Predicting stock price based on BP neural network model

Yishuai Tian1*, Botao Liu1, Boying Lv1

[1]College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang 443002, China.

ABSTRACT

The increase in the stock will affect the development of the economy, so the description and forecast of the capital are fundamental. Therefore, this paper first analyzes the historical data of Everbright Bank stock (601818) and explains the stock increase, the share affected by the exchange rate, and the index correlation analysis. Then a prediction model based on BP neural network is established. BP is used to train stock descriptors to predict stock prices for the first 30 trading days in early 2019. Combined with ARE and MSE for error analysis, the average relative error is 3.6%, and the mean-variance is 0.041, indicating that the prediction effect of the model is good. To make a specific contribution to the study of stock price prediction.

Keywords: Stock forecast; BP neural network; Stock analysis error check

*Correspondence to Author:

Yishuai Tian

College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang 443002, China.

1. Introduction

Stocks are the product of a market economy. The stock market is of considerable significance to the healthy development of the national economy today. The high returns and high risks of the stock market coexist, and the instability of the stock market easily damages the interests of investors. It can reasonably and adequately predict the fluctuations of the stock market, conduct reasonable supervision and regulation of the stock market, and arrange correct investment schemes, which will provide a reliable guarantee for the healthy development of the national economy [1]. The stock market is an extremely complex system. The combination and influence of many factors and many uncertainties make stock prediction very difficult, mainly manifested in the non-linear characteristics and instability of stock price fluctuations [2, 3].

Many scholars at home and abroad have actively invested in the research of stock market forecasting methods. For example, Massimiliano [4] and others used time series to predict changes in macroeconomic indicators affecting stocks. Kimoto [5] developed the Tokyo stock price system, combined with multiple regression statistical theory, and adopted an algorithm based on neural networks to predict the stock price. Liting Gu et al. [6] developed a new network cognitive system using a neural network algorithm based on fuzzy theory. To verify the predictive ability of the system, the stock price of Shanghai Second Textile Company was selected as the prediction object and the final network prediction. The accuracy was very high. Song Li et al. [7] added a particle swarm algorithm to the neural network and analyzed the stock index. The prediction error obtained was better than the traditional prediction method. Kuihe Yang [8] used the autocorrelation function to apply the variables that had a function mapping relationship with the output variables in the sample data to the input variable set of the new prediction model, which improved the prediction effect of the network. Dash et al. [9] used an integrated TOPSIS crow search-based classifier set for the application of stock price prediction. Jiang

et al. [10] combined the Stacking method to predict the direction of the stock price index derived from the mainstream price of an earlier month. Qiu et al. [11] used a recurrent neural network to build a stock price prediction model of long-term and short-term memory neural networks based on the attention mechanism. Li et al. [12] used the past stock price, volatility, and trading behavior to predict stock return volatility. With the development of the social economy, the demand for predictive analysis is also increasing. The biological and mathematical characteristics of neural networks have made it an excellent potential for application in stock market predictive analysis and risk assessment. Therefore, many scholars are still researching and improving network systems.

Therefore, this paper selected the historical data of China Everbright Bank (601818), analyzed and described it by fitting, combined multiple regression analysis with the neural network, used BP neural network to train stock description factors, and used this model to predict the stock of China Everbright Bank. The price changed and performed an error analysis at the end. To provide a basis for further research on the application of BP neural network in stocks.

2. A case study of Everbright Bank Stock

2.1 Data Analysis of Everbright Bank Stock

Through Sina Finance (https://finance.sina.com.cn/) to collect the data of Everbright Bank, we get the daily historical data of Everbright Bank (601818) from the listing of A-shares on August 18, 2010, to December 31, 2018, followed by the currency exchange rate historical data of China Everbright Bank from 2013 to 2018, the daily market index from 2010 to 2018 and the typical stock data of selected industries.

