A hybrid forecasting method for quantitative investments based on GA-VMD and SSA-DELM optimization

Ruitian Zhang[1, #], Zebang Deng[2,#], Yicheng Shen[2,#,*]

[1]School of Financial Technology, School of Hebei Finance University, Baoding, China

[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

[2] School of Accountancy, Zhejiang University of Finance & Economics, Hangzhou, China

[*]Corresponding author: vj162774@163.com

[#]These authors contributed equally.

Abstract. In the modern financial market, quantitative investment is a more advanced investment method, which has been widely used in the capital market with the characteristics of simplicity, efficiency, and plasticity. Compared with other countries, China started late in the quantitative field, and the strategies are not perfect enough to be studied in depth. To study quantitative investment in-depth, this paper proposes a GA-VMD-SSA-DELM quantitative model. By building an extreme learning machine (DELM) model optimized by the sparrow algorithm (SSA), and simultaneously using a variational modal decomposition model (VMD) optimized by the genetic algorithm (GA) for data noise reduction, a quantitative investment model is constructed, and a portfolio strategy is formed. The method first performs correlation tests on the data sources, uses the daily frequency trading indicators of Guizhou Maotai, Zhifei Bio, and Yangtze River Power for the years 2018-2022 as the database, and uses SPSS to conduct Pearson correlation coefficient analysis to establish the correlation between the data. Then the GA-VMD algorithm was used for data noise reduction. Finally, the SSA-DELM algorithm is used to establish a quantitative investment model, construct a portfolio, and draw relevant conclusions.

Keywords: Quantitative investment, Genetic algorithm, Variational modal decomposition, Sparrow algorithm, Extreme learning machine.

1. Introduction

Quantitative investment in securities is an important investment approach in the financial sector [1,2]. Financial markets have evolved rapidly in recent years, and the investment environment has become increasingly complex and digital. Machine learning and deep learning investment models are widely used and put into practice by institutional investors to help decision-makers. Therefore, using quantitative investment strategies is gradually becoming an emerging mainstream investment approach. Predicting short-term market fluctuations, i.e., trend investing, is essential to maintain investors' ability to achieve excess rates of return in the capital markets in the short term. Traditional investment methods rely on the subjective experience and judgment of the leading investor to make decisions, which is risky and uncertain in predicting short-term volatility, often making it difficult to achieve the desired returns. The use of quantitative investment models to predict short-term market volatility can provide a greater degree of hedge against the unsystematic risk posed by the investor body, while also providing investors with more comprehensive and reliable back-testing information.

Reviewing the development process of quantitative methods, Xu P. et al. [3] explored the future direction of multi-factor models based on the capital asset pricing model and the Fama-French five-factor model. Guida T. [4] proposed that quantitative trading would evolve from single-factor to multi-factor models based on capital asset pricing theory [5]. Based on the significance of information coefficients and the results of single-factor backtesting [6, 7], Wang X. et al. selected the most optimal stock selection factors from a BP factor library and used the random forest algorithm to construct a quantitative stock selection model. In a practical sense, Tang Z. et al. [8] made investment recommendations for investors to benefit from market volatility. Pan H. [9] explored the breadth of the use of multi-factor models in quantitative approaches.

In financial markets, early stock forecasting was mainly based on traditional time series, Hidden Markov [10] models and other methods. Yongpei H. [11] proposed an improved algorithm for estimating the parameters of the ARMA model, which achieved better forecasting results. Kun M. [12] used the ARMA model to forecast the stock prices of the Shanghai Composite Index, and the results showed that the ARMA model was more suitable for short-term forecasting. However, traditional forecasting models are challenging to achieve the desired results in terms of data processing. Very often, the data at hand is incomplete, noisy, discordant, or multi-modal. A new approach to signal processing, based on variational mode decomposition (VMD) and genetic algorithms (GA), can effectively solve the noise problem.

Genetic algorithms provide a general framework for solving non-linear, multi-model, multi-objective and other complex system optimization problems. They are a powerful tool for solving large-scale optimization problems with parallel search and population search. Genetic algorithms have also been widely used in securities investment analysis, and experts such as Richard J. [13,14] have written about genetic algorithms and investment strategies. A mathematical model of genetic algorithms for stock market investment analysis was presented by Li M. [15] and others, and Allen F. [16] applied genetic algorithms to study the technical trading rules of the S&P 500 index. Since genetic algorithms provide a general framework for solving nonlinear, multi-model, multi-objective and other complex system optimization problems, and that they have the characteristics of parallel search, population search and the advantages of VMD methods in dealing with complex nonlinear, non-stationary, multi-scale signals and genetic algorithms in classification, this paper applies them to the study of stock investment price trend noise reduction.

