

Comparison Of Methods for Predicting Changes in Stock Trading Volume

Qinxuan Que *

Glorious Sun School of Business and Management, Donghua University, Shanghai, China, 200051

* Corresponding Author Email: 18912251117@163.com

Abstract. Stocks occupy an important place in the financial sector, while machine learning, which uses data to solve problems, has been applied in several fields. In this paper, three methods of multivariate linear model, BP neural network and ELMAN neural network are adopted to predict the transaction volume change value of ZTE and compare the evaluation. The study shows that, in the short-term prediction of BP neural network and multiple linear model, the two methods have good prediction effect, the BP neural network is stronger; ELMAN neural network is worse, it is not recommended to predict the trading volume change value.

Keywords: Stock Volume Value, Multivariate Linear Model, BP Neural Network, ELMAN Neural Network.

1. Introduction

Under the background of the rapid development of economic globalization, the stock was born, which is the most important financial instrument and financing method under the condition of modern commodity economy. The high risk and strong liquidity of stocks, which is beneficial to the development of market economy and national economy and is of far-reaching significance to the improvement of the national financial system. In recent years, the prediction of the stock market has become a hot research topic, and the rapid development of computer technology and artificial intelligence has opened up a new way for the prediction of the stock market.

Yang Yubin (2019) [1] In the article "Forecast and Analysis of Shanghai Composite Index based on Machine Learning Method", he used machine learning algorithm to predict the daily closing price of Shanghai A-shares, and further discussed whether the stock prediction of machine learning method is reasonable. Yao Siyu (2019) [2] in the model based on multiple fusion and financial text emotional securities trading signal prediction, the theoretical analysis and data analysis combining two methods, create two is used to predict the inflection point signal model, through the comparative analysis of model prediction results, discuss the model adaptability and prediction effect. Zhou Lianfang (2021) [3] Using BP-KV model, the financial data, asset value and volatility training data of listed companies to estimate the asset value and volatility of non-listed private enterprises, and finally derive the default distance and expected default rate of private enterprises.

A large number of studies have shown that deep learning algorithms have more advantages in predicting the stock market, with higher correct prediction results, while traditional machine learning is slightly inferior in these aspects.

Stock turnover is an important basis for judging the trend of individual stocks. There is a certain correlation between trading volume and stock price, the larger the trading volume, the higher the stock price, the smaller the trading volume, the lower the stock price. But in practice, investors often do not accurately grasp the market movements. This paper will select three methods: multivariate linear model, BP neural network and ELMAN neural network to predict the trading volume change value of ZTE in 10 trading days and compare and analyze.

2. Data source and variable selection

Various indicators of ZTE (000063) in a total of 300 trading days from 2020.07.13 to 2021.09.30 (daily opening price x1, daily highest price x2, daily lowest price x3, daily closing price x4, daily

transaction amount x_5 , daily trading volume x_6 , daily trading volume change value y) are selected as sample data (Data source: Shenzhen Credit Data Service Platform), where the change value of trading volume = (trading volume of the day - trading volume of the previous day)/trading volume of the previous day, positive number indicates the growth of trading volume compared with the previous day, negative number indicates the decline of trading volume compared with the previous day. The first 290 data of the sample data are used as the training set, and the last 10 data are used as the test set.

3. Prediction of stock trading volume change based on multiple linear model

3.1. Model introduction

Model definition: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon$, $x_1, x_2, \dots, x_p$ is a nonrandom independent variable, y is the random dependent variable, β_0 is a constant term, $\beta_1, \beta_2, \dots, \beta_p$ is the regression coefficient, ε is a random error term[4].

3.2. Model fitting and optimization

(1) The model was fitted, and the regression equation of full variables was established.

Model restoration:

$$y = 0.288825 - 0.010336x_1 - 0.236737x_2 + 0.146379x_3 + 0.106553x_4 + 0.015614x_5 - 0.004144x_6$$

Model evaluation:

The model of $p\text{-value} = 4.302e-05$, the significance level is less than 0.05, the null hypothesis is rejected, and the model is valid. The model of adjusted $R^2 = 0.08017$, close to 0, the model fitting effect is poor.

