Stock Price Prediction Based on ARIMA-GARCH and LSTM

Xingdan Huang[1], Panlu You[2], Xiaolian Gao[2], and Dapeng Cheng[2](✉)

[1] School of Statistics, Shandong Technology and Business University, Yantai, China
[2] School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China
1259601112@qq.com

Abstract. Stock price prediction is a hot topic in the financial industry, and accurate stock price prediction is an important method to prevent risk and protect market stability. To this end, this paper constructs time series models and deep learning model, respectively, and compares the prediction results of the two types of models from the perspective of dynamic and static forecasting based on SSE index data. The results show that the forecasting methods of the models affect their forecasting effects, and the ARIMA-GARCH model has the highest average forecasting accuracy in static forecasting, while the LSTM model has the most accurate forecasting effect in dynamic forecasting, with an RMSE value of only 6.32%.

Keywords: static forecasting · dynamic forecasting · ARIMA · LSTM

1 Introduction

The essence of stock price trend prediction is to use the historical trading data information of stock market to predict the trend of stock price in the future time, and the current stock price prediction methods can be broadly divided into two categories, the first category is the time series model forecasting methods, based on statistical theory, by data mining the linear relationship between the series for prediction, the second category is deep learning and machine learning models, where deep learning models can not only handle large amounts of data but also have strong nonlinear processing and learning capabilities, which are more applicable than linear models.

Time series models include ARIMA [1] and GARCH [2]. As a classical time series model, the Autoregressive Integrated Moving Average (ARIMA) model performs very well in short-term forecasting of stock prices [3]. Deep learning models include LSTM [4], GRU [5], Transformer [6], etc., which have applications in stock forecasting. With the development of technology, scholars found that deep learning algorithms have good feasibility in stock prediction research, among which Long Short Term Memory (LSTM) model is the most prominent in time series data prediction, the results consistently show that LSTM models have high fit in forecasting [7–10]. In a comparative study of the two types of models, Zhang conducted a comparative study of the forecasting performance

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 438–448, 2023.
https://doi.org/10.2991/978-94-6463-198-2_45

of neural network models and ARIMA models and found that neural network models have higher forecasting accuracy than the traditional ARIMA models [11]. However, Majumder et al. used various methods including linear models and feedforward neural network models to forecast 35 stock data, and the results showed that the ARIMA models has the best average prediction accuracy [12].

Summarize the literature found that scholars in stock forecasting are using a prediction method to study, and did not consider the impact of the prediction method on the prediction results, this paper to further explore the effect of the model on different prediction methods, choose static forecasting and dynamic forecasting two prediction methods for forecasting. Static forecasting means one step forward forecasting, where the true value is chosen to make the next value of the forecast, while rolling forecasting is chosen to replace the true value with the fitted value for forward forecasting. The main contributions are as follows:

- Construction of ARIMA, ARIMA-GARCH and LSTM models, using both dynamic and static forecasting methods for forecasting.
- Comparing the forecasts of time series models and deep learning models under different forecasting methods, the results show that the forecasting methods affect the model forecasting accuracy.

2 Method

2.1 Principle of ARIMA Method

The ARIMA(p,d,q) model is essentially a non-stationary sequence transformed into a stationary sequence after a d-order difference, and finally an ARMA(p,q) model is built based on the stationary sequence. For a time series its ARMA(p,q) model equation is shown in (1), where the parameter c is a constant, the parameters $\varphi_{\mathrm{k}}(1 \leq k \leq p)$ and $\theta_{\mathrm{i}}(1 \leq i \leq q)$ are the coefficients of the autoregressive and moving average models respectively, ε_{t} for a white noise series.

$$\begin{aligned} u_t = c + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_p u_{t-p} \\ + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned} \tag{1}$$

2.2 Principle of GARCH (1,1) Method

Generalized autoregressive conditional heteroskedasticity is referred to as GARCH. GARCH model is a good solution to the heteroskedasticity of stock price series that fluctuate strongly at one time and less at another. In financial time series, scholars have concluded through a large number of empirical analyses that the best prediction is achieved at lag order p = q = 1, so the GARCH (1,1) model is chosen for subsequent predictions. The standard expressions for the GARCH (1,1) model are shown in (2) and (3).

$$y_t = x_t^T \gamma + \varepsilon_t \tag{2}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3}$$

In the above equation, where x_t is the vector of explanatory variables, γ is the vector of coefficients, ε_t is the disturbance term, ω is the constant term, and is the coefficient of the ARCH and GARCH terms, respectively, is the predicted variance of the previous period and σ_{t-1}^2 is the predicted variance of the previous period.

