International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 04 - Issue 07 || July 2018 || PP. 59-64

Empirical Study on Short-Term Prediction of Shanghai Composite Index Based on ARMA Model

Cai-xia Xiang[1], Ping Xiao[2*]
[1](*School of Hunan University of Humanities, Science and Technology, Hunan417000, China*)
[2](*School of Hunan University of Humanities, Science and Technology, Hunan417000, China*)
**Corresponding author. applefly13@126.com*

Abstract: This paper estimates the comprehensive forecasting model from the closing price of 370 Shanghai Composite Index from January 3, 2017 to July 10, 2018. Firstly, through the preliminary analysis of the data, the unsteady raw data is transformed to establish a new stationary time series, so as to establish the ARMA model to predict the closing price of the Shanghai Stock Exchange in the next three days. Since the Shanghai Composite Index is one of the most representative indexes in China, it can fully reflect the development trend of China's stock market to a certain extent. Therefore, it is practical significance to study the short-term changes in the daily closing price of the Shanghai Composite Index, understand the changes in the stock market, make investment decisions, and provide reliable information services and decision-making guidance for investors and decision makers.
Keywords: Time series; ARMA model; short-term forecast; Shanghai index

I. Introduction

There are many time series problems in real life, such as the impact of bank interest rate fluctuations, the trend of stock returns, and the changes in international exchange rates. The basic idea of the time series prediction method is to use the past behavior to predict the future when predicting the future change of a phenomenon, that is, to reveal the phenomenon of time variation through historical data of time series, and extend it, then make predictions for the phenomenon of the future. However, the ARMA model is the most commonly used model for fitting stable sequences and has important theoretical implications in the financial and equity sectors. It uses a finite parameter linear model to describe the autocorrelation structure of the time series, which is convenient for statistical analysis and mathematical processing.

Time series analysis has been widely used in many fields. For the application of ARMA model in time series, Liang Yan and XiaLetian[1] (2012) has done a corresponding research. FengPan and Cao Xianbing[2] (2011) in the "Empirical Study of Stock Price Analysis and Prediction Based on ARMA Model", first use the unit root test to determine the stability of original time series, if not, use the difference method to make it Smooth and pass the ADF test, then build the ARMA model and use guessing ideas to determine the values of p and q. Then ensure that the parameters of the model are significant. Finally, the rationality of the model is determined by the residual test, and the stock price is predicted by the model. By comparing the actual value with the predicted value, it can be seen that the model is more reasonable. Wu Chaoyang[3] (2010) combines the ARMA model with the gray model to predict the stock index, and improves the defect that the gray model GM(1,1) is not optimal. The improved gray model and ARMA model are obtained, and the prediction accuracy is obtained. MengKun, Li Li[4] (2016) "Empirical analysis of stock price based on ARMA model" better solved the modeling problem of non-stationary time series. Cong Zheng, Xu Jiaping[5] (2013) used the ARMA model to calculate the rural financial gap in Liaoning Province and solved the practical problems.

In this paper, the ARMA model is selected to predict the Shanghai Stock Index. Since the ARMA model is used to process stationary sequences, the nonlinear and non-stationary characteristics of the Shanghai Stock Exchange index data need to be smoothed first. Then, the daily closing price forecasting model of the Shanghai Composite Index is established, and the model is used to fit the daily closing price to infer its future trend.

II. Empirical Analysis and Prediction

2.1 Data Selection

Because the time series model usually requires a large number of data samples, a total of 370 data (closing price) is sampled from January 3, 2017 to July 10, 2018 on the Sina Finance official website. Based on the modeling theory of ARMA model, the daily closing price forecasting model of Shanghai Composite Index was established, and the short-term changes of the daily closing price of Shanghai Composite Index were studied.

2.2 Smooth processing of raw data

As the stock market volatility is relatively large, it can be seen from Figure 1.1 that the daily closing price of the Shanghai Composite Index has a certain trend, which is a non-stationary time series from January 3, 2017 to July 10, 2018, so it is necessary the raw data can be processed to be smooth. The sequence diagram of the first-order difference of the original data is drawn by Eviews, as shown in Figure 1.2.