In the single regression coefficient test of the model, the regression coefficients of some variables are not significant, and the model does not achieve the most concise effect, so the model needs to be improved.

(2) Model optimization, stepwise regression

By using stepwise regression, x_1 , x_3 , x_4 , x_5 and x_6 are removed successively to make the model reach the simplest form, and AIC value is reduced to -308.01.

Model restoration:

$$y = -0.0235949 - 0.190411x_2 + 0.2048847x_3 - 0.0012303x_6$$

Model restoration:

The model of $p\text{-value} = 3.407e-06$, the significance level is less than 0.05, the null hypothesis is rejected, and the model is valid. The model of adjusted $R^2 = 0.08402$, close to 0, the model fitting effect is still poor.

In the single regression coefficient test of the model, compared with the model before optimization, the significance of x_2 and x_6 variables after optimization is enhanced.

3.3. Model prediction

The daily trading volume change value of ZTE from September 15, 2021, to September 30, 2021 was predicted by the equation before and after optimization respectively, and the prediction results were shown in Part 6. It is found that the prediction error of the equation before optimization (that is, the full variable regression equation) is smaller than that of the equation after optimization. Then, the predicted value of the equation before optimization is compared with the predicted value of the BP neural network and the ELMAN neural network.

4. Prediction of stock trading volume change based on BP neural network

4.1. Model introduction

BP neural network is a kind of common neural network model. Its self-learning ability is strong, and it can optimize the network structure and improve the prediction accuracy by adjusting the network parameters [5-7]. A simple BP neural network consists of three layers: input layer, hidden layer, and output layer (see Figure 1).

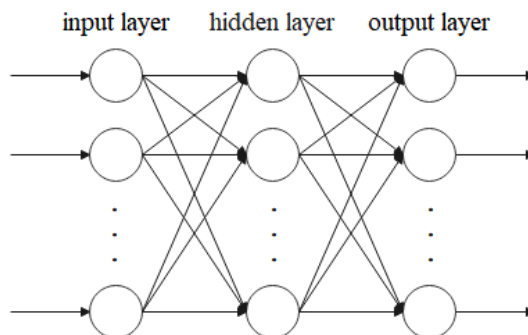


Figure 1. Structure model of BP neural network

4.2. Model building

(1) Divide training set and test set. The data of 300 samples were read, the training set was the first 290 samples, and the test set was the last 10 samples.

(2) Read data. The first six of the data are listed as characteristic indicators $x_1 \sim x_6$. The seventh column of the data outputs the index value y .

(3) Determine the number of neurons in the hidden layer. The number of hidden layer ganglion points l is determined according to the formula, which is:

$$l = \sqrt{m + n} + \alpha \quad (1)$$

m is the number of the input layer nodes, n is the number of the output layer nodes, α is generally taken as an integer between 1 and 10, by calculating the l as 3-13.

(4) The optimal BP neural network model is selected by means of mean square error (MSE) and goodness of fit (R-value) through training neural networks with different hidden layer ganglion points with MATLAB.

(5) The optimal BP neural network model is used to predict the change value of trading volume in the next 10 days.

4.3. Model prediction

The mean square error and goodness of fit corresponding to the number of nodes in different hidden layers are shown in Table 1.

Table 1 The mean square error and goodness of fit corresponding to the number of nodes in different hidden layers

	Number of hidden layer nodes										
	3	4	5	6	7	8	9	10	11	12	13
MS	0.278	0.413	0.389	0.443	0.375	0.396	0.386	0.329	0.162	0.241	0.345
E	8	2	7	2	7	3	0	4	3	5	5
R	0.408	0.352	0.245	0.339	0.275	0.339	0.368	0.414	0.558	0.415	0.303
	0	4	1	5	9	0	4	1	9	3	7

When the number of nodes in the hidden layer is 11, the mean square error is the smallest. The goodness of fit R value is closest to 1, indicating that this prediction model has better accuracy in describing experimental data.