2.1.1 Analysis on the change of Stock Price

The price of the stock in the circulation market is the market price of the stock in the complete sense, which is generally called the stock market price or the stock market. The changes in the stock price per share (based on the closing price) of each day from 2010 to 2018 are reflected in the line chart, and the closing price data is approximately fitted as a fluctuation
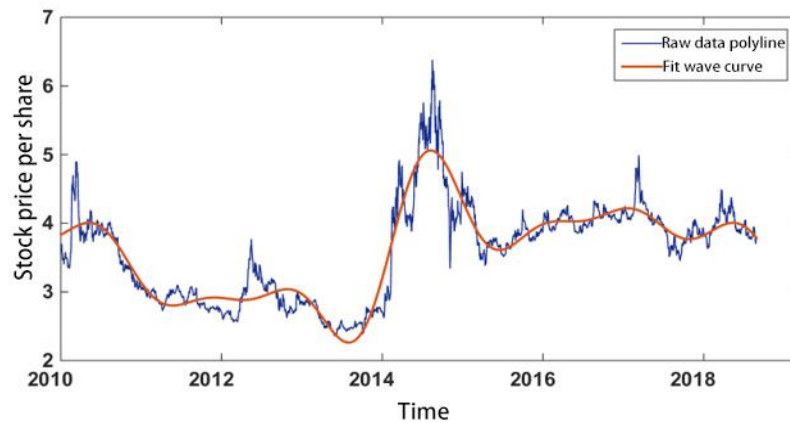
curve.

Fig. 1 data of closing price fitting volatility curve

It can be concluded from Fig.1 that the average price per share from 2010 to 2018 is relatively stable except for the sudden change of instability in some years. This reflects that the overall stability of bank stocks is not easy to rise and fall sharply. At the beginning of 2014, stock prices generally showed a trend of sharp rise and fluctuation, and showed a sharp decline in mid-2014, indicating that the stability of stock prices in the first half of 2014 was weak, attesting a sharp rise and fall in the extreme form. The period from 2011 to 2013 and 2015 to 2017 are the stable periods for Everbright Bank shares.

2.1.2 An Analysis of the relationship between RMB Exchange rate and Stock Price

As the daily return of stocks and the macroeconomic environment are closely related to the company's operating conditions, and the exchange rate can have an impact on both, it can also affect the stock price. Collect the stock prices of Everbright Bank and the US dollar exchange rate of RMB in Everbright Bank in 2017 and 2018, as shown in Fig.2.

Fig. 2 Line chart of stock price and RMB exchange rate

From Fig. 2, we can see that there is a negative correlation between the exchange rate and stock price. The overall relationship between the exchange rate and the bank stock price is a reverse movement, that is, a decline in the exchange rate. There is a two-way causal relationship between RMB exchange rate and bank stock price. For the interest rate difference between China and the United States, the impact on stock prices also changes due to the different exchange rate expectation environment and the operation of the central bank's foreign

exchange market.

2.1.3 Correlation Analysis of Stock Indexes

The lowest price, opening price, previous closing price, up and down, turnover, total market value, tradable market value, high and low difference% of the collected stocks, and the closing price number is 1Mui 13.The correlation among [13].The contingency table analyzes thirteen indexes, and the three-dimensional diagram is shown in Fig.3. It is found that the association of the lowest price, opening price, and previous closing price is almost close to 1, and the correlation coefficient of transaction volume, transaction amount, and the total market value fluctuates at 0.8-1.0. The other correlations are less than 0.8, indicating that the association is not significant.
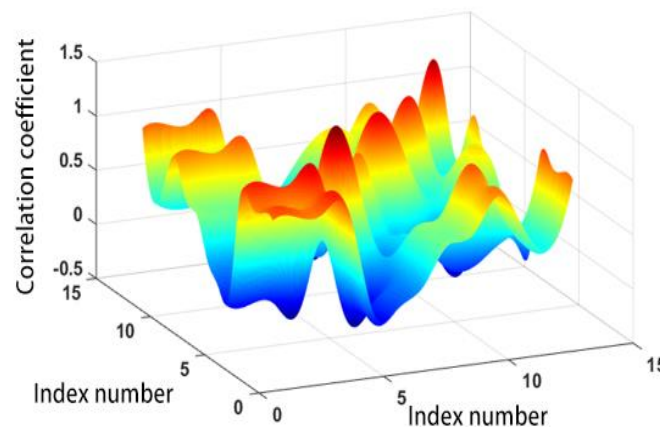