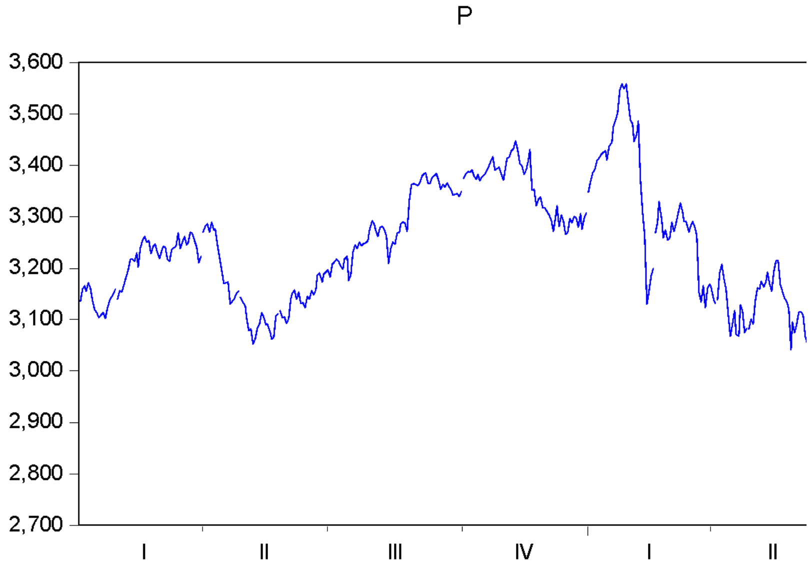


Figure 1.1The time series trend of the daily closing price of the Shanghai Composite Index

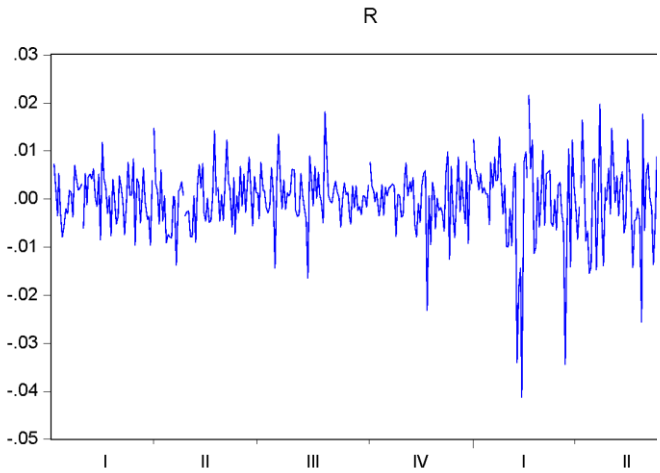


Figure 1.2 Sequence diagram after first-order difference

From the sequence diagram of the first-order difference of the original data in Figure 1.2, it can be roughly seen that the difference sequence may be stationary. Therefore, it is necessary to test the stability testof the first-order difference sequence by ADF. If it does not pass the test, it proves that the difference sequence is not stable. The results of the ADF test are shown in Figure 1.3:

Null Hypothesis: R has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-18.32770	0.0000
Test critical values:	1% level	-3.447914	

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 04 - Issue 07 || July 2018 || PP. 59-64

5% level	-2.869176
10% level	-2.570905

*MacKinnon (1996) one-sided p-values.

Figure 1.3 First-order differential post-ADF test results

The null hypothesis of the ADF test is that there is a unit root in the {R} sequence. According to the test, the unit root statistic ADF=-18.32770, whose absolute value exceeds the absolute value of the threshold value of -3.447914 at the significance level of 1%, and the P value is 0.0000, so the null hypothesis can be rejected, so after the first order difference The sequence does not have a unit root, so the sequence is a stationary sequence. Therefore, the sequence after the difference can pass the stability test. At this time, the smoothing process of the original data is completed.

2.3 Determine the applicable model and set the order

Observing the autocorrelation plot and the partial autocorrelation plot of the raw data of the closing price, as shown in Figure 1.4, the autocorrelation coefficient of the time series of the original data exhibits a trailing feature, which gradually decreases, and the previous order of the partial autocorrelation coefficient doesn't Within the 5% confidence interval, the original data sequence {p} is a non-stationary sequence.

Date: 07/20/18 Time: 23:06
Sample: 1/03/2017 7/10/2018
Included observations: 370

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.973	0.973	352.89	0.000
		2	0.943	-0.055	685.55	0.000
		3	0.909	-0.101	995.35	0.000
		4	0.871	-0.076	1280.8	0.000
		5	0.839	0.095	1546.3	0.000
		6	0.808	0.015	1793.4	0.000
		7	0.776	-0.073	2021.6	0.000
		8	0.746	0.018	2233.1	0.000
		9	0.711	-0.093	2426.0	0.000
		10	0.677	-0.005	2601.3	0.000
		11	0.645	0.021	2760.7	0.000
		12	0.614	0.010	2905.4	0.000
		13	0.585	0.019	3037.6	0.000
		14	0.557	-0.045	3157.6	0.000
		15	0.533	0.077	3267.8	0.000
		16	0.509	-0.034	3368.6	0.000
		17	0.492	0.125	3463.1	0.000
		18	0.476	-0.022	3551.6	0.000
		19	0.459	-0.038	3634.2	0.000
		20	0.444	0.017	3711.8	0.000

Figure 1.4 ACF and PACF diagram of raw data

The first-order difference data of the original data can be obtained and the correlation coefficient AC and the partial autocorrelation coefficient PAC can be observed to determine whether it is AR, MA or ARMA model.

