Forecasting Analysis of Shanghai Stock Index Based on ARIMA Model

Chenggang Li [1, 2], Bing Yang[1] and Min Li[1]

[1]Faculty of Finance, Guizhou University of Finance and Economics, Guiyang, 550025, China

[2]Faculty of Science, National University of Singapore, 117546, Singapore

Abstract. Prediction and analysis of the Shanghai Composite Index is conducive for investors to investing in the stock market, and providing investors with reference. This paper selects Shanghai Composite Index monthly closing price from Jan, 2005 to Oct, 2016 to construct ARIMA model. This paper carries on the forecast of the last three monthly closing price of Shanghai Stock Index that have occurred, and compared it with the actual value, which tests the accuracy and feasibility of the model in the short term Shanghai Stock Index forecast. At last, this paper uses the ARIMA model to forecast the Shanghai Composite Index closing price of the last two months in 2016.

1 Introduction

The stock market is an embodiment of the economic strength of the real industry. The Shanghai Composite Index is the embodiment of a country's overall entity economy level. In some ways, the stock price reflects a national economy strength. We can know the rise and fall of a country's real industry by observing the share price. The Shanghai Composite Index can reflect the most intuitive of the overall economy.

Therefore, the Shanghai Composite Index forecast has important significance for a country in latter regulation and control of the real economy, and the development of regional policy. But it is difficult to analyze and predict the Shanghai Composite Index by using the traditional structural method, and the result error is quite large. So choosing suitable *ARIMA* model to predict the Shanghai Composite Index can be satisfied with the results.

2 ARIMA(p,d,q) model

Setting X_t sequence is a random non-stationary time series. After the X sequence for D order difference, we can conclude:

$$Y_t = \nabla^d X_t \tag{1}$$

Among them, ∇ is the differential operation. d is the differential order [1]. After the difference, sequence Yt has the following structure model:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{2}$$

Integrated auto-regressive average moving is referred as the *ARIMA (p, d, q)* model. Among them, $\boldsymbol{p}$ is auto-regressive order. $\boldsymbol{d}$ is the moving average. $\{\varepsilon_t\}$ is the white noise sequence. φ_i indicates the auto regressive coefficient, and θ_i is the moving average coefficient.

The auto-regressive moving average model, which is first proposed by the American statistician G.E.P.Box and G.M.Jenkins in 1970, is widely used in time series data analysis [2]. It is a effect prediction method with high precision of short-term.

3.The application of the model in the forecasting analysis of Shanghai Stock Index

3.1 Data selection and processing

This paper selects Shanghai Composite Index monthly closing price from Jan, 2005 to Oct, 2016. All data are from the Shanghai Security Exchange [3]. All data processing and final data prediction are carried out using EViews7.2.

3.2 Stationarity test

We do the stationary test of Shanghai Composite Index monthly closing price by ADF test. The test results are shown in Table 1. From the test results, it can be considered that the original time series is not stationarity, but the first order differential sequence is stationarity.

Table 1. Stationarity test results.

Sequences	T statistics	P values	Conclusions
Original sequence	-2.0807	0.2528	Non-stationarity
First order Differential sequence	-6.5150	0.0000	Stationarity

3. 3 Parameters determination

There are three important parameters in the *ARIMA(p,d,q)* model, which are the auto-regressive order p, the difference order d and the moving average q. In the front, after one difference to determine the smooth of the sequence, we can get $\boldsymbol{d=1}$. Next, it is necessary to determine the value of parameter p and parameter q in order to determine the final model *ARIMA (p, d, q)*. By analyzing the auto-correlation and partial correlation graphs of time series $\{Y_t\}$, the value of p and q can be determined. The self-correlation and partial correlation test results of the sequence are shown in Figure 1.

From Figure 1, we can see that the partial correlation number is significantly truncated, while the self-correlation coefficient falls on the edge of 2 times standard deviation in the time lag of second and forth order [4]. In this case, it is very difficult to use the traditional Box-Jenkins method (auto-correlation partial auto-correlation function, residual variance, F test, the criterion function) to determine the model order. For this special case, this paper makes repeatedly a comparison of the parameters of the model to determine the significance of the parameters of different models.
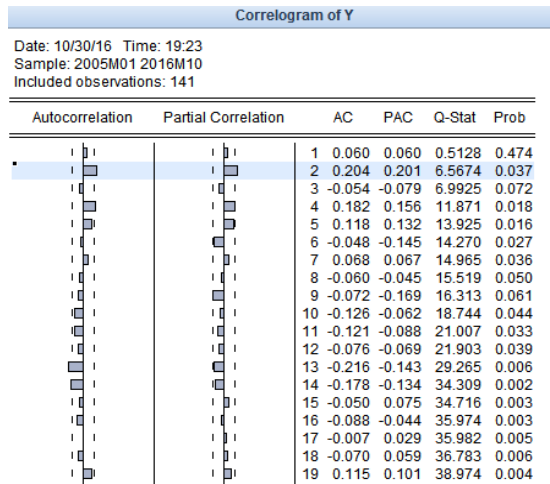
Fig. 1. Self-correlation and partial correlation of first order difference sequence.