Fig.3 Visualization based on the association of stock indicators

2.2 Establishment of regression Prediction Model based on BP Neural Network

BP (Backpropagation) neural network is a kind of multi-layer feedforward network trained by the error backpropagation algorithm, and it is one of the most widely used neural network models at present.[14] When describing the price change of the stock, only the historical data of the capital, that is, 13 indicators, are taken into account. When predicting the price change of the stock, the exchange rate, the exchange rate change, and the stock index are comprehensively considered. The types of stock data are too complicated, and the processing of dimensions determines the complexity of the later calculation, so we predict the opening price, closing price, highest price and lowest price of the selected stock, and replace the original data with more concise data. Improve computational efficiency.

Establishment of non-linear continuous function $g(x)$**.**The stock market is a complex non-linear system with strong randomness and nonlinearity, while the artificial neural network has an excellent adaptive ability to meet the stock market prediction. Let the closing price on the day K be X_K, then $X_K = g(X^K)$, Among them, $X^K = (X_{K-1}, \cdots, X_{K-T})^T$, As a r result, the non-linear continuous function formula is established as follows:
$$X_K = g(K) \quad (K = 1, 2, \cdots, m)$$

The determination of input and output nodes of the neural network. Considering the actual stock situation, it is reasonable to think that the recent trading data have a more significant impact on the future than the forward trading data, so to predict the time series $t+n$ data, the latest historical trading data from time $t+n-1$ becomes the current window, where n is the window size, that is, the number of nodes in the input layer. Limited to one-dimensional time series, only one input node is needed to represent the predicted value of $t+n$ time.

The establishment of hidden layer node P. The number of hidden layer nodes are related to the training speed, and the number of hidden layer nodes depends on the number of training samples P, sample noise and the complexity

of the rules contained in the samples. Let x represent the number of input nodes represents the number of output nodes. The formula is as follows: $P = \sqrt{\left(x^2 + y^2\right)}$

The establishment of the activation function S. The S-type activation function expresses the non-linear approximation ability of the BP neural network, and the range of S-type activation function can only be between $(0,1)$ or $(-1,1)$. In this stock forecast, the linear activation function is used between the hidden layer and the output layer.

For the selection of initial parameters, the neural network model is established by selecting the changes in the learning rate, initial weight, number of cycles, momentum factor, and adjusting parameters.

- Learning rate：in general, a more significant learning rate will make the network unstable and begin to diverge if it does not reach the error range; too small will lead to long training time and slow convergence. The overall value is between $(0.01 \sim 0.8)$, , and the learning rate is 0.6 in this training.

- Initial weight: to make the error as small as possible, it is necessary to choose a resonable threshold and weight. If it is too large, it is easy to fall into the saturation zone. Therefore, this training is predicted by random generating function, and the formula is as follows: $\partial = (0,1)$

The range of the initial weight is the random number in $(0,1)$.

- The number of cycles: because the number of cycles is related to other parameters, a larger number of cycles will get more consistent results. In this prediction, training begins with a step size of 50.

- Determination of momentum factor a：momentum factor is an improved parameter of the BP algorithm, and its value range is $(0,1)$. The area where the sample space is relatively flat is generally suitable for smaller a and larger a can obtain greater learning speed, but it may damage details, reduce the stability of the network and increase local saturation and other uncertain factors. In this training, take the momentum factor: $a = 0.8$

- The determination of the adjustment parameter r：the parameter in the activation function is a random number between, which is used to adjust the activation function to make it converge quickly. The intermediate value of the initial value range is selected in this stock prediction training model. The BP network structure of the stock market forecasting model in this paper is shown in Fig.4. Where X_t is the n input value, X_{t+n} is the output value, that is, the predicted value and the number of nodes in the hidden layer is $r(1 < r < n)$ nodes.
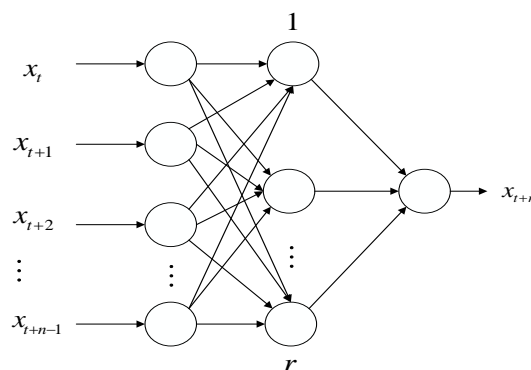