The BP neural network model with the number of nodes in the hidden layer of 11 is used to forecast the daily trading volume change value of ZTE for ten days from September 15, 2021 to September 30, 2021, and the prediction results are shown in Part 6.

5. Prediction of stock trading volume change based on ELMAN neural network

5.1. Model introduction

The ELMAN neural network consists of four layers: input layer (signal effect), output layer, hidden layer and undertaking layer. As shown in Figure 2, each of the two adjacent layers is connected by adjustable weights. In general, it is considered a special type of feed forward neural network with additional memory neurons and local feedback[8-10].

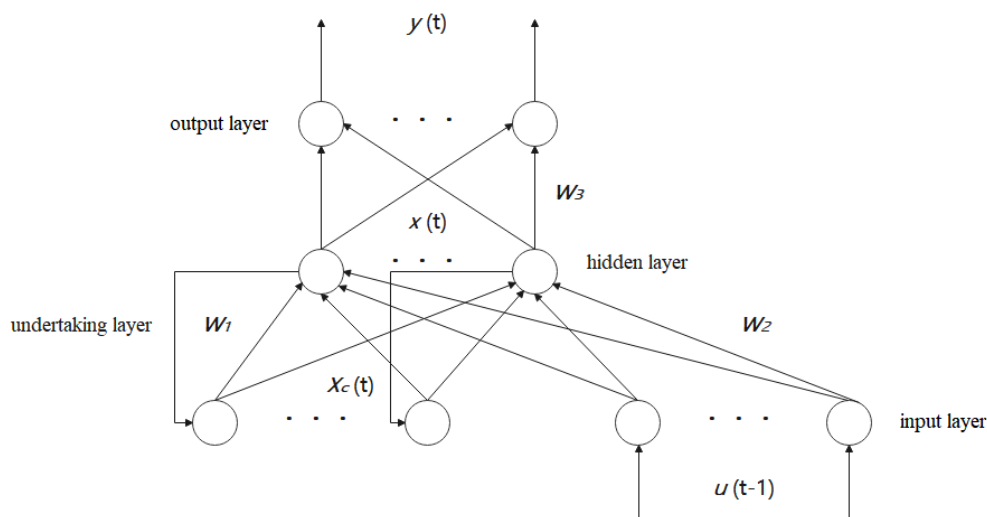


Figure 2. Structure model of ELMAN neural network

5.2. Model building

(1) Divide training set and test set. The data of 300 samples were read, the training set was the first 290 samples, and the test set was the last 10 samples.

(2) Read data. The first six of the data are listed as characteristic indicators $x_1 \sim x_6$. The seventh column of the data outputs the index value y .

(3) Data normalization. Normalization is an important process in building neural networks because it avoids the problem of algorithm convergence caused by unnecessary results or difficult training processes.

(4) Determine the number of neurons in the hidden layer.

(5) Select the optimal ELMAN neural network model.

(6) Forecast the change value of trading volume in the next 10 days through the optimal ELMAN neural network model.

5.3. Model prediction

According to Table 2, it can be seen that when the number of nodes in the hidden layer is 9, the mean square error MSE is the smallest, indicating that this prediction model has better accuracy in describing the experimental data.

Table 2 Means square error corresponding to the number of nodes in different hidden layers.

	Number of hidden layer nodes									
	3	4	5	6	7	8	9	10	11	12
MSE	0.0575	0.0552	0.0876	0.2563	0.1557	0.0546	0.0544	0.0545	0.0920	0.0568

This ELMAN neural network model is used to forecast the daily trading volume change value of ZTE for ten days from September 15, 2021, to September 30, 2021, and the forecast results are shown in Part 6.

6. Evaluation of model prediction effect

In order to visually compare the change value of ZTE's trading volume, the multiple linear model (MLM), BP neural network and ELMAN neural network predicted values of the 10 trading days from 2021.09.15 to 2021.09.30, Figure 3 is presented.