Date: 07/20/18 Time: 23:09
Sample: 1/03/2017 7/10/2018
Included observations: 369

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.044	0.044	0.7069	0.400
		2	0.033	0.032	1.1236	0.570
		3	0.073	0.070	3.1206	0.373
		4	-0.121	-0.129	8.5967	0.072
		5	-0.003	0.004	8.5999	0.126
		6	0.033	0.037	9.0106	0.173
		7	0.041	0.058	9.6504	0.209
		8	0.028	0.005	9.9503	0.269
		9	0.035	0.025	10.407	0.319
		10	-0.011	-0.015	10.455	0.402
		11	-0.001	0.008	10.455	0.490
		12	-0.044	-0.046	11.190	0.513
		13	0.001	0.011	11.190	0.595
		14	-0.080	-0.088	13.684	0.474
		15	0.005	0.018	13.695	0.549
		16	-0.031	-0.043	14.068	0.594
		17	0.024	0.044	14.292	0.646
		18	0.025	0.002	14.544	0.693
		19	-0.011	0.001	14.588	0.748
		20	0.028	0.018	14.891	0.783

Figure 1.5 ACF and PACF diagrams for the first-order differential post-sequence

(1) First observe the AC and PAC maps of the first-order differential data R. It can be seen from Fig. 1.5 that the autocorrelation and partial autocorrelation plots of the first order difference sequence {r} have no obvious tailings, so we can try to use the ARMA model to establish the model. The specific lag term p, q Values are also specifically determined using AIC criteria and T statistical significance.

(2) Try different models and determine the model ARMA (p,q) according to the principle of AIC and SC minimization. After comparing the different ARMA (p,q) models by multiple rounds, the values of corresponding AIC and SC can be obtained.

In general, there are many methods for the order of p and q in ARMA (p,q). The simple ARMA ordering method is based on the significant order of ACF and PACF graphs. As seen in Figure 1.5, four models are selected for comparison. The four models are ARMA (1, 1), ARMA (1, 2), ARMA (2, 1), and ARMA (2, 2). as follows:

Table 1.1 Indicators for each model

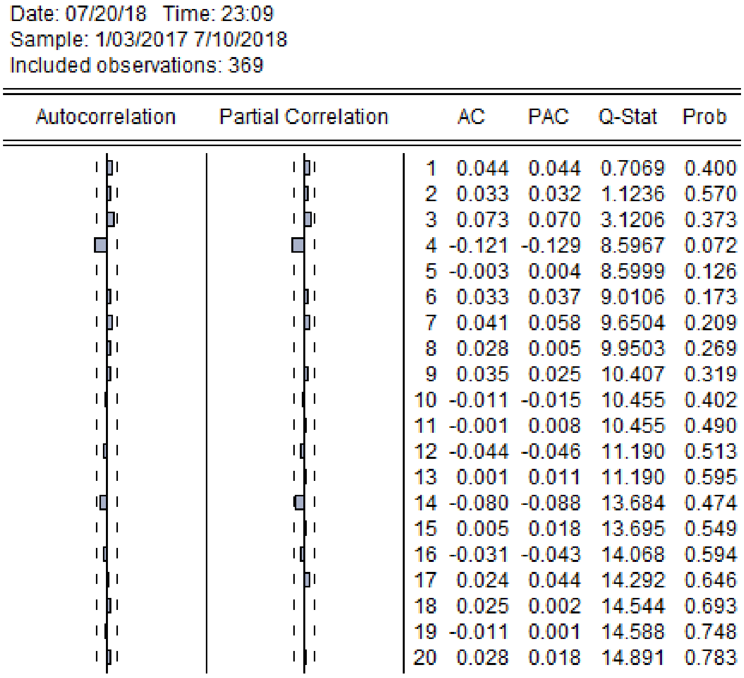
	AIC	SC	T statistical significance
ARMA(1,1)	-6.829032	-6.807792	Not obvious
ARMA(1,2)	-6.824769	-6.792909	Not obvious
ARMA(2,1)	-6.821662	-6.789738	Not obvious
ARMA(2,2)	-6.827274	-6.784708	Significant

From the test results of the above four models, the comprehensive test of T statistical significance and AIC criterion, the comparison shows that the coefficients in the ARMA (2, 2) model are very significant, and the AIC of this model The value is relatively small. Therefore, the original data sequence is modeled and predicted using the ARMA (2, 2) model.