Fig. 4 BP neural network structure of stock market forecasting model

Determination of training objective function SSE. Assuming the expected, predicted value of the system $x_K\left(k=1,\cdots,N\right)$, Corresponding error e. Of which

$e=\left(e_1,e_2,\cdots,e_N\right)=\left(x_1-x_1',x_2-x_2',x_N-x_N'\right)$, Then the sum of square error（SSE）is used as one of the performance evaluation indexes to evaluate the performance of the neural network prediction system, so that the x_k' matrix represents the actual output value of the corresponding network, and the x_k is the expected output value.

We can get:

$$SSE=\frac{1}{2}\sum_k\left(x_k-x_k'\right)\quad\left(k=1,2,\cdots,N\right)$$

The establishment of a fitting regression equation. The regression equation with a high fitting degree is obtained by substituting the respective variables and dependent variables into the neural network training. Regression variable. In the process of using the regression equation to describe the original data, the dependent variables substituted into the training are: the opening price, the closing price, the highest price, and the lowest stock price, and the independent variables are two description factors. In the process of predicting the stock price by using the regression equation, the dependent variables substituted for the training are the four prices, namely, the opening price, the closing price, the highest price, and the lowest price.

2.3 Solution of regression Prediction Model based on BP Neural Network

2.3.1 Using Neural Network to predict the change of Stock Price

This paper makes regression predictions on the opening price, closing price, highest stock price, and lowest stock price of Everbright Bank stocks, and uses artificial neural network algorithms to implement them in MATLAB.[15] The following uses the prediction of the opening price of China Everbright Bank stocks in the first 30 trading days of 2019 as an example to illustrate the solution process.

- BP training. As shown in Fig.5, for a given training mode input, the output node of the network is calculated. The three lines are: Train represents the performance of the MSE indicator in the BP training process in each generation, MSE indicator performance in the BP cross-validation process in each generation, and MSE indicator performance in the BP test process in each generation. The red line represents the BP training results
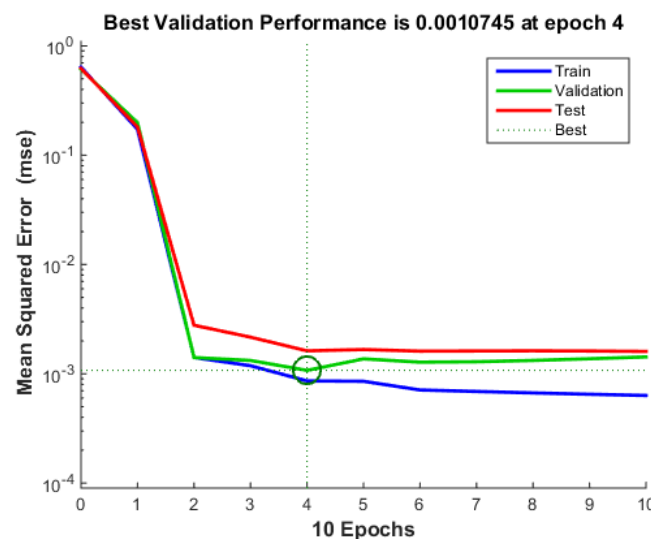