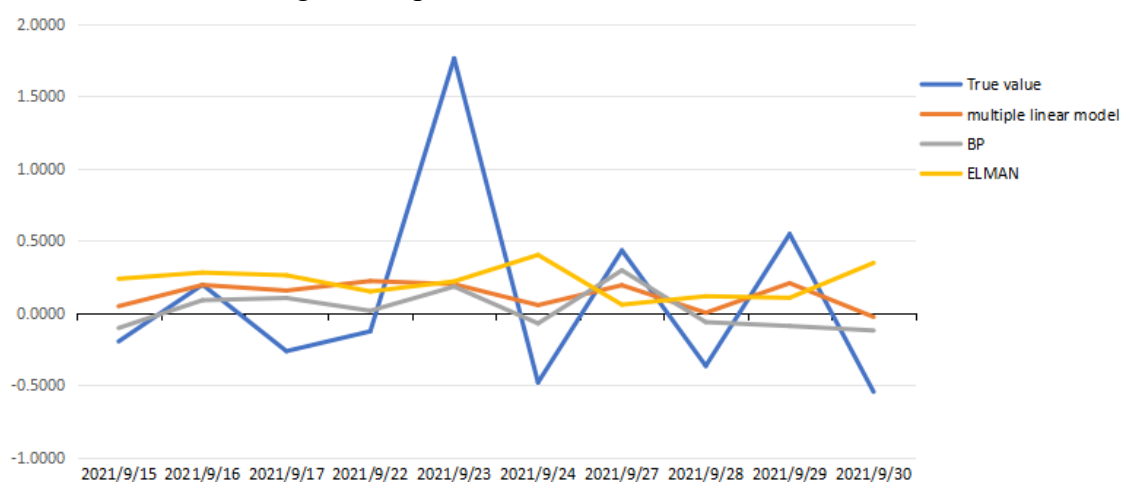


Figure 3. Forecast comparison of the change value of ZTE's trading volume

It can be seen that the prediction effect of the multiple linear model and BP neural network is slightly better, and the predicted value is roughly close to the real value, indicating that these two models can better predict the change value of stock trading volume. ELMAN neural network has a large gap with the real value at many points, and the deviation degree between the predicted value and the real value is relatively large, which indicates that ELMAN neural network has certain reference value for predicting the change value of stock trading volume.

Table 3 Comparison of trading volume changes and model errors

	True value	MLM Predicted value	BP Predicted value	ELMAN Predicted value	MLM AE	BP AE	ELMAN AE
2021/9/15	-0.1954	0.0480	-0.1026	0.2377	0.2434	0.0928	0.4331
2021/9/16	0.1966	0.1947	0.0896	0.2804	0.0019	0.1070	0.0838
2021/9/17	-0.2629	0.1561	0.1048	0.2617	0.4190	0.3677	0.5246
2021/9/22	-0.1257	0.2221	0.0164	0.1499	0.3478	0.1421	0.2756
2021/9/23	1.7614	0.2001	0.1837	0.2197	1.5613	1.5777	1.5417
2021/9/24	-0.4794	0.0555	-0.0709	0.4028	0.5349	0.4085	0.8822
2021/9/27	0.4336	0.1931	0.2960	0.0586	0.2405	0.1376	0.3750
2021/9/28	-0.3645	0.0025	-0.0622	0.1170	0.3670	0.3023	0.4815
2021/9/29	0.5470	0.2074	-0.0889	0.1062	0.3396	0.6359	0.4408
2021/9/30	-0.5437	-0.0260	-0.1198	0.3475	0.5177	0.4239	0.8912
MAPE					107.52	83.85	141.77
MAE					0.4573	0.4195	0.5929

According to Table 3, it can be seen that:

(1) The absolute error is used as a judgment index. Excluding 2021.9.23, because the data to be predicted are all less than 1, the absolute errors of the three models are not large, and they are all below 1. It can be seen that the multivariate linear model, BP neural network and ELMAN neural network can better predict the change value of stock trading volume.

