

Research on Combination Prediction of Shanghai Composite Index Based on IOWGA Operator

Lu Xinwen

Institute of Statistics and Applied Mathematics, Anhui University of Finance & Economics, Bengbu, 233030, China

Keywords: Arima, Lstm, Iowga Operator

Abstract: In this paper, the ARIMA(2,1,2) model and LSTM model are used to predict the Shanghai Composite Index, and then the IOWGA operator is used to establish a combined forecast model to further improve the model forecast accuracy. The empirical results show that the errors of the combined forecasting model are lower than those of the individual forecasting models, and the average precision is better than that of the individual forecasting models. In addition, the forecast results show that the Shanghai Composite Index will fluctuate slightly in the next five trading days.

1. Introduction

1.1 Research Background and Significance

With the continuous development of the market economy and the vigorous development of the domestic financial market, the stock market has become another important channel for the public to invest and manage money. However, stock prices are highly unstable due to the influence of various independent variables. Therefore, applying modern scientific methods and technical tools to predict the trend of stocks can reduce investment risks and obtain certain considerable returns.

In addition, the sudden outbreak of the new crown pneumonia epidemic in early 2020 has impacted the global economy. In particular, since the end of 2021, the Shanghai Composite Index has continued to fluctuate and fall, falling below 3,000 points all the way, setting another new low in 2020. As a multi-independent variable time series data, there are currently many methods for stock index forecasting. Although the single-item forecasting method explains the law of stock price changes to a certain extent, the forecasting effect and accuracy are average. Therefore, in order to take advantage of the characteristics of various forecasting methods, many scholars have proposed to combine various individual forecasting methods appropriately to further improve the accuracy and forecasting effect of the forecasting model. In this paper, the ARIMA and LSTM neural network models are selected, and the IOWGA operator is established to construct a combined forecasting model.

1.2 Literature Review

Previous studies on stock index forecasting focused on the use of single forecasting models, such as ARIMA, neural network, SVM, etc. Nowadays, many scholars focus on combinatorial

optimization of one-way forecasting^{[1][2]}. Liu Jiaming and Liu Haibin (2014) established a combined forecasting model of ARIMA and BP neural network, respectively fitting the linear and nonlinear residual parts of the time series, and then integrated them to improve the forecasting accuracy of the model^[3]. Cheng Changpin et al. (2012) decomposed the time series data with wavelet decomposition algorithm, separated the low-frequency and high-frequency information in the non-stationary sequence, used ARIMA and SVM model fitting respectively, and then superimposed the model prediction results to obtain the original sequence prediction value^[4]. Wu Dashuo (2020) proposed to improve the LSTM model stock index prediction based on genetic algorithm, and compared its prediction accuracy with BP neural network and LSTM network model through the demonstration of Nasdaq data^[5].

Combing the relevant literature, scholars have various research methods on stock index forecasting. The common methods include gray forecasting, ARIMA model, BP neural network, SVM model, etc. In this paper, two common single-item forecasting methods, ARIMA and LSTM, are used. The IOWGA operator establishes a combined forecasting model, and empirically tests the effectiveness of the combined forecasting.

2. Model Introduction

2.1 IOWGA Operator

Assuming n two-dimensional arrays $(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots, \langle v_n, a_n \rangle)$ and $W = (w_1, w_2, \dots, w_n)$ be the induced ordered weighted geometric mean operator, that is, the weight vector associated with the IOWGA operator, and W satisfies $\sum_{i=1}^n w_i = 1$. Let $IOWGA_w(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots, \langle v_n, a_n \rangle) = \prod_{i=1}^n a_{v-index(i)}^{w_i}$,

where $v_1, v_2, v_3, \dots, v_n$ is the induced variable, $\prod_{i=1}^n a_{v-index(i)}^{w_i}$ is the subscript corresponding to the i -th largest number of $v_1, v_2, v_3, \dots, v_n$ arranged in descending order, and $IOWGA_w$ is the n -dimensional induced ordered weighted geometric mean operator, that is, the n -dimensional IOWGA operator.