Fig.5 BP network algorithm implementation process simulation

● Provide training mode, training network, until the learning requirements are met, you can see that when the number of training is 55 times, the MSE reaches the expected order of magnitude, and the training stops at this time, and the effect is shown in Fig.6.
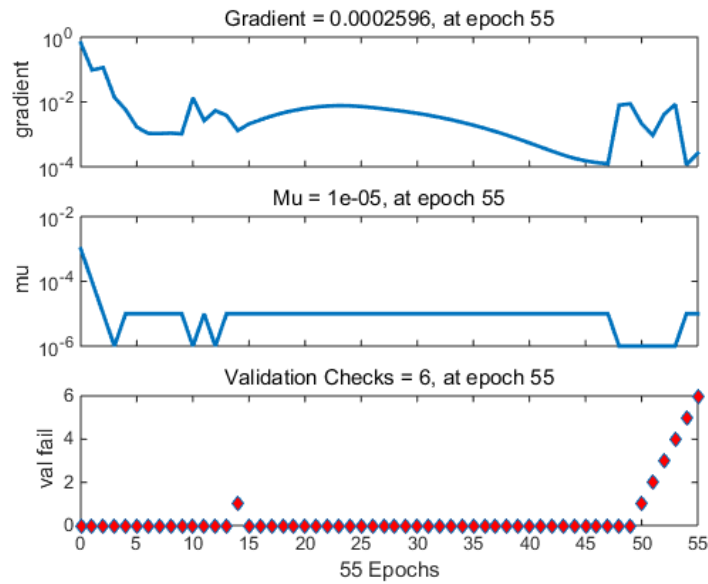

Fig.6 BP network algorithm realizes training mode

● Regression simulation. Calculate the error of the unit in the same layer, correct the weight and threshold, regression simulation, and get the following regression diagram. Get the opening price before anti-standardization
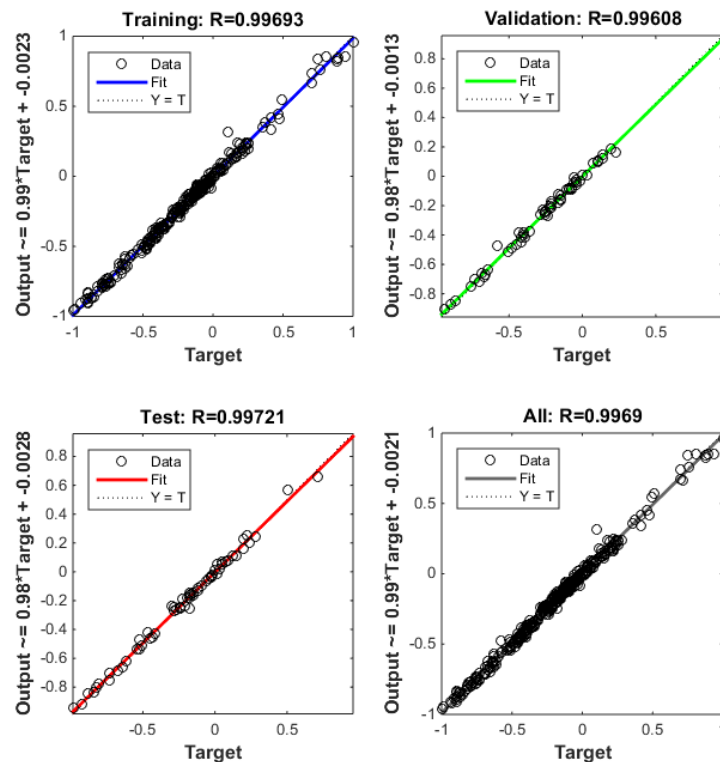

Fig.7 Regression Analysis based on BP Network algorithm

The BP neural network is used to predict the scores of the two factors in the first 30 working

trading days in 2019, and substituted into the regression fitting equation to obtain the standardized matrix of the highest price, the lowest price, the opening price and the closing price in the first 30 trading days in 2019. Referring to the comparison of the actual stock data, we can see that

the stock forecast price curve of Everbright Bank is highly consistent with the actual price curve, and there is only a small difference in the trend predicted on the 5th to 10th trading day, which is within the range of error acceptance.
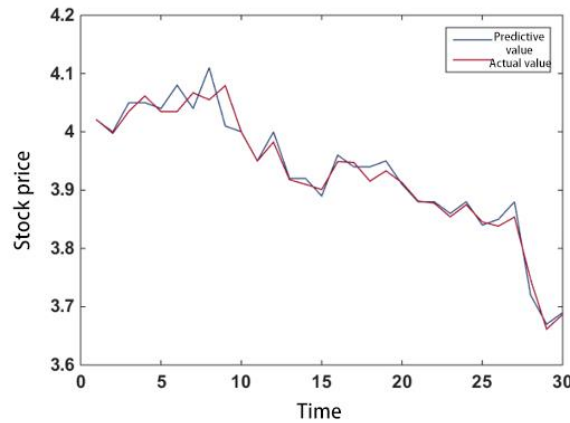