Extreme learning machines (ELMs) are a class of machine learning methods based on Single-hidden Layer Feedforward Neural Networks (SLFNs), which are widely used in various prediction problems [17].

However, a single-limit learning machine is unstable in classifying data. To overcome this drawback, more and more researchers are considering the use of integrated ELMs. Furthermore, the combination of these factors motivates us to adopt ELMs as the base classifier in integrated learning with a new approach, namely the Sparrow Search Algorithm (SSA) search algorithm with the deep extreme learning machine (DELM) algorithm. The artificial bee colony algorithm is an optimization technique that simulates the foraging behavior of bees, is simple in structure, has fewer parameters, is easy to implement, and has been successfully applied to various practical problems. Incorporating DELM can improve the algorithm's performance and provide more reliable support for the algorithm results. The main contributions of this paper are summarised as follows:

This paper considers the impact of correlation between different stock price factors and investigates the problem of the impact of different factors on changes in price volatility trends using multiple data sets. Unlike most quantitative modeling studies, this paper proposes a noise reduction method combining genetic algorithm (GA) parameter optimization for VMD (GA-VMD) to optimize the validity and reliability of the prediction model results. In this paper, we develop a hybrid model called ABC-DELM to forecast the price movements of securities, in which we use an artificial bee colony algorithm to find the optimal trend

2. Model methodology

2.1 GA-VMD

2.1.1 GA

Due to external environmental interference and the influence of transmission tools, the signal is susceptible to noise interference in the transmission process, resulting in distortion of the original signal, or even causing the object signal to be completely covered. The presence of noise will bring difficulties to the detection and identification of the target signal, and will also affect the accuracy of the analysis results. Therefore, a GA algorithm with an excellent noise reduction effect is needed for noise reduction of the data source.

GA (Genetic Algorithm, GA) originated from studying computer simulations performed on biological systems. It is a stochastic global search and optimization method that mimics the evolutionary mechanism of organisms in nature, drawing on Darwin's theory of evolution and Mendel's theory of genetics. It is an efficient, parallel, global search method that automatically acquires and accumulates knowledge about the search space during the search process. It adaptively controls the search process to find the best solution. The algorithm follows the principle of "survival of the fittest, survival of the fittest". It is a type of randomized search method that draws on natural selection and genetic mechanisms in nature.

2.1.2 VMD

As the stock data is susceptible to external factors and contains numerous complex information, the traditional wavelet analysis is susceptible to white noise, while classical modal decomposition is easily disturbed, which greatly impact the decomposition of data and over. SSA algorithm is suitable for processing and analyzing nonlinear, non-stationary, and chaotic signals. It can extract the trend or periodic components embedded in the time series, thus decomposing the series into a few interpretable and independent components, which facilitates the prediction.

Variational Modal Decomposition (VMD) is a new type of time-frequency analysis algorithm that can decompose multicomponent signals into multiple but component signals, avoiding problems such as spurious components encountered in the iterative process. The decomposition process of VMD is the solution process of the variational problem, in which the intrinsic mode function (IMF) is defined as a bandwidth-constrained amplitude-modulation function. The function of the VMD algorithm is to construct and solve the constrained variational problem by decomposing the original signal. The function of the VMD algorithm is to decompose the original signal into a specified number of IMF components by constructing and solving a constrained variational problem.

2.1.3 GA-VMD

To optimize the VMD algorithm using GA, we need to go through six steps: coding, population initialization, calculation of individual fitness, selection, crossover, and variation, and finally, make the population evolve into a better-adapted population. In this optimization process, "calculating individual fitness" plays a key role. First, we need to define a proper function to calculate the fitness value to evaluate the difference between individuals and the optimal value. The larger the fitness, the higher the probability that individuals will be inherited by the next generation. The envelope entropy is usually chosen as the fitness function of GA. The envelope entropy can well reflect the uncertainty of the signal, and the larger the entropy value, the greater the signal instability.