2.2 Prediction Accuracy

For the prediction of the same economic problem, assuming that there are n kinds of prediction methods, x_{it} is the true value ($i = 1, 2, \dots, n; t = 1, 2, \dots, N$) of the i -th prediction method at time t , and z_{it} represents the accuracy of the i -th prediction method at time t , $z_{it} \in [0, 1]$:

$$z_{it} = \begin{cases} 1 - |(x_t - x_{it}) / x_t|, & |(x_t - x_{it}) / x_t| < 1 \\ 0, & |(x_t - x_{it}) / x_t| \geq 1 \end{cases} \quad (1)$$

2.3 IOWGA Combination Model

According to the prediction accuracy and prediction value of n prediction methods at time t , n two-dimensional arrays are constructed: $(\langle z_{1t}, x_{1t} \rangle, \langle z_{2t}, x_{2t} \rangle, \dots, \langle z_{nt}, x_{nt} \rangle)$. n kinds of prediction methods at time t are ordered by prediction accuracy in descending order, and $z-index(i)$ represents the i -th largest prediction accuracy subscript at time t , let:

$$IOWGA_w(\langle z_{1t}, x_{1t} \rangle, \langle z_{2t}, x_{2t} \rangle, \dots, \langle z_{nt}, x_{nt} \rangle) = \prod_{i=1}^n x_{z-index(it)}^{w_i} \quad (2)$$

Let $e_{z-index(it)} = \ln x_t - \ln x_{p-index(it)}$, then the total combination forecast logarithmic error square sum S of N period is:

$$\begin{aligned} S &= \sum_{t=1}^N (\ln x_t - \ln \prod_{i=1}^n x_{p-index(it)}^{w_i})^2 \\ &= \sum_{t=1}^N (\ln x_t - \sum_{i=1}^n w_i \ln x_{p-index(it)})^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j (\sum_{t=1}^N e_{z-index(it)} e_{z-index(jt)}) \end{aligned} \quad (3)$$

Therefore, the combined prediction model based on IOWGA based on the sum of squared logarithmic errors can be expressed as the following model:

$$\begin{aligned} \min S(W) &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j (\sum_{t=1}^N e_{z-index(it)} e_{z-index(jt)}) \\ s.t. &\begin{cases} \sum_{i=1}^n w_i = 1, \\ w_i \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (4)$$

Let $E_{ij} = \sum_{t=1}^N e_{z-index(it)} e_{z-index(jt)}$, $i, j = 1, 2, \dots, n$, then $E = (E_{ij})_{n \times n}$ is called the combined prediction logarithmic error information square matrix of n -order IOWGA, so the model can be expressed in matrix form:

$$\begin{aligned} \min S(W) &= W^T E W \\ s.t. &\begin{cases} \sum_{i=1}^n w_i = 1 \\ W \geq 0 \end{cases} \end{aligned} \quad (5)$$

2.4 Error Evaluation Index

RMSE, SSE, MAE, MRE, and MAP can all evaluate the prediction deviation.

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (observed_t - predicted_t)^2} \quad (6)$$

$$SSE = \sum_{t=1}^N (x_t - \hat{x}_t)^2 \quad (7)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t| \quad (8)$$

$$MRE = \frac{1}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (9)$$

$$MAP = \frac{1}{N} \sum_{t=1}^N \left(1 - \left| \frac{x_t - \hat{x}_t}{x_t} \right| \right) \quad (10)$$

3. Shanghai Composite Index Forecast

3.1 Variable Selection and Data Sources

This paper selects the time-series data of the closing price of the Shanghai Composite Index from 2005/1/4 to 2022/5/31 as the research object, and initially observes its trend chart.