Fig. 8 Comparison between the predicted value and actual value

2.3.2 Error Analysis of BP Neural Network Prediction

To evaluate the prediction results of stock opening price by BP neural network more comprehensively and objectively, the average relative

error（ARE）and mean square error（MSE）are used to evaluate the prediction effect. The formula is as follows:

Average relative error formula:

$$ARE = \frac{1}{n} \left| \frac{\hat{y}_k - y_k}{y_k} \right| \quad (k = 1, 2, \cdots, N)$$

Mean square error:

$$MSE = \frac{1}{n} \sum_{K=1}^{m} (y_k - \hat{y}_k)^2 \quad (k = 1, 2, \cdots, N)$$

Where y_k represents the actual value of the valuation, $\hat{y}_k$ represents the predicted value of the valuation, and n represents the number of forecast samples. Comparing the data of the first 30 trading days predicted by the BP neural network with the real data, the line chart is as follows, and the average relative error（ARE）and mean square error（MSE）), are calculated to get an average relative error of 3.6% and a mean square error of 0.041. It can be proved that the regression prediction effect of a neural network is better, and the model is robust.

2.3.3 Error analysis of the stock description

Through the above regression results, Everbright Bank stocks can be described, but the regression fitting will have certain errors, so by calculating the determination coefficient of multiple non-linear regression, recorded as, to judge the effect of regression fitting.

The calculation of the determination coefficient. In multiple regression analysis, the determination coefficient is numerically equal to the square of the channel coefficient between the original data sequences, that is,

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where: define the sum of total deviation squares $SS_{tot} = \sum_i (Y_i - \bar{Y})^2$, Define the sum of total residual squares $SS_{reg} = \sum_i (f_i - \bar{Y})^2$; define the sum of regression squares $SS_{res} = \sum_i (Y_i - f_i)^2$; define f_i as the fitting value Y_i as the original data.

Realized by MATLAB software, the determination coefficient of multivariate non-linear regression fitting is 0.9172, and the average error is 7.7%. The larger the determination coefficient, the better the degree of regression fitting, and the error is smaller.[16] the regression equation is fitted to the original data image. According to the regression equation, taking the two description factors as the x-axis and y-axis variables and the corresponding stock price as z-axis variables, the regression surface is obtained by using MATLAB to visualize the data, and the real stock price data points are drawn, as shown in Fig.9.
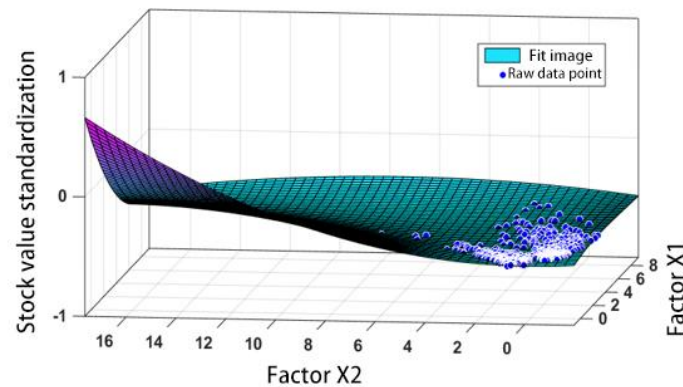

Fig. 9 regression fitting degree of the share price of Everbright Bank

It can be seen from Fig.9 that the original data points are basically on the regression surface. That is, the average value of the opening, closing, highest price and lowest price of Everbright Bank stock from 2010 to 2018 is similar to that of Everbright Bank stock price, and the regression fitting effect is good. Therefore, by observing the fit between the appropriate image and the original data points, we can directly analyze that the error of the regression equation is small.