2.2 SSA-DELM

2.2.1 SSA

SSA (Sparrow Algorithm) is a novel population optimization algorithm proposed by Jiankai Xue et al. in 2019. In this algorithm, the sparrows in the population are often divided into two parts: joiners and discoverers, with each sparrow becoming a discoverer and another sparrow becoming a joiner, their identities are constantly changing. In addition, each sparrow will be given an initial position and a corresponding fitness.

In the SSA algorithm, the magnitude of the discoverer's food search capability is represented by the degree of adaptation, and the discoverer's position is updated during the algorithm's operation in the following manner.

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t}\exp\left(-\frac{i}{\alpha i_{\text{ter max}}}\right), \ R_2 < S_T \\ X_{i,j}^{t} + \varphi L, \ R_2 \geq S_T \end{cases} \tag{1}$$

Where t denotes the number of iterations; j denotes the number of dimensions.$i_{ter\ max}$ represents the maximum number of iterations. α is a random number belonging to the interval (0,1]. Φ is a random number that obeys normal distribution.S_T represents the safety value.R_2 represents the warning value.$R_2 < S_T$ represents the population is in a safe area where the finder can randomly predate, and$R_2 \geq S_T$ represents the presence of predators around the population and the need to move to a safe area for foraging.

If a joiner in the population finds a finder foraging for better food, it will immediately grab the food. If the joiner is successful, it will get the food from the finder; if not, it will continue to monitor the finder for food. The location of the finder is updated as follows.

$$X_{i,j}^{t+1} = \begin{cases} \varphi \exp\left(\frac{X_{worse} - X_{i,j}^{t}}{i^2}\right), & i < n/2 \\ X_P^{t+1} + \left|X_{i,j}^{t} - X_P^{t+1}\right| A^{*} L, & \text{other} \end{cases} \tag{2}$$

WhereX_P represents the best position in which the discoverer is located.X_{worse} represents the worst position in the current population. A denotes a 1 * d matrix, and each element is randomly assigned 1 or -1.$i > n/2$ denotes the ith accession with a low fitness value that does not acquire food and needs to go elsewhere to forage.

2.2.2 DELM

The extreme learning machine is a single hidden layer feedforward neural network (SLFN). The input weights and thresholds of the hidden layer are randomly generated in the training of the extreme learning machine network, and the output weights can be calculated only by the generalized inverse matrix principle to complete the learning process, thus the ELM network has the advantages of fast learning speed and strong generalization ability.

Encoder (AE) is an unsupervised learning algorithm that can be used for high-latitude complex data processing and feature learning. If the encoder idea is applied to the extreme learning machine, the total input and output of the network model can be made the same. If $n = \mathring{n}$, ELM-AE can achieve equal dimensional feature expression, and if $n < \mathring{n}$, it can achieve high latitude feature value expression.

For high dimensional feature representation with dimensionality reduction, the output weight of the hidden layer $\boldsymbol{\beta}$ can be expressed as

$$\boldsymbol{\beta} = \left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} + \frac{1}{C}\right)\boldsymbol{H}^{\mathrm{T}}\boldsymbol{X} \tag{3}$$

For the equal dimensional feature expression, the output weight of the hidden layer $\boldsymbol{\beta}$ can be expressed as

$$\boldsymbol{\beta} = \boldsymbol{T}\boldsymbol{H}^{-1} \tag{4}$$

DELM (Deep Extreme Learning Machine) is a deep learning network formed by superimposing multiple ELM-AEs, which improves the accuracy and generalization ability of the model by better mapping of data features due to the superimposition of multiple hidden layers. In the training process of DELM model, the training data is often used as the output of the first ELM-AE layer to obtain the output weight $\boldsymbol{\beta}_1$. The output in the hidden layer of the DELM-AE model is used as the in and out of the second ELM-AE model to complete the training for each ELM-AE model layer in this way. The DELM model is adopted to train and predict each modality after decomposition, which can improve the model's accuracy.

2.2.3 SSA-DELM

Multiple ELM-AE models superimpose the DELM model, and the output proportions of ELM-AE are used to initialize the DELM. In the ELM-AE training process, only the proportions and thresholds of the output layer are calculated by the least-squares method for correlation. The input layer is randomly generated by the orthogonal matrix, which makes the prediction results of the DELM model have great volatility. Adopting SSA (sparrow algorithm) for the optimization of the model can reduce such effects to some extent, and the specific steps are as follows.

