ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

Research on Stock Price Prediction of BP Neural Network Based on Factor Analysis

Biao Liu, Chenglin Xu

School of Mathematics, Yunnan Normal University, 650500, Kunming, China

Abstract: *In this thesis, the real estate industry is focused on, the financial data of 27 listed real estate companies in China is chosen, and the method of factor analysis is used to carry out the dimensionality reduction processing for the collected 20 variables. Finally, six main factors are extracted. Besides, the stock prices before and after the factor analysis are predicted and studied respectively based on the BP neural network prediction model. In line with the results, the better prediction accuracy for the BP neural network model established with the main factor determined by factor analysis as the independent variable and the stock price as the dependent variable is embodied. In contrast with the research carried out by other scholars, the BP neural network is established from more dimensions, so the prediction accuracy is greatly enhanced.*

Keywords: *factor analysis; BP neural network; stock price prediction*

1. Introduction

When it comes to the stock market, it has occupied a significant position in China's financial market system, and a country's economic situation can be reflected by the fluctuation of stock prices to a certain extent. While bringing us high returns, stocks also have high risks. For the purpose of striving for the largest returns and minimizing the existing risks, the investment analysis on stocks should be necessarily carried out. At present, EEMD method, time series analysis, machine learning and so on are the main methods for stock price prediction. In particular, Wu Yuxia and Wen Xin [1] carry out the short-term forecasting research on stocks on the basis of the ARIMA model, and the good model is shown from the results. Long Huidian [2] establishes the ARIMA model based on wavelet analysis has enhanced the prediction accuracy. What's more, Huang Hongyun, Wu Libin, and Liu Huachun [3] put forward an Elman neural network with dynamic feedback function, which is able to better predict stock prices. As all time series analysis methods, the above methods have high prediction accuracy, but the revealing of the internal relationship is difficult. After that, with the gradual emergence of BP neural networks, in which Jian Rongjie [4] sets up a BP neural network model through the extraction of key factors from the stock's own data characteristics; Ouyang Jinliang and Lu Liming [5] make the proposal of the combination of additional momentum method and dynamically adjusted learning rate method to improve the BP algorithm. Chen Yi [6] makes the optimization by adding an Adam algorithm based on BP neural network. Over the years, there are a large number of scholars having introduced wavelet neural networks to carry out the research on stock prices. According to the research of Chen Qiao [7], the wavelet analysis is found to be able to better identify the frequency change trend of stock index series. A combined prediction method based on wavelet decomposition network is proposed by Sun Bingjie, Tang Rui, Zuo Yi, Huang Ming, and so on[8], which significantly improves the prediction accuracy in contrast with that of traditional neural networks.

A certain predictive effect is played by the above models and algorithms on the future trend of stock price changes, but their shortcomings lie in the direct causal relationship between stock price changes and the performance of the listed company. The grasp of the internal causes of stock price changes is difficult based on the simple stock data analysis. In view of it, 16 financial indexes are chosen by Zhu Yongming and Shao Kangyun [9] to first extract principal components and then predict stock prices based on BP neural networks, in which the final prediction accuracy is about 81.2%. Although the improvement of the accuracy is realized, it fails to consider four important factors. In this thesis, the factor analysis is firstly carried out to reduce the dimensions of the selected 20 financial indexes, so the multiple collinearity between various variables is reduced. Then, a BP neural network model is established between the obtained six main factors and stock prices. Besides, the BP neural network model established by using 20 variables and stock prices is compared to break through the accuracy of

ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

the neural network.

2. Financial evaluation indexes of listed real estate companies

In this thesis, the financial data of 120 real estate listed companies from the WIND database is collected, and the data of the third quarter of 2022 is taken as the research object. The data is integrated to ultimately retain 27 real estate companies for research. By taking the basic characteristics of real estate enterprises into account, the following six aspects are divided for types of indexes: such as capital operation ability, profitability indexes, equity expansion ability, business operation ability, development ability indicators, and cash flow indicators.

