Dual Thrust strategy based on investor attention

In this section, the Dual Thrust strategy proposed in Section 4.1 is optimized. In Section 4.2, K_1 and K_2 used to determine the breakout interval are set to the same value of 0.2. In this section, K_1 and K_2 will be determined by the function of investor attention to avoid the subjectivity of manual selection of K_1 and K_2 values.

The attention-driven purchase hypothesis suggests that an increase in attention to a stock means that more investors pay attention to the stock, and investors are limited by their own cognitive resources to choose and invest from the stocks they pay attention to, so an increase in attention also means an increase in potential investment demand for the stock, which in turn boosts the trading volume and price of the stock (Barber & Odean, 2008). Therefore, this paper believes that when investors pay more attention to the stock index futures prices have a rising trend, it is best to execute the buying operation. At this time, K_1 can be reduced, making long positions trigger more easily. Similarly, when investors pay less attention, the price of stock index futures tends to decline. It is better to carry out the selling operation. At this time, K_2 can be reduced, which makes the short positions more likely to trigger.

In order to better build K_1 and K_2, first of all, we normalize investors' attention.

$$Att_t' = \frac{Att_t - \min(Att)}{\max(Att) - \min(Att)} \in [0,1], \tag{15}$$

where, Att_t' represents the investor's attention after normalization on day t. Then build K_1 and K_2 as follows:

$$K_1 = (1 - Att_t') * 0.1, \quad K_2 = 0.3 * Att_t'. \tag{16}$$

Among them, the formula of K_1 multiplies the coefficient of 0.1, which is because the volume of each transaction is not high in this paper, and one transaction is board lot. Therefore, the coefficient can be set lower to make it easier to buy and exchange high risks for high benefits. The formula of K_2 multiplies the coefficient of 0.3, which can avoid frequent selling operations to obtain more holding income. Aggressive investors can set the coefficient of K_1 lower and K_2 higher. On the contrary, conservative investors can set the coefficient of K_1 higher and K_2 lower. The optimized Attention-based Dual Thrust strategy established by K_1 and K_2 is determined by the above method, which solves the core problem of how to determine the breakout interval and effectively avoids the risk brought by subjective selection.

Then, perform the back test based on the optimized Dual Thrust strategy, and the following results can be obtained, as shown in Figure 21.

As shown in the first part of the above figure, the initial capital is 1,000,000. After quantitative investment, the total capital reaches 1,987,601, and the net income is 987,601. As shown in the second part of the figure above, there is only two negative trades. It should be noted that there are not only two transactions, because only when all the futures purchased