3. Example measurement

3.1 Forecast data

In order to construct a portfolio strategy, data forecasting is required. This paper selects the traditional three major sectors of the Chinese stock market - alcohol, electric power, and pharmaceuticals - as the research objects, and selects representative stocks in each sector - Guizhou Maotai, Yangtze River Electric Power, and Jiffy Bio - as experimental subjects, with 1038 trading records of each from January 2018 to April 2022 as the database. Pearson data analysis was performed using SPSS to investigate the correlation between the three. The correlation formula is as follows.

$$\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]}{\sqrt{\sum_{n}^{i=1}\left(X_i-\mu_X\right)^2}\sqrt{\sum_{n}^{i=1}\left(Y_i-\mu_Y\right)^2}} \tag{5}$$

Pearson correlation coefficient The range of values is [-1,1], and the parameters close to both ends have a strong correlation, while the correlation parameters close to 0 are less correlated. The results of the runs are shown in Table 1.

Table 1 Pearson correlation analysis

Ticker		GZMT	CJDL	ZFSW
GZMT	Pearson	1	0.865	0.929
	Sig		0	0
CJDL	Pearson	0.865	1	0.746
	Sig	0		0
ZFSW	Pearson	0.929	0.746	1
	Sig	0	0	

Considering the situation that stocks may influence each other, a correlation analysis needs to be conducted for the three stocks. Taking Guizhou Maotai as an example, the Pearson correlation values of Yangtze River Power and Jiffy Biologicals and their Pearson correlation values are 0.865 and 0.929, which are positively correlated, and the values are close to the benchmark correlation level with a high correlation degree. The rest of the results based on different evaluation benchmark stocks are similar. Thus, it can be concluded that the three stocks have a high degree of correlation and have a greater degree of influence on the portfolio return in quantitative investment, which needs further study.

Certain factors within individual stocks can also impact returns, so correlation analysis of variables within stocks is needed. Here, Guizhou Maotai is used as a representative, and the closing price, high price, low price, up/down range, turnover, effective turnover ratio, and PE P/E ratio are selected as the influencing indicators for validity analysis, and the experimental results are shown in Table 2. With the closing price as the benchmark, each indicator, and its Pearson correlation values were 1, 0.999, 0.014, 0.625, -0.287, 0.937. The highest price, the lowest price, turnover, PE P/E ratio and the closing price were positively correlated, and the value is close to the benchmark correlation level, with a high degree of correlation.. Effective turnover rate and the closing price were negatively correlated. The correlation between the effective turnover rate and the closing price is negative, with a low degree of influence. The correlation between the upside and downside and the closing price is close to 0, with a low degree of correlation and a low degree of influence. The rest of the results based on different benchmark indices are similar. It can be concluded that there is a specific correlation between the impact indicators among individual stocks, which impacts the portfolio return and needs to be further studied.

Table 2 Pearson correlation analysis

Factor		Closing Price	Highest Price	Lowest Price	Up or Down	Turnover	Effective turnover rate	PE P/E Ratio
Closing Price	Pearson	1	1	0.999	0.014	0.625	-0.287	0.937
	Sig		0	0	0.641	0	0	0
Highest Price	Pearson	1	1	0.999	-0.004	0.635	-0.278	0.937
	Sig	0		0	0.890	0	0	0
Lowest Price	Pearson	0.999	0.999	1	-0.005	0.614	-0.298	0.936
	Sig	0	0		0.867	0	0	0
Up or Down	Pearson	0.014	-0.004	-0.005	1	-0.019	-0.023	0.029
	Sig	0.641	0.890	0.867		0.550	0.462	0.353
Turnover	Pearson	0.625	0.635	0.614	-0.019	1	0.477	0.602
	Sig	0	0	0	0.550		0	0
Effective turnover rate	Pearson	-0.287	-0.278	-0.298	-0.023	0.477	1	-0.246
	Sig	0	0	0	0.462	0		0
PE P/E Ratio	Pearson	0.937	0.937	0.937	0.029	0.602	-0.246	1
	Sig	0	0	0	0.353	0	0	

3.2 GA-VMD

To verify the validity of the experimental conclusions in this paper, the GA-VMD method is used to perform noise reduction on the simulated signal and stock data. The GA algorithm is used to optimize the VMD parameters, and the IMF decomposition is obtained by decomposing the signal according to the optimized parameters, combining it with the envelope entropy to remove the noise components, and reconstructing the remaining components to obtain the noise-reduced signal finally.