3. The evaluation of financial indexes of enterprises based on factor analysis method

3.1 Data pre-processing

The data is sorted out, the duplicate data is deleted from the original data set, the missing and abnormal values are handled. The standardized processing for the data is further conducted to facilitate the analysis of the correlation between various indexes.

3.2 Factor analysis through the use of the principal component method

R language[10] is used to carry out the factor analysis [11] on the above 20 indexes, and the comparison results are shown in Table 1.

Table 1: Cumulative contribution rate of principal components

	Vars	Vars.Prop	Vars.Cum
Factor1	6.25	0.31251	31.25
Factor2	3.064	0.15322	46.57
Factor3	2.689	0.13443	60.02
Factor4	1.659	0.08293	68.31
Factor5	1.582	0.07912	76.22
Factor6	1.36	0.06801	83.02

The contribution rate of each principal component can be seen from the table above. Here, a case where the cumulative contribution rate is 83.02% is chosen, that is, these six principal components can represent 83.02% of the information amount for the original 20 financial indexes. Therefore, the extracted six factors can be indicated as the main factors influencing the above 20 indexes.

From the results of the factor analysis, the loads of total asset turnover, current asset turnover, and quick ratio on factor f_1 are found to be 0.81336, 0.91619, and 0.79566 respectively, which all reflect capital operation ability, so f_1 is named as the capital operation ability factor. The loads of total asset reward rate, main business profit rate, ratio of profits to cost and expense, and inventory turnover rate on factor f_2 are 0.64206, 0.89159, 0.88915, and 0.60829, respectively, which can reflect the profitability of the enterprise, so f_2 is named the profitability factor. The loads of asset liability ratio, net assets per share, capital reserve per share, and undistributed profit per share on factor f_3 are 0.32256, 0.94089, 0.57664, and 0.85945, respectively, which can reflect the equity expansion ability, so f_3 is named the equity expansion ability factor. The loads of the turnover rate of accounts receivable, current ratio, and growth rate of main business income on factor f_4 are 0.36805, 0.55890, and 0.78638 respectively, which can reflect the enterprise's operating capacity, so f_4 is named the enterprise's operating capacity factor. The loads of return on net assets, growth rate of net profit, and cash content of net profit on factor f_5 are 0.60431, 0.7625, and 0.83825 respectively, which can reflect the development ability of the enterprise, so f_5 is named as the development ability factor. The loads of total asset growth rate, net cash flow per share, and cash content of main business income

ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

on factor f_6 are 0.882784, 0.606436, and 0.777151 respectively, which can reflect cash flow, so f_6 is named as a cash flow factor.

3.3 Factor score

This is recorded as the weight of each factor in line with the proportion of the total variance explained by each factor, and the calculation formula for the comprehensive score can be obtained as follows:

$$F = 0.2080f_1 + 0.1736f_2 + 0.1254f_3 + 0.1103f_4 + 0.1076f_5 + 0.1052f_6$$

The weighted least square method is adopted to get the score of each stock on each factor [12] through calculation. In accordance with Table 1, it is found that the quick ratio, ratio of profits to cost and expense, growth rate of main business income, and cash flow of main business have high load values on f_1, f_2, f_4 and f_6. The factor coefficients of the comprehensive score formula are combined, these four independent variables are indicated to have a high impact on the stock score, especially the quick ratio of profits to cost and expense.

4. Establish a model to make the prediction and analysis

4.1 BP neural network

As an algorithm that imitates the structure of biological neural networks, BP neural network [13] is constituted by input layers, hidden layers, and output layers, in which the close connection of neurons between layers is shown, and each layer of neurons is assigned a weight. Two processes of forward propagation and error back propagation are mainly divided for its learning process[14].In the process of forward propagation, the passing of sample data is conducted from the input layer, a weight is randomly assigned to the neurons of each hidden layer, and then the passing of data is done through the hidden layer to the output layer. It will enter the error back propagation process if the error between the output value and the actual value is higher than the threshold. Error back propagation relies on a certain signal to make the transmission of output errors to hidden layers, and the errors are allocated to neurons in each layer, allowing each layer to adjust the weight of neurons to achieve the goal of reducing errors. This process is repeatedly conducted, the learning and training process [15]of the network is the adjustment process of neuron weights. The BP neural network terminates operation until the output error value decreases below the threshold value or the training times reach the preset goal, in which Figure 1 shows the concrete structure [16].