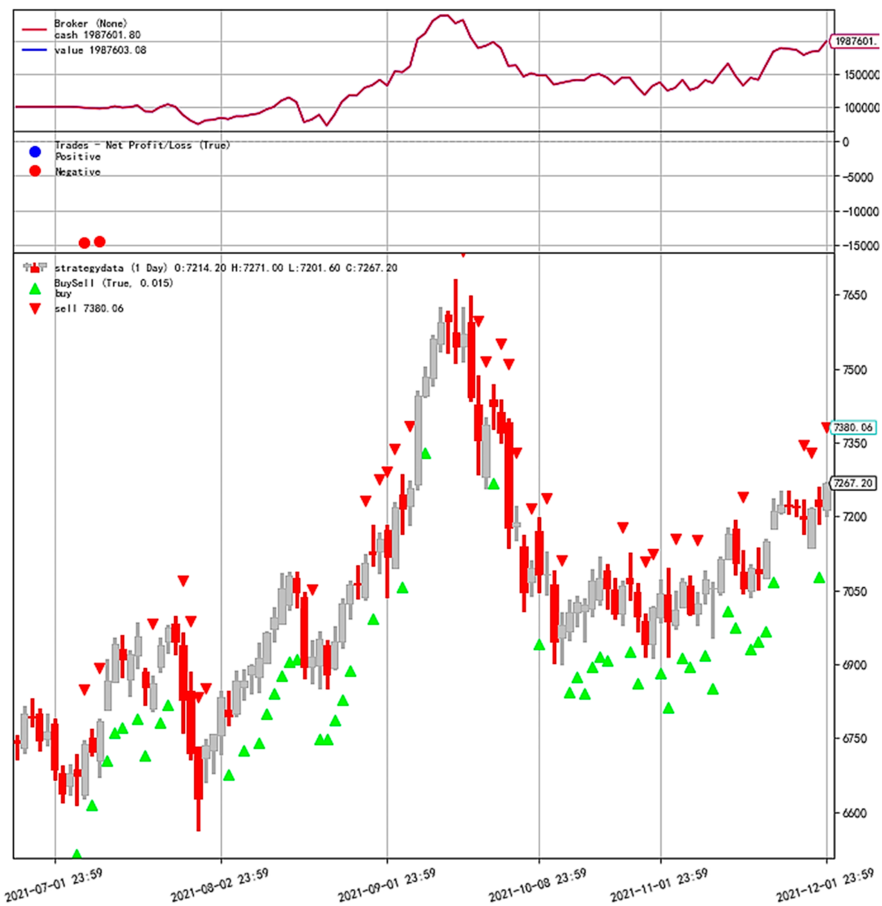

Figure 21. Backtesting results of optimization strategy

before are cleared, can it be counted as a complete transaction. However, when this paper executes the sell operation each time, it does not clear all the positions, so there are only two transactions in the figure. The third part of the figure above shows the time of buying and selling operations. The green triangle represents the buying operation and the red triangle represents the selling operation.

Similarly, we can get the cumulative rate of return change of the optimization strategy during the backtesting period, as shown in the Figure 22.

Then, the optimization strategy is compared with the strategy implementation results in Section 4.2, and the following Table 14 is obtained.

As shown in the below table, the annualized rate of return and cumulative rate of return of the optimized strategy are 436.9% and 98.8% respectively, which are higher than the original strategy. At the same time, the risk of the optimized strategy is also higher, with the maximum drawdown up to 50.5%. At the same time, the Sharp ratio and Calmar ratio are also lower than the original strategy. To sum up, in order to obtain higher investment returns, we can use investors' attention to determine the values of K_1 and K_2, which is practical and can improve the return on investment. In the optimized Attention-based Dual Thrust strategy, K_1 and K_2 change with the investor attention index, and the coefficients can also be set according to the different risk preferences of investors, effectively controlling risks and obtaining higher returns.

Figure 22. Cumulative rate of return of the optimization strategy

Table 14. Comparison table

Evaluation indicator	original strategy	optimization strategy
Annualized rate of return	370.9%	436.9%
Cumulative rate of return	88.4%	98.8%
Sharpe ratio	2.26	1.95
Calmar ratio	12.34	8.66
Maximum drawdown	30.1%	50.5%

4.4. Strategy comparison

In order to verify that the Attention-based Dual Thrust strategy proposed in this article is effective, this section will implement a blank strategy, and then compare it with the strategy results proposed in this paper.

The blank strategy buys one hand stock index futures on the first day of the backtesting period, clears positions on the last day, and does not carry out any buying or selling operations in the middle. The comparison results of the original Attention-based Dual Thrust strategy, the optimized Attention-based Dual Thrust strategy and the blank strategy are shown in the following Table 15.

Table 15. Blank strategy comparison table

Evaluation indicator	original strategy	optimization strategy	blank strategy
Annualized rate of return	370.9%	436.9%	30.6%
Cumulative rate of return	88.4%	98.8%	11.5%
Sharpe ratio	2.26	1.95	1.56
Calmar ratio	12.34	8.66	2.75
Maximum drawdown	30.1%	50.5%	11.1%

It can be seen from Table 15 that the quantitative investment strategy proposed in this paper has much higher returns than the blank strategy. Although the maximum drawdown of the blank strategy is the lowest, which means that the risk is the lowest, but the two comprehensive indicators for measuring risk and return, the Sharp Ratio and Calmar Ratio, are lower than the quantitative investment strategy proposed in this paper. To sum up, compared with the blank strategy, the two Attention-based Dual Thrust strategies constructed in this paper will obtain higher returns with lower risks, which has certain practical significances.

Conclusions

Traditional time series analysis is only suitable for dealing with stationary sequences. For non-stationary data, it is often necessary to transform it into stationary data through differential transformation or other transformations, and then analyze it. The CEEMDAN-BP neural network algorithm proposed in this paper can effectively deal with non-stationary sequences. Through experiments, it is found that by applying the investor attention index to the prediction of the closing price of stock index futures can make response adjustments according to changes in market sentiment and the occurrence of uncertain events, thereby improving the prediction accuracy of the model. The optimized quantitative decision-making model uses the investor attention index to realize the dynamic selection of the breakout interval in the dual-thrust strategy trading algorithm, effectively avoiding the possible risks caused by subjective selection, and meeting the different risk preferences of investors.

In addition, it also helps the government and financial regulators correctly grasp investors' attention behavior, improve the market regulatory framework, avoid investors' irrationality causing drastic market fluctuations, and further play the role of financial support and guidance to the real economy. Future research can be extended from the following aspects: (1) Continue to discuss the influence of high and intermediate frequency signals obtained by CEEMDAN technology on the prediction model. (2) Comprehensively consider leverage ratio, principal, position management, etc., in the quantitative investment strategy, and further optimize the model.

Acknowledgements

This work was supported by Guanghua Talent Project of Southwestern University of Finance and Economics, and the National Natural Science Foundation of China under Grant No. 72071135 and also supported by the Fundamental Research Funds for the Central Universities (No. JBK2304110).