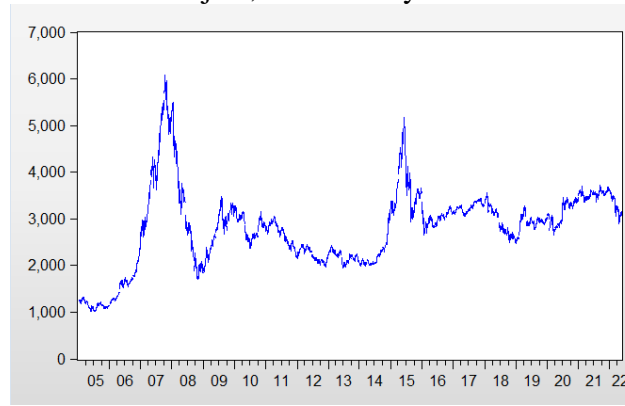


Figure 1: Sequence chart of Shanghai Composite Index from 2005/1/4 to 2022/5/31

3.2 ARIMA (p,d,q) Prediction Model

ARIMA differential autoregressive moving average model is a combination of autoregressive model, moving average model and difference method. Among them, p is the number of autoregressive items, q is the number of moving average items, and d is the order of difference made to make the data a stationary sequence. Its model can be expressed as:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (11)$$

It can be seen from Figure 1 that the time-series data of the Shanghai Composite Index is non-stationary, and the P value of the ADF test is 0.1118 higher than 0.05, which further indicates that the time-series data is non-stationary, and it can be transformed into a stationary sequence by d times difference and ADF test.

The P value of the ADF test result after the first-order difference is less than 0.05, which indicates that a stationary sequence can be obtained after the first-order difference, and the Q statistic with the probability $P < 0.05$ can judge that the first-order difference sequence is non-white noise. The autoregressive and moving average orders were preliminarily judged from the ACF and PACF correlograms, and the p value was determined to be 2 and the q value to be 2 by AIC and BIC.

According to the number of items obtained and the order of difference, the ARIMA (2,1,2) model is constructed, and the results are obtained and judged: the t values of AR(1), AR(2), MA(1) and MA(2) are respectively -2.201728, -21.19044, 2.058227 and 17.24893, the t test passed; the DW value was 1.936351, indicating that the model does not have serial autocorrelation; at the same time,

the model passed the White test, indicating that the model does not have heteroscedasticity. Through the above tests, it shows that ARIMA (2, 1, 2) is suitable for the prediction of Shanghai Composite Index in recent years. The estimated form of its model is:

$$\Delta Y_t = -0.086043\Delta Y_{t-1} - 0.786569\Delta Y_{t-2} + 0.091878\varepsilon_{t-1} + 0.726932\varepsilon_{t-2} \quad (12)$$

So it can be deduced that the prediction formula of the ARIMA (2, 1, 2) model is:

$$Y_t = Y_{t-1} - 0.086043\Delta Y_{t-1} - 0.786569\Delta Y_{t-2} + 0.091878\varepsilon_{t-1} + 0.726932\varepsilon_{t-2} \quad (13)$$

According to the model (13), the data in the sample period can be predicted and compared, and the specific results are shown in Figure 2.

3.3 LSTM neural network

LSTM is a kind of recurrent neural network, which is widely used in sequence forecasting, such as weather forecasting, stock market forecasting and so on. It mainly consists of five different parts: unit hidden, hidden state, input gate, forget gate and output gate.

In this paper, the LSTM neural network is used to predict the Shanghai Composite Index. Using Python software, a total of 4179 trading days from 2005/1/4 to 2022/3/15 are used as training samples. The forget gate of the model uses the sigmoid function to selectively pass through the variable information. The input gate uses the combination of the sigmoid layer and the tanh layer to process and update the data. The output gate uses the sigmoid function to determine the part that needs to be output, and after processing with the tanh layer. The data are multiplied to determine the output information. In terms of model training, the loss function is MSE, the optimization algorithm is Adam, the model training batch data is 1, the number of training is 100, and the number of neuron layers is 8. The final prediction results are shown in Table 1.