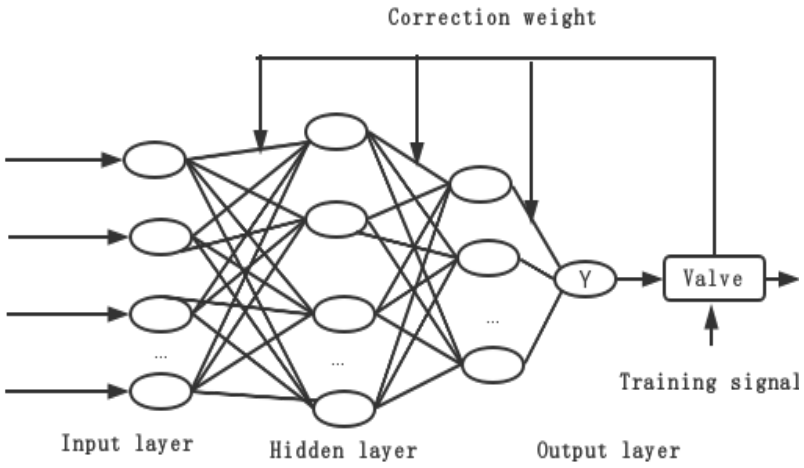

Figure 1: Structure of BP neural network

4.2 Data pre-processing

Before applying BP neural network to make the prediction of stock prices, the normalization processing for data is necessary[17]. The main purposes are as follows: Firstly, the errors caused by large

ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

differences in the order of magnitude between data can be eliminated. Secondly, the convergence speed of BP neural network can be accelerated to reduce learning and training time. The adopted normalization formula is: $\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$, in which $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the original data, x_i is the original data, $\hat{x}_i$ is the data after normalization. Since the output curve of Sigmoid function has the good properties of fast change frequency in the middle and flat at both ends, such characteristics are in line with the signal output form of neurons. Moreover, the saturated nonlinear characteristics and differentiability of S-type function greatly improve the nonlinear mapping ability of neural network. Therefore, Sigmoid- function $f(x) = \frac{1}{1+e^{-x}}$ is selected as the activation function of output layer, and double tangent S-type function $f(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$ is selected as the activation function of hidden layer function[18]

4.3 Experimental principle

(1) The input of the node in the input layer: X_i;

(2) The output of the hidden layer node: $y_i = f\left(\sum_i w_{ij} x_i + \delta_i\right)$;

(3) Output layer node output: $\phi_l = f\left(\sum_i u_{li} y_i + \delta_l\right)$;

(where, w_{ij} is the hidden layer connection weight, δ_i is the node threshold, and u_{li} is the output layer connection weight)

(4) Performance calculation: The error index of this period can be calculated after training for a cycle, and the mean square error is selected for replacement here:

$$E = \left(\sum_{q=1}^{Q} (E_q)^2\right)^{1/2} \Big/ Q, \text{ in which } E_q = \frac{1}{2} \cdot \sum_{i=1}^{n} (d_{ik} - Y_{ik})^2$$

If the error accuracy fails to satisfy the conditions, that is: $E > \varepsilon$, the back-propagation procedure will be activated. The operation will not be terminated until the training times reach the set value or the error is lower than the set target after repeated training.

4.4 Experimental process and conclusion

In this thesis, the realization of stock price prediction selects MATLABR2022a [[References:]], in which the training frequency is set to 1000 times, and the target accuracy is 0.001. The tan-sigmoid and purelin are used as two functions of the network hidden layer in the algorithm. Model 1: The establishment of BP neural network is carried out in accordance with 20 variables in the original data and the average value of the third quarter's stock price; Model 2: The establishment of BP neural network is conducted by taking the six main factors after factor analysis as independent variables and in line with the average value of the third quarter's stock price.