References

Anadu, K., Kruttli, M., McCabe, P., & Osambela, E. (2020). The shift from active to passive investing: Risks to financial stability? *Financial Analysts Journal*, *76*(4), 23–39. https://doi.org/10.1080/0015198X.2020.1779498

Andrei, D., & Hasler, M. (2015). Investor attention and stock market volatility. *Review of Financial Studies*, *28*(1), 33–72. https://doi.org/10.1093/rfs/hhu059

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

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

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

Caldeira, J. F., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Brazilian Review of Finance*, *11*(1), 49–80. https://doi.org/10.12660/rbfin.v11n1.2013.4785

Camerer, C. F., & Loewenstein, G. (2004). Behavioral economics: Past, present, future. In *Advances in behavioral economics* (pp. 3–51). Princeton University Press. https://doi.org/10.1515/9781400829118-004

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, *519*, 127–139. https://doi.org/10.1016/j.physa.2018.11.061

Chan, E. P. (2021). *Quantitative trading: How to build your own algorithmic trading business*. John Wiley & Sons.

Chen, C., Zhang, P., Liu, Y., & Liu, J. (2020). Financial quantitative investment using convolutional neural network and deep learning technology. *Neurocomputing*, *390*, 384–390. https://doi.org/10.1016/j.neucom.2019.09.092

Covel, M. (2006). *Trend following: How great traders make millions in up or down markets.* FT Press.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Daniel, K., Hirshleifer, D., & Teoh, S. H. (2002). Investor psychology in capital markets: Evidence and policy implications. *Journal of Monetary Economics*, *49*(1), 139–209. https://doi.org/10.1016/S0304-3932(01)00091-5

Deng, C., Zhou, X., Peng, C., & Zhu, H. (2022). Going green: Insight from asymmetric risk spillover between investor attention and pro-environmental investment. *Finance Research Letters*, *47*, 102565. https://doi.org/10.1016/j.frl.2021.102565

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Fang, J., Gozgor, G., Lau, C. K. M., & Lu, Z. (2020). The impact of Baidu Index sentiment on the volatility of China's stock markets. *Finance Research Letters*, *32*, 101099. https://doi.org/10.1016/j.frl.2019.01.011

García, F., Guijarro, F., Oliver, J., & Tamošiūnienė, R. (2018). Hybrid fuzzy neural network to predict price direction in the German DAX-30 index. *Technological and Economic Development of Economy*, *24*(6), 2161–2178. https://doi.org/10.3846/tede.2018.6394

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

Han, L., Zhang, R., Wang, X., Bao, A., & Jing, H. (2019). Multi-step wind power forecast based on VMD-LSTM. *IET Renewable Power Generation*, *13*(10), 1690–1700. https://doi.org/10.1049/iet-rpg.2018.5781

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. https://doi.org/10.1016/j.eswa.2019.01.012

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1–3), 337–386. https://doi.org/10.1016/j.jacceco.2003.10.002

Huang, J., Li, Y., & Yao, H. (2022). Nonparametric mean-lower partial moment model and enhanced index investment. *Computers & Operations Research*, *144*, 105814. https://doi.org/10.1016/j.cor.2022.105814

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., & Blank, K. (2009). On instantaneous frequency. *Advances in Adaptive Data Analysis*, *01*(02), 177–229. https://doi.org/10.1142/S1793536909000096

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A century of evidence on trend-following investing. *The Journal of Portfolio Management*, *44*(1), 15–29. https://doi.org/10.3905/jpm.2017.44.1.015

Jiang, Y. (2022). Prediction model of the impact of innovation and entrepreneurship on China's digital economy based on neural network integration systems. *Neural Computing and Applications*, *34*(4), 2661–2675. https://doi.org/10.1007/s00521-021-05899-7

Ji, C., Zhang, C., Hua, L., Ma, H., Nazir, M. S., & Peng, T. (2022). A multi-scale evolutionary deep learning model based on CEEMDAN, improved whale optimization algorithm, regularized extreme learning machine and LSTM for AQI prediction. *Environmental Research*, *215*, 114228. https://doi.org/10.1016/j.envres.2022.114228

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218–226). Prentice-Hall.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, *10*(2), 169–181. https://doi.org/10.1016/0925-2312(95)00020-8

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera-Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, *25*(5), 716–742. https://doi.org/10.3846/tede.2019.8740

Krollner, B., Vanstone, B., & Finnie, G. (2010, April). Financial time series forecasting with machine learning techniques: A survey. In *18th European Symposium on Artificial Neural Networks* (*ESANN 2010*)*: Computational Intelligence and Machine Learning* (pp. 25–30). Bruges, Belgium.