2.3 Principle of LSTM Method

LSTM is an improvement on recurrent neural network, which solves the gradient disappearance gradient explosion problem of recurrent neural network by adding a linear dependency between two states in a targeted way. A schematic of the LSTM network structure is shown in Fig. 1.

In the schematic diagram of the network structure, f_t, i_t and o_t represent the forgetting gate, the input gate, and the output gate, respectively. The three gating units and the state space are represented as follows (4–9).

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{c}_t = tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{7}$$

$$o_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_0) \tag{8}$$

$$h_t = o_t \cdot tanh(c_t) \tag{9}$$

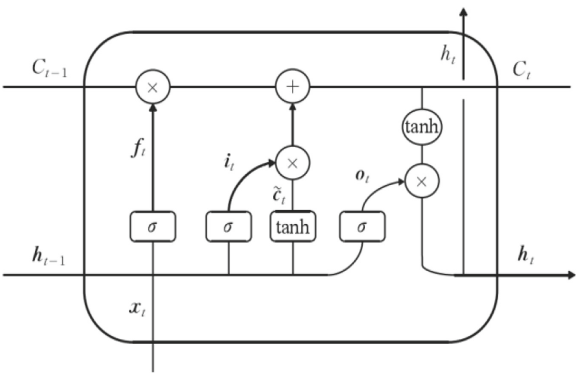

Fig. 1. LSTM network structure diagram

The forgetting gate is determined by the combination of the current moment input Xt, the previous moment output ht-1 and the previous moment state Ct-1 to determine which part of the information is discarded. The input gate determines which part of the information goes to the current moment state, and the output gate determines the final output of the neural unit.

3 Experiments

3.1 Experimental Data

The data are obtained from the China Stock Market Accounting Research (CSMAR). In this paper, a total of 2288 days of SSE Composite Index closing price data from January 4, 2010 to July 8, 2019 is selected as the modeling interval, the static forecast interval is 120 days of stock closing price data from July 9, 2019 to December 31, 2019, and the dynamic forecasting interval is the closing price of the stock for the three days from July 9, 2019 to July 11, 2019.

3.2 Stock Price Forecasting Based on ARIMA Model

1) **Data preprocessing**
   The original stock closing price time series $\{P_t\}$ was tested for stationarity and the p-value in the ADF test was $0.229159 > 0.05$, the original time series was non-stationary. Therefore, the series was smoothed, and the t-value of the ADF test after first-order differencing was -8.66420, which was much smaller than the critical values at 1%, 5%, and 10% significance levels, and the probability was 0.00, rejecting the original hypothesis that the time series was smooth after first-order differencing. The differenced time series was tested for white noise and the p-value was $0.00 < 0.05$, rejecting the original hypothesis, so the time series was considered non-white noise series at 95% confidence interval.
2) **Modeling analysis**
   The model order of the smooth time series is determined, and the order is determined by the ACF and PACF plots of the smooth series observed in Fig. 2 and Fig. 3, the final model order is determined by the BIC criterion for the p and q values under different values, and the BIC value is found to be the smallest at $p = 3$ and $q = 2$ after model comparison analysis, and the ARIMA(3,1,2) model is determined, and the model expression is as follows. From the above fitting results, the final equation of the ARIMA (3,1,2) model can be derived as

$$\begin{aligned} R_t = 0.235R_{t-1} - 0.9596R_{t-2} + 0.1046R_{t-3} \\ +\varepsilon_t - 0.1706\varepsilon_{t-1} + 0.8985\varepsilon_{t-2} \end{aligned}$$

In the above equation, $R_{\mathrm{t}} = P_t - P_{t-1}$ and $\varepsilon_{\mathrm{t}} \sim N(0, 1801)$

Using the function for white noise test, we get $p = 0.923 > 0.05$, and accept the original hypothesis, so the time series is considered as white noise series at 95% confidence