After carrying out the tests many times, the number of the optimal hidden layer nodes in Model 1 is determined to be 10, and the number of the optimal hidden layer nodes in Model 2 is determined to be 4. The error diagrams of the two models are shown in Figure 2 and Figure 3, and the comparison of various indexes is shown in Table 2.

ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

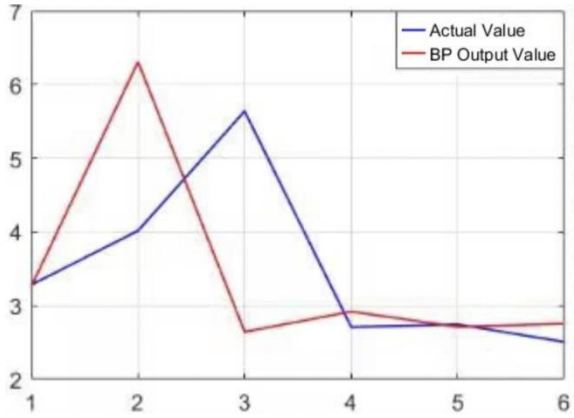

Figure 2: Prediction results of model 1

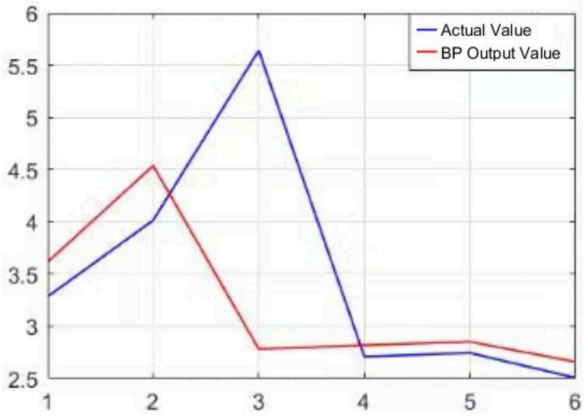

Figure 3: Prediction results of model 2

Table 2: Prediction results of two models

Name	Mean absolute deviation	Prediction accuracy
Model 1	1.0264	78.3907%
Model 2	0.6862	85.3670%

Through the contrast of the two models, the predicted accuracy of firstly carrying out factor analysis for variables and then taking advantage of BP neural network is much higher than that of directly using BP neural network to predict the variables. In the research carried out by scholars Zhu Yongming and Shao Kangyun, the quick ratio, patio of profits to cost and expense, growth rate of main business income and cash content of main business were not considered. Under the condition of not involving in these variables, the prediction accuracy of the model without conducting the extraction of the principal factors was 59.2% and the prediction accuracy of the model after the extraction of principal factors was 81.2%. Through the contrast of their conclusions, the accuracy obtained in this thesis has been improved by about 19% and 4% respectively, so the above four financial indexes are shown to bring a nonnegligible impact on the prediction of stock prices. These four indexes play a crucial role in measuring the comprehensive ability of an enterprise, which can also indirectly determine the trend of stock price under the influence of the psychology of stockholders[17].

5. Conclusion

At the time of using BP neural network to make the prediction of stock prices, the original variables should be processed through factor analysis to reduce the multicollinearity between variables, so the accuracy of the model can be greatly improved. Meanwhile, the multi-dimensional analysis should be conducted, in which the quick ratio, ratio of profits to cost and expense, growth rate of main business income and cash flow of main business should be considered by the stock price forecast model. The shortcomings of this thesis lie in the small sample data, causing the low accuracy of BP neural network. In the future work, more stock data can be collected for data analysis to enhance the accuracy and universality of the model.