Kuang, Y., Singh, R., Singh, S., & Singh, S. P. (2017). A novel macroeconomic forecasting model based on revised multimedia assisted BP neural network model and ant Colony algorithm. *Multimedia Tools and Applications*, *76*(18), 18749–18770. https://doi.org/10.1007/s11042-016-4319-9

Li, X. (2006). Temporal structure of neuronal population oscillations with empirical model decomposition. *Physics Letters A*, *356*(3), 237–241. https://doi.org/10.1016/j.physleta.2006.03.045

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

Liu, F., Kang, Y., Guo, K., & Sun, X. (2021). The relationship between air pollution, investor attention and stock prices: Evidence from new energy and polluting sectors. *Energy Policy*, *156*, 112430. https://doi.org/10.1016/j.enpol.2021.112430

Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies*, *27*(6), 1797–1829. https://doi.org/10.1093/rfs/hhu019

Makridakis, S., & Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*, *16*(3), 147–163. https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

Mangram, M. E. (2013). A simplified perspective of the Markowitz Portfolio Theory. *Global Journal of Business Research*, *7*(1), 59–70. https://ssrn.com/abstract=2147880

Markowitz, H. M. (1968). *Portfolio selection: Efficient diversification of investments*. Yale University Press.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2021). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 1–36. https://doi.org/10.1111/joes.12429

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

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), 1–17. https://doi.org/10.3390/su12166293

Mullainathan, S., & Thaler, R. H. (2000). *Behavioral economics* (NBER Working Paper No. 7948). https://doi.org/10.3386/w7948

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

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Pruitt, G., & Hill, J. R. (2012). *Building winning trading systems with Tradestation* (2nd ed.). (Wiley Trading Book 542). John Wiley & Sons. https://doi.org/10.1002/9781119204954

Rodriguez, D. (2020). *Backtrader*. https://www.backtrader.com/

Sampath, V. S., O'Connor, A. J., & Legister, C. (2022). Moral leadership and investor attention: An empirical assessment of the potus's tweets on firms' market returns. *Review of Quantitative Finance and Accounting*, *58*(3), 881–910. https://doi.org/10.1007/s11156-021-01012-0

Sansa, N. A. (2020). The impact of the COVID-19 on the financial markets: Evidence from China and USA. *Electronic Research Journal of Social Sciences and Humanities*, *2*(2), 29–39. https://doi.org/10.2139/ssrn.3567901

Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, *45*(5), 571–583. https://doi.org/10.1016/j.ipm.2009.05.001

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 106181. https://doi.org/10.1016/j.asoc.2020.106181

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Shen, D., Zhang, Y., Xiong, X., & Zhang, W. (2017). Baidu index and predictability of Chinese stock returns. *Financial Innovation*, *3*(1), 4. https://doi.org/10.1186/s40854-017-0053-1

Siami-Namini, S., & Namin, A. S. (2018). *Forecasting economics and financial time series: ARIMA vs. LSTM*. arXiv. https://doi.org/10.48550/arXiv.1803.06386

Smales, L. A. (2021). Investor attention and global market returns during the COVID-19 crisis. *International Review of Financial Analysis*, *73*, 101616. https://doi.org/10.1016/j.irfa.2020.101616

Statcounter GlobalStats. (2020). *Search Engine Market Share China*. https://gs.statcounter.com/search-engine-market-share/all/china/#yearly-2020-2020-bar

Su, F., & Wang, X. (2021). Investor co-attention and stock return co-movement: Evidence from China's A-share stock market. *The North American Journal of Economics and Finance*, *58*, 101548. https://doi.org/10.1016/j.najef.2021.101548

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets. *BIS Quarterly Review*, *3*, 113–131.

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

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, *41*, 17–35. https://doi.org/10.1016/j.jbankfin.2013.12.010

Yeh, J. R., Shieh, J. S., & Huang, N. E. (2010). Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis*, *2*(02), 135–156. https://doi.org/10.1142/S1793536910000422

Yu, L., Wang, S., & Lai, K. K. (2008) Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), 2623–2635. https://doi.org/10.1016/j.eneco.2008.05.003

Zhang, D., Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, *36*, 101528. https://doi.org/10.1016/j.frl.2020.101528

Zhang, D., & Lou, S. (2021). The application research of neural network and BP algorithm in stock price pattern classification and prediction. *Future Generation Computer Systems*, *115*, 872–879. https://doi.org/10.1016/j.future.2020.10.009

Zhang, Y., Chu, G., & Shen, D. (2021). The role of investor attention in predicting stock prices: The long short-term memory networks perspective. *Finance Research Letter*s, *38*, 101484. https://doi.org/10.1016/j.frl.2020.101484

Zhu, B., Wang, P., Chevallier, J., & Wei, Y. (2015). Carbon price analysis using empirical mode decomposition. *Computational Economics*, *45*(2), 195–206. https://doi.org/10.1007/s10614-013-9417-4