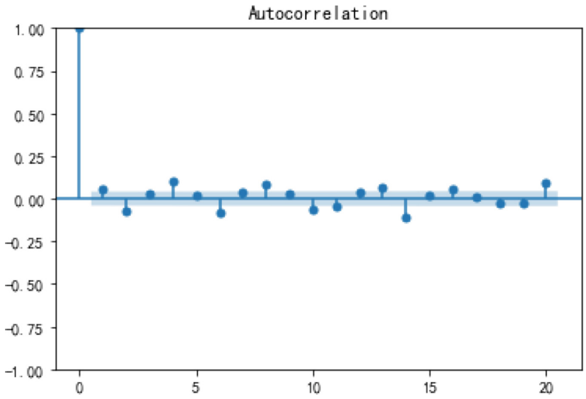

Fig. 2. Graph of ARIMA model autocorrelation results

interval, the ARIMA (3,1,2) model has the ability to extract useful information, and the parameters of the ARIMA (3,1,2) model passes the significance level test, the model can be used for forecasting.

3) **ARIMA (3,1,2) model prediction**
   The static forecasting results using the constructed ARIMA(3,1,2) model are shown in Fig. 4. The pictures show that the predicted and true stock price are very close and the model predicts roughly the same trend and the average error of the 120-day stock price is calculated with an RMSE value of 21.73042 and an MAE value of 17.21933.

ARIMA model dynamic forecasting results are shown in Table 1. In the table, T + 1, T + 2, T + 3 represent the closing prices of stocks on July 9, 2019, July 10, 2019 and July 11, 2019, respectively. It can be seen that the prediction results are closest on the first day, after which the error becomes larger.

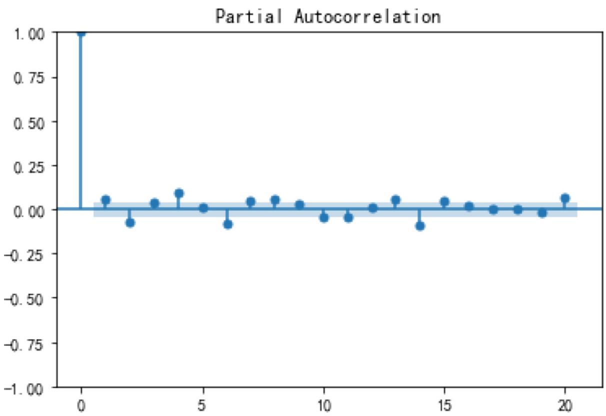

Fig. 3. Graph of ARIMA model partial autocorrelation results

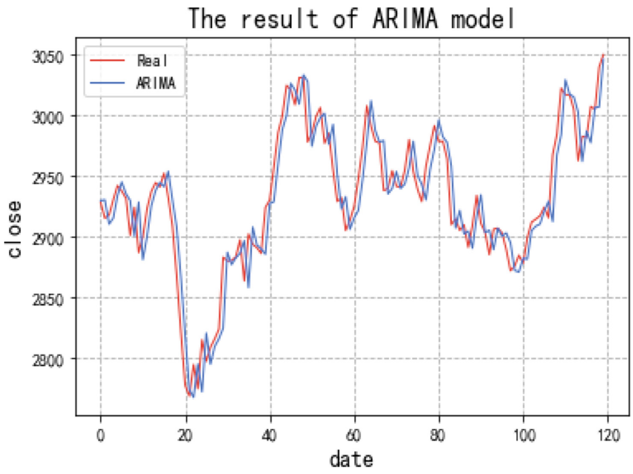

Fig. 4. Graph of ARIMA model prediction results

Table 1. ARIMA dynamic forecasting results

Value	DATE		
	T + 1	T + 2	T + 3
True	2928.23	2915.30	2917.76
Predicted	2929.705	2931.94	2927.9

3.3 Stock Price Forecasting Based on ARIMA-GARCH Model

1) **Build model**
The ARIMA (3,1,2) model fitted residual series have fluctuation aggregation, further ARCH effect test on the residual squared series $p = 0.00 < 0.05$, there is ARCH effect, which satisfies the GARCH modeling condition. Model comparison found the most accurate prediction under the GARCH (1,1)-GED distribution, and the specific results in combination with the ARIMA model are shown in Table 2.