(2) The average absolute percentage error as a judgment index. MAPE of BP neural network =83.85%, less than 100%, MAPE of multiple linear model =107.52%, and MAPE of ELMAN neural network =141.77%, all greater than 100%. The data shows that: BP neural network has the best effect in predicting the change value of ZTE's trading volume, while the multivariate linear model and ELMAN neural network have the worse effect.

(3) The average absolute error as a judgment index. The MAE of BP neural network is 0.4195, the MAE of multivariate linear model is 0.4573, and the MAE of ELMAN neural network is 0.5929. The data show that when predicting the change value of ZTE trading volume, BP neural network has the best prediction accuracy, while the multivariate linear model is average and the ELMAN neural network has the worst prediction accuracy.

In general, when predicting the change value of ZTE's trading volume, the prediction result of BP neural network is better than that of multivariate linear model, and the multivariate linear model is better than that of ELMAN neural network, and the prediction accuracy of the former two is slightly higher. BP neural network is recommended to predict the change value of stock trading volume, and ELMAN neural network is not recommended.

7. Conclusions and recommendations

This paper uses R language, MATLABR software, multivariate linear model, BP neural network and ELMAN neural network to predict the change value of ZTE's stock trading volume in 10 trading days. The results show that BP neural network and multiple linear model are recommended for short-term prediction of stock trading volume changes. BP neural network is better than other methods. The prediction effect of ELMAN neural network is poor, so it is not recommended to use this method to predict the change value of stock trading volume.

The research in this paper is still in the initial and experimental stage, and there is still a lot of work to be done to improve the prediction method. Areas for continued research and resolution in the future include:

(1) Expand the data set. The scale of data selected in this paper is relatively small, only 300 sample data. If the scale of data is expanded, not only short-term prediction can be made in the future, but also long-term prediction can be made, and the three methods may provide more accurate prediction results.

(2) Add variables. The characteristic indicators selected in this paper are of a single type and are all numerical indicators. In the future, macro factors and specific data can be combined to make the prediction effect better.

(3) Add prediction objects. In this paper, the three forecasting methods are only applied to one stock, not to multiple stocks. In the future, several stocks can be selected, and the three methods may present different advantages and disadvantages of the forecast.

References

- [1] Yang Y B. Prediction and analysis of Shanghai Composite Index based on machine learning method [D]. Zhongnan University of Economics and Law, 2019.
- [2] Yao Siyu. Prediction of securities trading signals based on multi-model fusion and financial text emotion [D]. Xiamen University, 2019.
- [3] Zhou Lian-Fang. Analysis of private enterprise bond default risk based on improved BP-KMV model [D]. Southeast university, 2022. DOI: 10.27014 /, dc nki. Gdnau. 2021.003646.

- [4] Li Xiaoning. Application of multiple linear regression and time series model in stock forecasting [J]. Science and Technology Entrepreneurship Monthly, 2019, 32(02): 153-155.
- [5] Zhao Hanlin. Research on stock price prediction method based on ARMI-BP neural network and sentiment analysis [D]. Shenyang university of technology, 2021. DOI: 10.27322 /, dc nki. Gsgyu. 2021.000864.
- [6] Zhang Rumeng, Zhang Huamei. Comparative analysis of BP neural network and ARMA-GARCH model in stock prediction [J]. Journal of Science, Normal University, 2021, 41(05): 14-20.
- [7] Lu Xiyan, Liang Qingmei. Prediction and Analysis of Stock Trends Based on Neural Networks [J]. Modern Business, 2019 (09): 89-90.
- [8] Yang Zhenhua. Research on stock prediction based on deep learning and text Mining [D]. Donghua university, 2020.
- [9] Wu Manman, Xu Jianxin. Short-term stock prediction model based on EMD improved Elman neural network [J]. Computer Engineering and Science, 2019, 41(06): 1119-1127.
- [10] Wang Jie Financial Physics Model and Financial Market Stochastic Prediction System [D]. Beijing Jiaotong University, 2018.