To take the VMD method for data noise reduction, it is first required to find the optimal combination of parameters for this signal by using the GA algorithm, and the process of the closed-loop of the adaptation curve is shown. From Fig. 1, as can be seen, the number of iterations is carried out to the 2nd time, after which the optimal adaptation degree is reached. At this time, the signal is decomposed by VMD according to the optimal adaptation combination to achieve the best decomposition effect. The four IMF components obtained after VMD decomposition are shown in Fig. 2 shown, and it can be seen from the figure that the low-frequency oscillations mainly appear in the first few orders of modes.

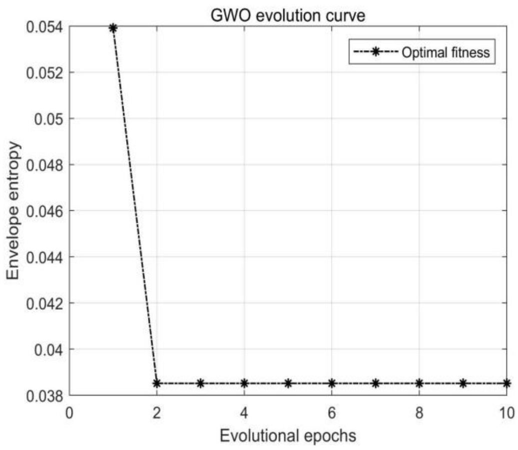

Figure 1 Adaptation-iteration number change graph

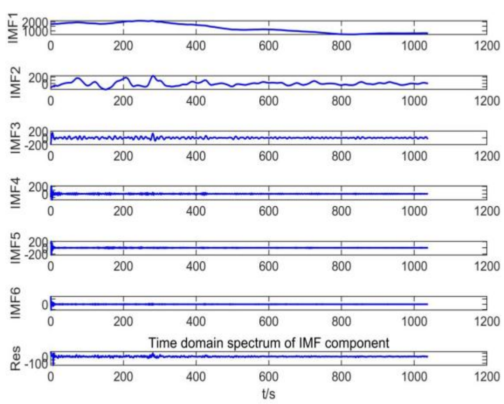

Figure 2 GA-VMD signal component diagram

In order to accurately filter the signal and noise components, the envelope entropy of each component needs to be calculated. Set the appropriate population size popsize, iter number of iterations iter, number of variables dim, crossover probability pc, and variation probability pm, and take popsize=6, iter=10, dim=2, pc=0.8, and pm=0.3 after several tests. Calculate the envelope entropy for each component, and the calculation results are shown in Table 3.

Table 3 Table of envelope entropy values

IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
1.1609	1.6820	0.8476	3.4375	0.0539	0.0693

After optimization, the center frequencies of each mode and the components of each mode are shown in Fig. 3. As shown in the figure, the overlap between each model component and its center frequency is small, and the decomposition effect is more excellent. It can be seen from the figure that the peak difference between each modal component and the center frequency is large, which indicates that there is no modal mixing phenomenon, the model optimization effect is good, the data noise reduction effect is excellent, and the output results can be used for model construction.