2.4 ARMA model residual test

After the parameter estimation, the applicability of the fitted model should be tested. Essentially, a white noise test is performed on the residual sequence of the model. If the residual sequence is not white noise, there is some important information that has not been extracted, so the model should be reset. If the sample autocorrelation coefficients of the residual sequence fall within the random interval, there is no statistically significant autocorrelation, then the residual sequence can be said to be completely random. The residuals were tested using Eviews software and the results are shown in Figure 1.6:

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 04 - Issue 07 || July 2018 || PP. 59-64



Date: 07/20/18 Time: 23:09
Sample: 1/03/2017 7/10/2018
Included observations: 369

	AC	PAC	Q-Stat	Prob
1	0.044	0.044	0.7069	0.400
2	0.033	0.032	1.1236	0.570
3	0.073	0.070	3.1206	0.373
4	-0.121	-0.129	8.5967	0.072
5	-0.003	0.004	8.5999	0.126
6	0.033	0.037	9.0106	0.173
7	0.041	0.058	9.6504	0.209
8	0.028	0.005	9.9503	0.269
9	0.035	0.025	10.407	0.319
10	-0.011	-0.015	10.455	0.402
11	-0.001	0.008	10.455	0.490
12	-0.044	-0.046	11.190	0.513
13	0.001	0.011	11.190	0.595
14	-0.080	-0.088	13.684	0.474
15	0.005	0.018	13.695	0.549
16	-0.031	-0.043	14.068	0.594
17	0.024	0.044	14.292	0.646
18	0.025	0.002	14.544	0.693
19	-0.011	0.001	14.588	0.748
20	0.028	0.018	14.891	0.783

Figure 1.6 ARMA (2, 2) model residual test results

As can be seen from the above figure, the AC and Q statistic tests for their residuals found that the residual autocorrelation is substantially close to zero. In addition, the probability values of a column on the right side of the figure are greater than 0.05, indicating that all Q values are less than the chi-square distribution threshold of the test level of 0.05. It can be seen that the residual information is less and the model fit is basically the same. That is, the random error term of the model is a white noise sequence. Therefore, it is appropriate to establish this model.

2.5 Model prediction analysis

Using the established ARMA (2, 2) model to predict the closing price of the Shanghai Composite Index, because the price changes of the stock are relatively large, the short-term forecast can get better results, but the long-term forecasting effect will have a larger error. Therefore, this paper mainly conducts short-term forecast of stock prices, and predicts the closing price of the three days from July 11, 2018 to July 13, 2018.

Table 1.2 Comparison of actual and predicted values

Date	Actual value	Predictive value	Absolute error
7/11/2018	2777.771	2827.645	0.017638
7/12/2018	2837.659	2827.703	0.003520
7/13/2018	2831.184	2827.768	0.001207

It can be known from the above table that there is a certain error in the prediction, but compared with the previous fluctuation trend, the model established in this paper is more accurate, and has a good practical significance for the short-term prediction of the closing price of the Shanghai Stock Exchange.

III. Conclusion

In this paper, the historical closing price of the above index is used as a time series to analyze and discover the trend of stock price changes, and to predict the closing price of future trading days, the error is small, and good results are achieved. It provides a certain reference for short-term investment activities of small and medium investors, and has certain practical significance. However, the time series of the stock price is a non-stationary time series, and a large amount of information is lost after the difference is made, so the estimated coefficient of the estimated time series model is not high. Furthermore, the prediction of the ARMA model can only be used for short-term predictions. The error generated by the longer time will be larger, and the reference value for long-term investment is not large.

In summary, the ARMA model can not only solve the modeling problem of non-stationary time series, but also can be used for time series prediction. With the help of Eviews software, the ARMA model can be easily applied to the research and prediction of time series problems such as finance, providing decision-making guidance and assistance to decision makers. Of course, due to the complexity of the financial time series, good simulations require further research and discussion.

References

[1]. Liang Yan,. Xia Letian., Application of Time Series ARMA Model, *Journal of Chongqing University of Technology(Natural Science)*,2012, (08):106-109.

[2]. Feng Pan,.Cao Xianbing,, An Empirical Study of Stock Price Analysis and Forecast Based on ARMA Model, *Mathematics in Practice and Theory,* 2011, (40) 22: 85-89.

[3]. Wu Chaoyang., Improved Grey Model and Stock Index Prediction of ARMA Model, *Journal of Intelligent Systems,* 2010, 5(3): 277-280.

[4]. Meng Kun. Li Li, An Empirical Analysis of Forecasting Stock Price Based on ARMA Model, *Journal of Hebei North University(Natural Science Edition)*,2016,32(5).

[5]. Cong Zheng. Xu Jiaping., Research on Rural Financial Gap Calculation in Liaoning Province Based on ARMA Model, *Journal of Shandong Agricultural University*,2013(4).

[6]. WANG Peng, Lü Yongjian., Research on the Relationship between Stock Market Volatility and Related News Factors Based on Message Model, *Mathematical Statistics and Management,* 2013(6).

[7]. Chen H F. Zhao W X., New Method of Order Estimation for ARMA/ARMAX Processes, 2010, 48(6).