ISSN 2616-5902 Vol. 5, Issue 10: 140-145, DOI: 10.25236/AJBM.2023.051021

References

[1] Wu Yuxia, Wen Xin. Short-term Stock Price Prediction Based on ARIMA Model [J]. Statistics & Decision, 2016 (23): 83-86
[2] Long Huidian, Yan Guangle. Research on GDP Time Series Prediction Based on SARIMA, GM (1,1) and Integrated Model of BP Neural Network [J]. Journal of Applied Statistics And Management, 2013,32 (05): 814-822
[3] Liu Huachun. Stock Market Decision-making Model Based on Elman Neural Network [J]. Computer Application, 2009, 29 (S1): 152-154
[4] Jian Rongjie. Stock Price Prediction Based on BP Neural Network and Correlation Coefficient [J]. China Journal of Commerce, 2017 (29): 24-25
[5] Ouyang Jinliang, Lu Liming. Application of Comprehensive Improved BP Neural Network Algorithm in Stock Price Prediction [J]. Computer & Digital Engineering, 2011, 39 (02): 57-59+97
[6] Chen Yi. Research on Stock Trend Prediction Based on Optimized BP Neural Network [J]. Science & Technology Information, 2020,18 (15): 198-200
[7] Chen Qiao, Liang Lili. Application of Wavelet Analysis in Stock Index Analysis [J]. Chinese & Foreign Entrepreneurs, 2015 (16): 45-47
[8] Sun Bingjie, Tang Rui, Zuo Yi, Huang Minghe. Research on Neural Network Stock Prediction Based on Wavelet Analysis [J]. Computer & Digital Engineering, 2016,44 (06): 1031-1034+1106
[9] Zhu Yongming, Shao Gengyun. Prediction of Stock Price Trend Based on BP Neural Network -By Taking Listed Companies of Real Estate Development as Examples [J]. Finance and Accounting Monthly, 2013 (14): 76-79
[10] Fei Yu, Guo Minzhi, Chen Yijuan. Multivariate Statistical Analysis -- Based on R. Beijing: China Renmin University Press, 2020
[11] Huang Xinjian, Li Ruoshan. Performance Evaluation Model of Real Estate Industry Based on Principal Component Analysis [J]. Statistics & Decision, 2006 (10): 163-164
[12] Li Yu, Zhang Li. Empirical Study on Financial Performance Evaluation of Real Estate Listed Companies—Based on Factor Analysis [J]. China Real Estate, 2020 (36): 35-39
[13] Shen Yuan. Application of BP Neural Network in Stock Price Prediction [J]. Knowledge Economy, 2010 (23): 40
[14] Xu Xingjun, Yan Gangfeng. Analysis of Stock Price Trend Based on BP Neural Network [J]. Zhejiang Fiance, 2011 (11): 57-59+64
[15] Yuan Bingqing, Cheng Gong, Zheng Liugang. Basic Principles of BP Neural Network [J]. Digital Communication World, 2018 (08): 28-29
[16] Zhao Mengchen, Huang Dongchao, Feng Yuhao, Hai Shiyang. Application of Artificial Neural Network in Stock Price Prediction [J]. Times Finance, 2012 (15): 242
[17] Jian Rongjie. Stock Price Prediction Based on BP Neural Network and Correlation Coefficient [J]. China Journal of Commerce, 2017 (29): 24-25
[18] Wang Yunxia. Based on Principal Components——Stock Prediction of BP Neural Network [C]//. Proceedings of 2013 the Fourth International Conference on Information,Communication and Education Application(ICEA 2013) Volume 31. [Publisher Unknown], 2013:395-401
[19] Lu Tianyu, Du Lana, Wang Haiyuan, Wang Yinqiu, Tao Mingwan, Zhang Xuewu. Prediction of Stock Price Change Trend Based on Principal Component Analysis and Neural Network Combination [J]. Computer Knowledge and Technology, 2019, 15 (06): 170-172+177
[20] Lv Junlin. Research on the Influence of Investor Sentiment and Momentum Effect on Stock Price Prediction [J]. Industrial Innovation Research, 2022 (19): 99-10