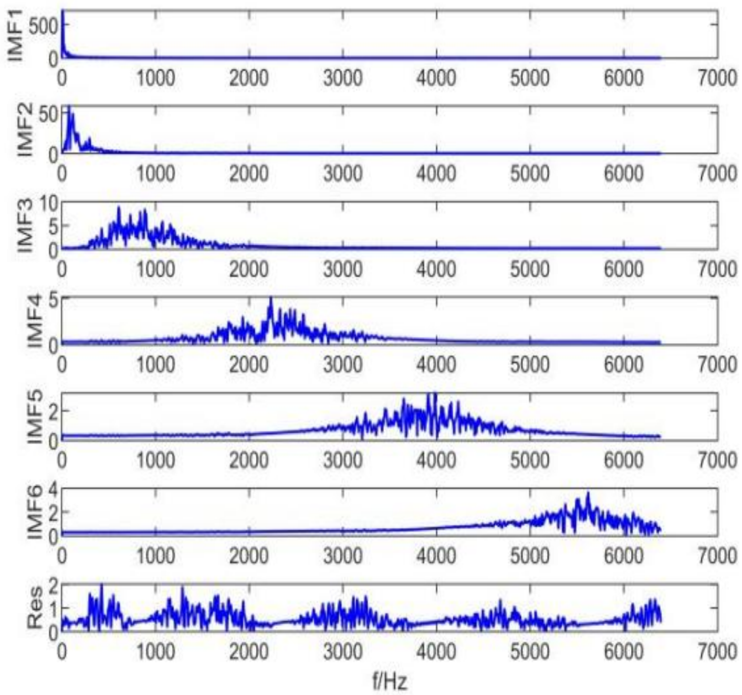

Figure 3 GA-VMD components of each mode

3.3 SSA-DELM Experiment

To further illustrate the effectiveness of the SSA-DELM model, common forecasting models DELM in the past literature were selected for comparison, where the focus was on testing the decomposed signal components IMF1, IMF2, IMF3, IMF4, IMF5 and IMF6 by testing them separately.

Take Guizhou Maotai data as an example. A total of 1,038 transaction records from January 2018 to April 2022 are selected for the database. As the daily stock price is affected by numerous factors, the target input for this part of the experiment is the six closing price components after the previous GA-VMD decomposition and noise reduction, and the data for each IMF component are predicted separately. The first 600 data of each component were selected as the training set and the last 300 data as the prediction set, making a balanced ratio of 2:1 between the two, effectively taking into account a sufficient number of training samples as well as a reasonable prediction range. In the process of processing the data, the choice of time dimension also determines the prediction accuracy. Suppose the time dimension chosen is too long. In that case, the contribution of the historical data too far away from the current prediction data will be small, making the training of the model ineffective, thus affecting the prediction accuracy of the model. Conversely, it will lead to the loss of some of the data features and the lack of information on the modal decomposition process, which will not be able to map the historical relationship between the training data and the test data, thus affecting the prediction accuracy of the model. The use of daily frequency data, therefore, provides a good trade-off between the integrity of the data features and the contribution of historical data to the current prediction data.

The parameters of the SSA-DELM method were set as follows: the population size of sparrow popsize was in the range of [30, 70]; the maximum number of iterations was in the range of [50, 70]; the lower boundary of DELM weights was -1, and the upper boundary was 1; the DELM model chose sigmoid as the activation function, the number of implied layers was 1 or 2, and the number of nodes in each layer was 4. variance as the fitness function. The SSA method showed good convergence speed and global search ability in the parameter search process for each IMF component. Convergence was reached when the number of experimental iterations was close to between [20, 30].

There are more error evaluation indicators in stock price forecasting, and this paper uses mean square error (MSE) as well as mean relative error (MAPE) to measure the model. The MAPE expression is as follows:

$$\sigma_{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{|\hat{y}(i)-y(i)|}{y(i)}\right| \times 100\% \tag{6}$$

In this expression, n is the number of data in testing set; y_i is the real price of stock in timing i; $\hat{y}_i$ is the predicted price of stock in timing i. MAPE evaluates the overall accuracy; MSE evaluates the error fluctuations. The smaller the value of MSE, the smaller the error and the better the prediction. The smaller the value of MAPE, the smaller the overall fluctuation of the error. The test errors obtained by summing all the components are organized in the table shown in Table 4.