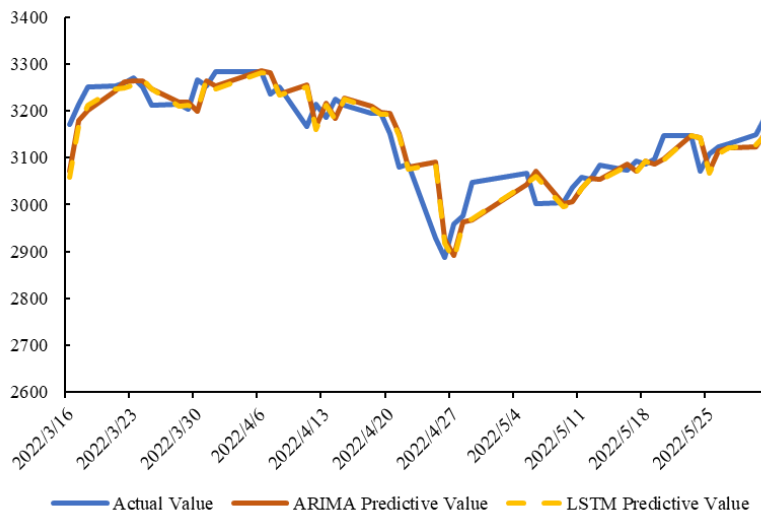


Figure 2: ARIMA and LSTM prediction model accuracy comparison

3.4 IOWGA combined forecasting model

3.4.1 Model Construction

Taking the prediction accuracy of the above two individual prediction methods as the induced value, the induced value and the predicted value are reordered and sorted to obtain the error information matrix E. Then the optimal solution of model (5) is solved to obtain the optimal weight

vector, and the IOWGA combined prediction model is constructed.

The predicted value and accuracy are sorted according to the accuracy from small to large, and the logarithmic error information matrix E of the model is obtained:

$$E = \begin{pmatrix} 0.011605 & 0.010639 \\ 0.010639 & 0.009832 \end{pmatrix} \quad (14)$$

Using the logarithmic error information matrix E, the sum prediction model of the IOWGA operator is constructed based on the minimum sum of logarithmic error squares and S: the minimum value of S is 0.0098.

$$\begin{aligned} \min S &= 0.011605\omega_1^2 + 0.021279\omega_1\omega_2 + 0.009832\omega_2^2 \\ \text{s.t.} &\begin{cases} \omega_1 + \omega_2 = 1 \\ \omega_1 \geq 0, \omega_2 \geq 0 \end{cases} \end{aligned} \quad (15)$$

The optimal solution is obtained through Matlab: $w = (w_1, w_2) = (0.0002, 0.9998)$, and the logarithmic error square sum $\min S = 0.0098$, which means that the weight $w_1 = 0.0002$ is assigned to the predicted value with the lowest accuracy of the two individual predictions, and the weight $w_2 = 0.9998$ is assigned to the predicted value with the highest accuracy, and the weight of the two predicted values is taken The geometric mean is used as the combined forecast over the sample period. The combined model constructed is:

$$\hat{x} = 0.0002x_{\alpha-index(1t)} + 0.9998x_{\alpha-index(2t)} \quad (16)$$

According to the model (16), the prediction value and prediction accuracy of the combination prediction model are obtained, see Figure 3 for details.

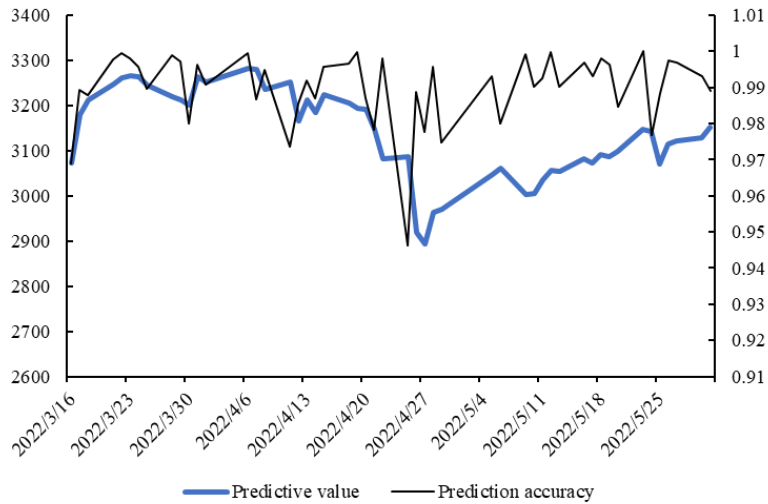


Figure 3: Predicted values of IOWGA combined forecasting model

3.4.2 Model Evaluation

Taking *RMSE*, *SSE*, *MAE*, *MRE* and *MAP* as error metrics, the forecasting effects of each forecasting model and the combined forecasting model of IOWGA are calculated and compared. The specific results are shown in Table 1.