Table 2. ARIMA-GARCH models of different orders

Models	all parameters significant	AIC
ARIMA(3,1,2) - GARCH(1,1)	No	9.724255
ARIMA(2,1,3) - GARCH(1,1)	No	9.723432
ARIMA(1,1,1) - GARCH(1,1)	**Yes**	**9.720953**
ARIMA(2,1,2) - GARCH(1,1)	No	9.723409

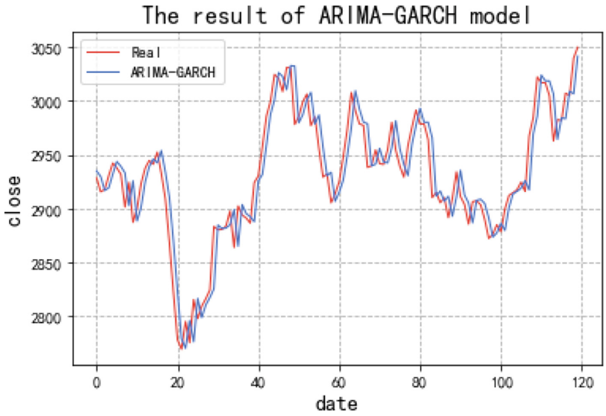

Fig. 5. Graph of prediction results of ARIMA-GARCH model

2) **Model Testing**
From the above table, we can find the best model of ARIMA (1,1,1)-GARCH(1,1), whose model expression is:

$$\begin{cases} R_{\mathrm{t}} = 143.3951 + 0.999965\ R_{t-1} + \varepsilon_t - 0.996755\varepsilon_{t-1} \\ \sigma_t^2 = 7.637566 + 0.053408\ \varepsilon_{t-1}^2 + 0.941288\sigma_{t-1}^2 \end{cases}$$

The sum of the coefficients of the ARCH term and the GARCH term in the variance equation is less than 1, which satisfies the constraints of the GARCH model. The coefficients are significant at 5% significance level except for the intercept term, and the ARCH effect test p-value = 0.2131 > 0.05, which accepts the original hypothesis that there is no ARCH effect in the series, indicating that the model is set correctly, and finally the model is applied to forecast stock prices.

3) **ARIMA (1,1,1)-GARCH(1,1) model prediction**
The static forecasting results of the ARIMA-GARCH model are shown in Fig. 5, Using the constructed model ARIMA (1,1,1)-GARCH (1,1) for the static forecasting, the mean RMSE value for the 120-day predicted data is 21.65515 and the mean MAE value is 16.98030.The prediction effect of the model has improved over the ARIMA model effect, and the difference between the predicted and true values in the graph is not significant.

The dynamic prediction results of the ARIMA-GARCH model are shown in Table 3. It can be found that the error of the model's prediction value on the second and third day becomes larger.

3.4 Stock Price Prediction Based on LSTM Model

The LSTM model in this paper is implemented in the Tensorflow 2.0 framework, and a total of 2288 days of data from January 4, 2010 to July 8, 2019 is selected as the training set in the prediction, and a total of 120 days of data from July 9, 2019 to December

Table 3. Prediction RESULTS OF ARIMA- GARCH MODEL

Value	DATE		
	T + 1	T + 2	T + 3
True Value	2928.23	2915.30	2917.76
Predicted value	2934.758	2936.158	2937.564

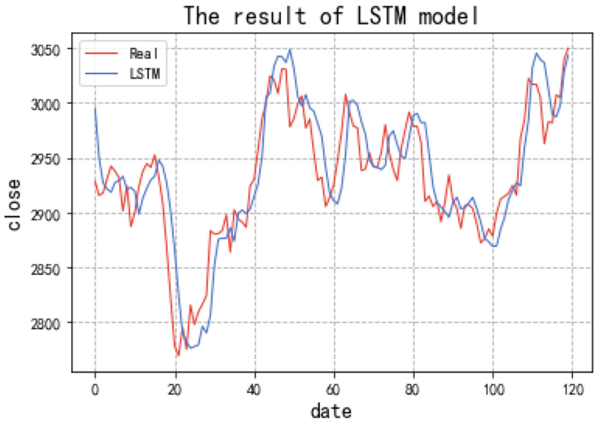

Fig. 6. Graph of prediction results of LSTM model

31, 2019 as the validation set. The model was selected to determine the parameters by trial-and-error method, and after several attempts the final parameters for optimal model performance were set as follows: the learning rate was set to 0.001, the Epoch value was 50, the BATCH_SIZE was 64, the number of hidden layers was 4, and the time step was 3, i.e., every three days of data predicted the closing price of the next day. The daily closing price is predicted with output_dim set to 1.

The data are normalized to eliminate the effect of dimensionality, the processed data are trained, and finally the trained model is used to predict the closing price of the stock, and the model predicts the following results.