Table 4 Table of prediction error values for the test set

	MAPE/ %	MSE/ ¥
SSA-DELM	5.898	42699.044
DELM	22.335	183369.357

The comparison of the results of the summed prediction values of all components is shown in Fig.6, and the prediction results of SSA-DELM are significantly better than those of DELM. Meanwhile, comparing the prediction results of each specific IMF component test set reveals that the prediction errors of SSA-DELM in six components, IMF1, IMF2, IMF3, IMF4, IMF5, and IMF6, are smaller than those of the DELM prediction model. Taking IMF1 as an example with Fig.7, Fig.8, and Fig.9, the article shows that SSA-DELM predicts better values than DELM, and the former error is smaller than the latter, and the iterative process of optimal fitness value change. It is shown that SSA-DELM outperforms the DELM prediction model in predicting the stock price (closing price). Compared with the DELM model, the SSA-DELM model has better prediction results, which is due to the search optimization capability of SSA to perform the search optimization process for the input parameters of the hidden layer of DELM, which avoids the errors arising from the random generation of parameters and thus improves the prediction accuracy of the model; with the error correction based on the GA-VMD-SSA-DELM model, the prediction accuracy is further improved, which is due to the existence of error components in the GA-VMD decomposition. Thus the GA model is used to further predict the error series. The prediction results of the error series are overlaid with the initial prediction results of the stock price time series for correction processing to obtain the final prediction results.

Overall analysis, the GA-VMD-SSA-DELM quantitative model proposed in this paper can better predict the trend of Guizhou Maotai stock price (closing price) on different components of MSE and MAPE indicators, which thoroughly verifies the effectiveness of the model. The model's superiority is also fully verified by comparing it with the traditional quantitative model.

4. Conclusions and discussion

In this paper, aiming at the core problem in the financial investment field-the core problem of stock price forecasting and modeling in the stock market, the noise reduction method combined with genetic algorithm (GA) parameter optimized VMD (GA-VMD) is applied to decompose and reorganize the data. Moreover, the high-frequency, intermediate-frequency and low-frequency subsequences with simpler fluctuation characteristics are generated, which creates favorable conditions for the model to further extract complex patterns in the sequences and significantly reduce the original data noise. Besides, the SSA-DELM model is applied to the experiment to forecast the final result. By comparing to other methods, GA-VMD-SSA-DELM shows advantages in the following ways:

GA-VMD creates favorable conditions for the model to further extract complex patterns from the sequences, and significantly reduces the interference of noise and complex patterns in the original data. It makes the idea of decomposing and reorganizing the stock price data by GA-VMD and then constructing the prediction model separately and summing up to obtain the overall prediction value reasonable and feasible.

In this paper, based on the GA-VMD noise reduction decomposition of stock price series, an advanced hybrid forecasting model SSA-DELM is developed and adopted. Combining the features of the DELM model that can better achieve the mapping of data features, and the features that can improve the accuracy and generalization ability of the model, the construction idea of SSA-DELM stock price forecasting model is reasonable and feasible. Finally, three representative companies in the domestic stock market are selected as experimental data, namely Guizhou Maotai, ZhiFei Biological and Yangtze River Power. Some traditional forecasting models with proven forecasting effects are used as the comparison method for the experiment, including the forecasting model that passes DEIM directly after VMD decomposition and noise reduction at the end, and the forecasting model that passes DELM after VMD decomposition and noise reduction. The prediction validity and adaptability of the GA-VMD-SSA-DELM model were evaluated quantitatively in multiple dimensions. The experimental results confirm that the GA-VMD-SSA-DELM forecasting model exhibits low error and high accuracy in the case of stock price forecasting.

However, the selection of parameters such as the number of iterations and the number of populations in the debugging process of the SSA-DELM prediction model in this paper is still subjective. Meanwhile, the fitting effect of SSA-DELM also needs to improve further the accuracy for the volatility presented by the data decomposed by GA-VMD with noise reduction, so the optimization treatment of model parameters needs to be further studied.