3. Evaluation of the model

3.1 Model advantages

(1) The advantages of model neural network and regression analysis can not only reveal the entire conditional distribution of response variables but also simulate the non-linear characteristics of the financial system to meet the forecasting requirements of the stock market.

(2) The BP neural network can be trained to approximate any non-linear continuous function with arbitrary precision; that is, the BP neural network has strong non-linear mapping ability.

(3) BP neural network has a high degree of self-learning and self-adaptation ability, which can effectively predict such a volatile stock market.

3.2 Disadvantages of the model

(1) Because the BP learning rule adopts the method of decreasing the error function to converge, it cannot be guaranteed to converge to the optimal local solution. At this time, the actual output value and the expected output value of the network may produce a significant error.

（2）The BP neural network has a considerable sample dependence problem in prediction, which may cause deviations in the prediction results due to the particularity of the sample situation.

4. Conclusion

This article first describes the historical data of China Everbright Bank's stocks. It can be analyzed that China Everbright Bank's stock price market and the economic transaction rate factor of the stock market show a reverse change

trend. That is, during the period of low market transaction rate, China Everbright's stock price has rising expectations. Conversely, during periods of high transaction rates, Everbright Bank's stock price has expectations of falling. Moreover, at a time when the macro stock price has an inevitable upward trend, China Everbright Bank's stock price will also increase expectations. Furthermore, the BP neural network is used to train and predict the stock price. The analysis error shows that the overall error of the prediction model is small, the calculated average error is only 3.6%, and the mean square error is only 0.041. This can indicate that the regression prediction effect using the BP neural network is excellent. The predicted value is highly reliable.

References

1. Sha Wang. Research on the Application of BP Neural Network in Stock Forecast [D]. Central South University, 2008.

2. Murphy, John J. Intermarket Technical Analysis[J].

3. Deboeck G J. Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets[M]. 1994.

4. Massimiliano, Marcellino, And, et al. Macroeconomic forecasting in the Euro area: Country specific versus area-wide information[J].

5. Kimoto T, Asakawa K, Yoda M, et al. Stock market prediction system with modular neural network: Neural Networks, 1990., 1990 IJCNN International Joint Conference on, 1990[C].

6. Liting Gu Zhongxing Ye. Two neural network methods for the classification of stock market change patterns [J]. Journal of Shanghai Jiaotong University, 1995 (02): 100- 104.

7. Song Li, Lijun Liu, Chen Gu. Comparative study on prediction models of chaotic time series [J]. Computer engineering and application,2009,45 (32):53-56.

8. Kuihe Yang, Baoshu Wang, Lingling Zhao. Selection of input variables in prediction model based on neural network [J]. Computer science, 2003(08):139-140.

9. Dash R, Samal S, Dash R, et al. An integrated TOPSIS crow search based classifier ensemble: In application to stock index price movement prediction[J]. Applied Soft Computing Journal, 2019.

10. Jiang M, Liu J, Zhang L. An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms[J]. Physica A: Statistical Mechanics and its Applications, 2019.

11. Jiayu Q, Bin W, Changjun Z. Forecasting stock prices with long-short term memory neural network based on attention mechanism.[J]. PloS one, 2020,15(1).

12. Liu L, Pan Z. Forecasting stock market volatility: The role of technical variables[J]. Economic Modelling, 2020,84.

13. Hideki T, Reina H, Saori K, et al. [A factor analysis method for contingency table data with unlimited multiple choice questions].[J]. Shinrigaku kenkyu : The Japanese journal of psychology, 2016,86(6).

14. Islam B, Baharudin Z, Nallagownden P. Development of chaotically improved meta-heuristics and modified BP neural network-based model for electrical energy demand prediction in smart grid[J]. Neural Computing and Applications, 2017,28(1).

15. Zhiyuan M, Wei Z, Zhongbing L, et al. Ultrasonic characterization of thermal barrier coatings porosity through BP neural network optimizing Gaussian process regression algorithm.[J]. Ultrasonics, 2020,100.

16. Yingchao Xu, Xiangyou Wang, Xiang Yin, et al. prediction of potato processing quality based on multiple non-linear regression analysis [J]. Journal of Agricultural Machinery, 2018, 49 (04): 366, 373.