For the sequence Y, we try several different models, such as *ARMA (2.2)*, *ARMA (4.4)*, etc. The results of t test (P value) of detecting the parameters of different models are used to select the best model.

According to the *ARMA (2,2)* model, the regression analysis results are shown Table 2.

Table 2. Regression analysis of *ARMA (2,2)*

Variable	Coefficient	Std.Error	t-Statistic	Prob
C	2795.319	390.4668	7.158915	0.0000
AR(1)	0.383696	0.201107	1.907921	0.0585
AR(2)	0.492393	0.194084	2.537005	0.0123
MA(1)	0.63379	0.194721	3.511645	0.0006
MA(2)	0.340521	0.08479	4.016054	0.0001
R-Squared	0.914277	Mean dependent var		2643.74
Adjust R-squared	0.911737	S.D.dependent var		937.99
SE of regression	278.6684	Akaike info criterion		14.132
Sum squared resid	10483574	Schwarz criterion		14.238
Log likelihood	-984.3088	Hannan-Quinn criter		14.175
F-statistic	359.9612	Durbin-Wast on stat		2.0367
Prob(F-statistic)	0.0000			

According to the *ARMA (4,4)* model, the regression analysis results are shown in Table 3.

Table 3. Regression analysis of ARMA (4,4).

Variable	Coefficient	Std.Error	t-Statistic	Prob
C	2720.817	46.94.33	57.96331	0.0000
AR(1)	0.618826	0.081399	7.602379	0.0000
AR(2)	0.779381	0.09541	8.168787	0.0000
AR(3)	0.303961	0.086531	3.513458	0.0006
AR(4)	-0.74622	0.066057	-11.29967	0.0000
MA(1)	0.367347	0.120799	3.040993	0.0029
MA(2)	-0.339289	0.043801	-7.746205	0.0000
MA(3)	-0.972002	0.046586	-20.86487	0.0000
MA(4)	-0.024311	0.115923	-0.209719	0.8342
R-Squared	0.925244	Mean dependent var		2664.049
Adjust R-squared	0.920607	S.D.dependent var		929.2741
SE of regression	261.8386	Akaike info criterion		14.03633
Sum squared resid	8844170	Schwarz criterion		14.22724
Log likelihood	-959.5066	Hannan-Quinn criter		14.11391
F-statistic	199.5754	Durbin-Wast on stat		2.000099
Prob(F-statistic)	0.0000			

Through the comparison of parameters significant t test results (P value) and the fitness of R-squared of the *ARMA (2.2)* and *ARMA (4.4)* model, it can be obtained, the t test is significant and R-squared is optimal in $p = 4$ and $q = 4$. The model *ARIMA (4, 1, 4)* is established.

3.4 ARIMA(p,d,q) model checking

After the completion of the model construction, it is necessary to carry out adaptive testing, to determine whether it is appropriate. For the model *ARIMA(p,d,q),* it is appropriate to determine whether the model is suitable by testing whether the residual sequence is white noise sequence. When the residual series is not a white noise sequence, it is shown that there is information contained in the residual error. The established model is certainly not the final model, and other parameters can not fully represent [5]. It is necessary to further estimate model. If the residual sequence is white noise, it is shown that the proposed model is suitable.

White noise test results (For space consideration, the results are not list) show that the auto-correlation coefficient of the residual sequence is in random interval, there is no residual serial correlation, and they are close to zero in each lag order auto-correlation and partial auto-correlation values. All of the Q-statistics is not significant, and the P values of lag of the Q-statistics are significantly greater than 0.05. So the residual sequence passes the white noise test. It is white noise sequence. So the *ARIMA(4,1,4)* model can be used for further analysis and prediction.

3.5 Model estimation

From the previous calculation, the final model is determined as the model *ARIMA(4,1,4)*. Through the Eviews7.2 software, we get the coefficients of the *ARIMA(4,1,4)* model. The results are shown in Table 4.