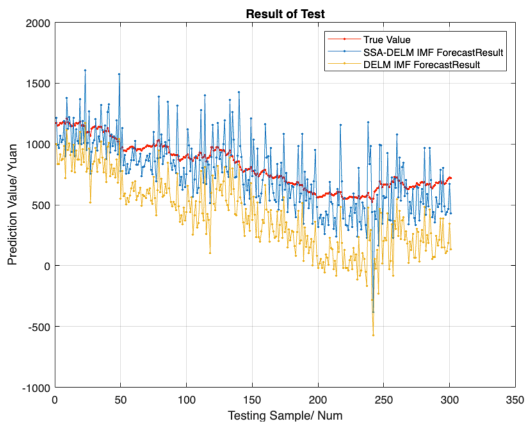

Figure 6 Comparison of IMF Component Aggregate Forecast Results

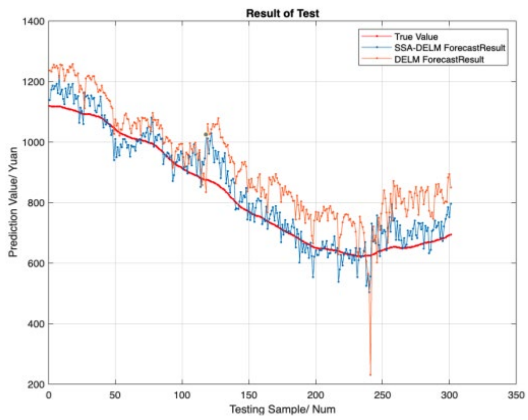

Figure 7 IMF1 component prediction results

Figure 8 IMF1 component prediction error

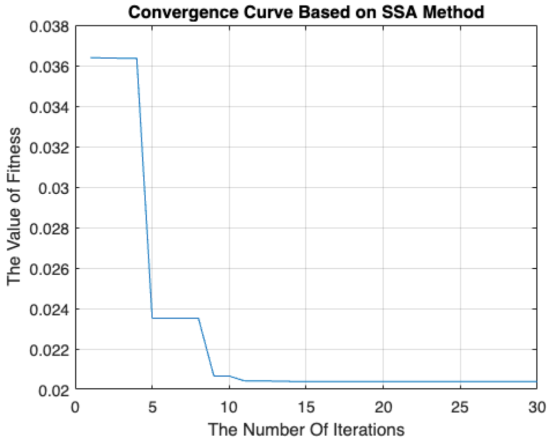

Figure 9 Iterative process of changing the optimal fitness value of IMF1 component SSA-DELM

References

[1] ANAGNOSTOPOULOS I.Fintech and regtech:Impact on regulators and banks[J].Journal of Economics and Busines，2018 (100):7-25.

[2] LEE I， SHIN Y J.Fintech:Ecosystem， busines models， investment decisions， and chalenges[J].Busines Horizons，2018，61(1): 35-46.

[3] Xu P, Wang M, Liu J. Establishment and Optimization of Multi-factor Model[J]. 2020.

[4] Guida T. Big data and machine learning in quantitative investment[M]. John Wiley & Sons, 2019.

[5] Putnam B. Death by Simulation[M]//Economics Gone Astray. 2019: 161-170.

[6] Zhang,Ling.Equilibrium investment strategy for a defined contribution pension plan under stochastic interest rate and stochastic volatility[J].JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS,2020,368.

[7] Wang X, Yin S. Research on Quantitative Investment Strategy Based on Random Forest Model and Risk Management[C]//2021 4th International Conference on Information Systems and Computer Aided Education. 2021: 2126-2131.

[8] Tang Z, Cheng Y, Wang Z. Quantified Investment Strategies and Excess Returns: Stock Price Forecasting Based on Machine Learning[J]. Academic Journal of Computing & Information Science, 2021, 4(6).

[9] Pan, H. (2021, July). Chinese Value Investing Theory and Quantitative Technology. In 2021 IEEE 19th International Conference on Industrial Informatics (INDIN) (pp. 1-6). IEEE.

[10] Wang Chen. Stock price index forecasting based on hidden Markov model [D] . Jinan: Shandong University, 2018.

[11] He YP. Improvement of ARMA model parameter estimation algorithm and its application in stock forecasting [J]. Journal of Chongqing Institute of Technology (Natural Science Edition), 2009, 23(2): 109 - 112.

[12] Meng Kun, Li Li. An empirical analysis of stock price prediction based on ARMA model [J]. Journal of Hebei North College: natural science edition,2016,32(5):55-60.

[13] Jones R. The trading game: playing by the numbers to make millions[M]. John Wiley & Sons, 1999.

[14] Bauer R J, Dahlquist J R. Technical Markets Indicators: Analysis & Performance[M]. John Wiley & Sons, 1998.

[15] Li MQ, Zhang JF, Kou Jisong. Application of genetic algorithms in the study of stock market investment strategies (strategies)[J]. Systems Engineering Theory and Practice,1998(08):20-26.

[16] Allen,,F et. al. Using genetic algorithms to find technical trading rules[J].JOURNAL OF FINANCIAL ECONOMICS,1999,51(2):245-271.

[17] HUANG G B. Extreme learning machine: A new learn ing scheme for feedforward neural networks[J]. Proc. intl Joint Conf.on Neural Networks, 2004.