Using the constructed LSTM model for static prediction, the results are shown in Fig. 6, with an average predicted RMSE value of 30.279924 and MAE value of 23.70677 for the 120-day data. It can be seen from the figure that the model does not fit well, there is a delay in the LSTM model prediction, and the model error is large.

The data were dynamically predicted for the next 3 days using the LSTM model, and the prediction results of the model are shown in Table 4.

3.5 Comparison and Analysis of Model Results

1) **Static forecasting**

The results of comparing the static forecasts of the model through the 120-day closing price forecasts are shown in Table 5. From the table, it can be seen that in the static

Table 4. LSTM dynamic forecasting results

Value	DATE		
	T + 1	T + 2	T + 3
True Value	2928.23	2915.30	2917.76
Predicted value	2929.0315	2920.254	2908.034

prediction, the ARIMA-GARCH model has the best average prediction accuracy, and its RMSE and MAE values are the lowest, and significantly lower than the deep learning model. Overall it seems that the average prediction accurathe time series model outperforms the average prediction accuracy of the deep learning model in the case of static prediction.

2) **Dynamic forecasting**
In the dynamic forecasting, direct future 3-day data based on historical data forecasts and fitted values. The results are shown in Table 6.

In dynamic prediction, the future values are unknown and need to fill the prediction by fitting the values, so the model is required to have a high learning fitting ability. The deep learning model has a strong learning ability and the prediction effect is better than other models, and it can be seen from the table that the LSTM model has the best fitting effect in the RMSE value and MAE value for dynamic prediction.

Table 5. Model static prediction results

Models	RMSE	MAE
LSTM	30.279924	23.706770
ARIMA-GARCH	**21.65515**	**16.98030**
ARIMA	21.73042	17.21933

Table 6. Model dynamic prediction results

Models	RMSE	MAE
LSTM	**6.319**	**5.160**
ARIMA-GARCH	11.286	9.422
ARIMA	17.029	15.730

4 Conclusion

This paper forecasts stock prices in terms of static forecasting and dynamic forecasting. In the time series forecasting model, ARIMA (3,1,2) model and ARIMA (1,1,1)-GARCH (1,1) model are constructed. In the deep learning prediction model, the LSTM model was constructed. The results show, in static prediction, the overall prediction accuracy of the time series model is higher than that of the deep learning model. However, in dynamic prediction, the deep learning model performs better. Ultimately, it was found that the prediction method of the model affects the prediction performance of the model.

Acknowledgment. General special subject of the "Fourteenth Five Year Plan" for education and teaching in Shandong Province (2021CYB012).

References

1. G. Caginalp, G. Constantine. Statistical inference and modelling of moment-um in stock prices[J]. Applied Mathematical Finance, 1995, 2(4).
2. Bicong Ji, Pinyi Zhang. Financial time series forecasting based on ARIMA-LSTM model[J].Statistics and Decision Making,2022:145–149.
3. Zhenghong Cha. Statistical analysis and forecast of SSE Composite Index[J]. Journal of Shanghai Maritime University, 1999(04):82-89.
4. Xiaoya XI, Heibin QIN, Zhijuan LU. Research on stock price change prediction based on LSTM neural network model--Baidu stock price as an example[J]. National Circulation Economy, 2022(16):102-105.
5. Feng Wu, Cong Xie, Chunmei Xie. Stock price prediction algorithm based on differential variance GRU gradient [J].Modern Computer ,2022,28(10):18–24+71.
6. Chen, S. L., Wang, X., Zhou, C. J.. Multi-factor stock forecasting based on GA-Transformer model[J]. Journal of Guangzhou University (Natural Science Edition),2021,20(01):44-55.
7. B.L. Jiang, J. Yue, S.J. Lan. Research on stock prediction system based on LSTM neural network[J]. Journal of Hebei College of Construction Engineering,2021,39(04):165-170.
8. Liang YJ, Song DF. Stock prediction based on LSTM and sentiment analysis[J]. Technology and Innovation, 2021(21):126-127.
9. I.W. Hu. Stock prediction based on optimized LSTM models [J]. Computer Science, 2021, 48(S1): 151-157.
10. Yang Can, Zhai Junjie, Tao Guihua. Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory[J]. Mathematical Problems in Engineering, 2020.
11. Zhang G. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003:159–175.
12. M. M. R. Majumder, M. I. Hossain and M. K. Hasan," Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), CoxsBazar , Bangladesh,2019 ,pp. l- 5