Table 4. Coefficient estimation of ARIMA(4,1,4) model

Variable	Coefficient	Std.Error	t-Statistic	Prob
C	2749.145	178.5409	15.39784	0.000
AR(4)	0.622969	0.079429	7.843075	0.000
MA(4)	0.373878	0.098223	3.806418	0.002

From the above table, the mathematical specification of the model *ARIMA(4,1,4)* is

$$Y_t = 2749.145 + 0.6229Y_{t-2} + \varepsilon_t + 0.3738\varepsilon_{t-2} \tag{3}$$

3.6 Forecasting analysis of Shanghai Composite Index

Firstly, we use the *ARIMA(4,1,4)* model to forecast the closing price of Shanghai Composite Index from July, 2016 to October, 2016. The relative errors between the predicted and actual values can be calculated to determine the feasibility and accuracy of the model [6]. The relative errors are calculated by the following formula:

The relative error = (actual value - prediction value / actual value)

The predicated value of the closing price of Shanghai Composite Index and the relative errors are shown in table 5 and Figure 2.

Table 5. Predicated value and error analysis.

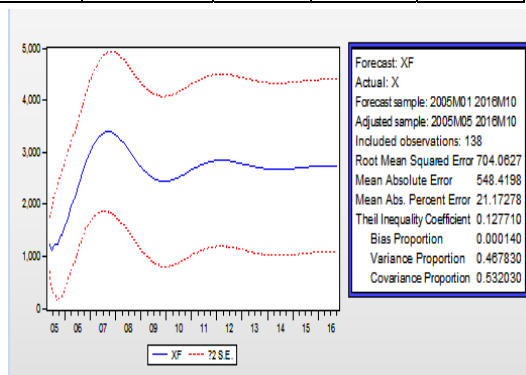| Data	Predicative value	Actual value	Relative error	Average error
2016-08	3010.56	3085.49	-2.4%	
2016-09	306073	3004.70	1.8%	2.4%
2016-10	3005.10	3100.49	3.1%	

Fig. 2. Short term prediction.

Comparing the predicated value and the actual value, we can see that the average error is 2.4% less than 5%. Seen from the graph, the Theil inequality coefficient is 0.1277, which shows that the model has better prediction effect, and its decomposition table shows that the error ratio is close to zero, and the variance is small.

Secondly, we use the closing price of the Shanghai composite index from January, 2005 to October ,2016 to forecast the closing price of November, 2016. The main calculation method is the deferred method. We take the forecast closing price data of November, 2016 data back, combined the data from January 2005 to October, 2016 to forecast the closing price of December, 2016. The deferred method can get more accurate results [7]. The predicated closing price of the Shanghai Composite Index in November and December, 2016 are shown in Table 6.

Table 6. Predicted results.

Time	Nov,2016	Dec,2016
Predictive value	3273.52	3195.7

4 Conclusions

Changes of the Shanghai Stock Index are the focus of investors and investment institutions. The forecasting of the Shanghai Stock Index is the hot-spot and difficulty in the financial sector. In the varieties of the methods, the *ARIMA* model is one of the typical representative, and has its own unique advantages in the prediction of the Shanghai Stock Index. *ARIMA* model is good at capturing the linear characteristics of the data, and has quite ideal effect for the short-term prediction.

In this paper, the *ARIMA* model is used to analyze the Shanghai Composite Index in October, 2016 and predict the changes of the Shanghai Composite Index in November and December, 2016. The results show that in the last two months of 2016, the Shanghai Composite Index will have a small rise. This gives investors a strong basis for the future of the market judgment, and can let investors in the stock market reduce the loss impacting by the risk of future uncertainty.

References

1. Q. Wang. Time series analysis and application, Southwest Jiao- Tong University Press, (2008).

2. J. Yu. Econometrics software - the use of Eviews , University of International Business and Economics Press, (2006).

3. Y. Wang. Applied time series analysis ,China Renmin University Press, (2005).

4. B. Zhang, X. Liu. Dynamic ecological footprint simulation and prediction based on ARIMA Model: a Case study of Gansu Province, China, Acta Ecologica Sinica,31, (2011).

5. Z. Xiong. Research on RMB exchange rate forecasting model based on combining ARIMA with Neural Networks , The Journal of Quantitative & Technical Economics, 6, (2011).

6. S. Shao. Forecast research on ARIMA model for China's health insurance, Journal of Baoji University of Arts and Sciences. 3, (2016).

7. L. Xiao. Forecast of resident's consumption level based on seasonal ARIMA model, Statistics and decision making.8, (2016).