Table 1: Comparison of model errors

Error Metric	ARIMA Model	LSTM Model	Combined Forecasting Model
<i>RMSE</i>	45.0711	45.34203	43.28617
<i>SSE</i>	101570.218	102794.965	101570.155
<i>MAE</i>	32.5773179	33.4809052	32.57634
<i>MRE</i>	0.010476522	0.010756872	0.01046
<i>MAP</i>	0.9895235	0.989243128	0.98005

According to Table 1, the error measurement index values of the combined forecasting model based on the IOWGA operator are lower than those of the single forecasting model, and the forecasting effect of the combined forecasting model is better. Its average prediction accuracy is as high as 0.9901, which is higher than the average prediction accuracy of the other two items. Therefore, the establishment of a combined prediction model based on the IOWGA operator is conducive to further improving the prediction accuracy of the Shanghai Composite Index.

3.4.3 Prediction Results

Since there is no real data for the next 5 trading days, it is impossible to compare the forecasting accuracy of the individual forecasting models, and thus it is impossible to sort them out. Referring to the relevant literature, according to the weight values assigned by each individual forecast model in the combined forecast model on each trading day, the weights of the ARIMA model and the LSTM model can be obtained by resummation and arithmetic mean. The combined forecast model based on the IOWGA operator can be obtained. According to the obtained model, the forecast value of the Shanghai Composite Index for the next 5 trading days is calculated. The forecast results are shown in Table 2.

Table 2: Predicted values of each model for the next 5 trading days

Time	ARIMAModel	LSTM Model	Combined Forecasting Model
2022/6/1	3189.749769	3184.865069	3187.307419
2022/6/2	3185.503598	3184.600207	3185.051902
2022/6/6	3183.255365	3180.809963	3182.032664
2022/6/7	3186.788717	3179.088376	3182.938547
2022/6/8	3188.253087	3176.465226	3182.359156

4. Conclusion

In this paper, the ARIMA (2, 1, 2) model and LSTM model are used to predict the Shanghai Composite Index, and then the combined forecasting model is used to predict the stock index. The results show that due to the limitations of the model itself, the forecast accuracy of the single forecast model needs to be improved, while the forecast accuracy of the combined forecast model based on the IOWGA operator is 0.9901, and the forecast error is lower than that of each single forecast.

In addition, the forecast value of the Shanghai Composite Index in the next five trading days is obtained through the combined forecasting model (see Table 2). The results show that the Shanghai Composite Index will fluctuate slightly and decline in the next five trading days. The reasons for this phenomenon may be the sluggish economic market environment, company operating conditions and policy factors, etc. Overall, the stock market may be in a short market state.

Since only two single-item forecasting models are selected in this paper, the stock forecast is affected by multiple independent variables. There are linear and nonlinear parts in the time series of the Shanghai Composite Index, which cannot be optimized separately. There is room for

improvement in the model, and the forecasting accuracy can be further improved.

References

- [1] Yao Jinhai. *Research on Stock Index Forecasting Based on the Combination of ARIMA and Information Granularized SVR* [J]. *Operational Research and Management*, 2022, 31(05):214-220.
- [2] Liu Ming, Shan Yuying. *Prediction of stock index closing price based on EMD-LSTM model* [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2021, 35(12):269-276.
- [3] Liu Jiaming, Liu Haibin. *The Application of Combined Forecasting Model in Short-term Stock Index Forecasting—An Empirical Analysis of Shanghai Composite Index as an Example* [J]. *China Business*, 2014(05):93-96.
- [4] Cheng Changpin, Chen Qiang, Jiang Yongsheng. *Stock Price Forecast Based on ARIMA-SVM Combination Model* [J]. *Computer Simulation*, 2012, 29(06):343-346.
- [5] Wu Dashuo, Zhang Chuanlei, Chen Jia, Xiang Qihuai. *Improved LSTM Neural Network Stock Index Prediction Analysis Based on Genetic Algorithm* [J]. *Computer Application Research*, 2020, 37(S1):86-87+107.