

An Empirical Study on Stock Price Based on Arima Mode

Yang Xiaoguang
Dalian Maritime University
Dalian, China
e-mail:xgyang57@163.com

Lu Yuzhen
Dalian Maritime University
Dalian, China
e-mail:dlyuzhen@dlnu.edu.cn

Yu Ting*
Dalian Maritime University
Dalian, China
e-mail:mengxier1@163.com
*corresponding author

Chu Zhiyuan
Dalian Maritime University
Dalian, China
e-mail:chuzhiyuan6@sina.com

Abstract—the model in time series analysis are widely used in the field of economy, it can describe the historical data change over time, and analyze data to make predictions. The ARIMA is one of the important models in time series models. This paper summarizes several smooth data processing methods. Using the A-Information Criterion and the principle of the precision of the model, it gives the step and method to build ARIMA model. The ARIMA model is used to analyze the data of China Merchants Bank(600036) shares at the opening price (2013/01/04-2013/10/18) and to predict the next five days (2013/10/21-2013/10/25) stock opening price data. In contrast with the actual data, the model prediction error is smaller, indicating that ARIMA model is very suitable for short-term forecasts.

Keywords- time series;ARIMA model;forecast;stock prices

I. INTRODUCTION

With the rapid development of science and technology progress of financial market, the development of Chinese stock market is moving from immature towards mature gradually, and become one of the most important capital markets in our country. Both stock investors and managers are looking forward to less investment return for more. We think the stock information is reflected in the stock price, through understanding and analyzing of the stock price, we can know the information of the stock market and dig out the useful information. So change of the stock market causes different price of the stock market information. The information randomness leads to today's and tomorrow's price were independent with each other. Thus the paper uses time series analysis to analyze and forecast the stock price. ARIMA model is the one of the most important models in time series which can well predict the stock prices.

Using the model in time series analysis to analyze and forecast the data is not uncommon. In 2011, Anderson Paul L^[1] used the ARMA model to forecast the data. It's

also can be seen that using the model in time series analysis for analysis and prediction of stock price in Lina shao's paper, Fusheng Quan's paper, Jun Deng's paper and Pan Feng's paper from 2007 to 2011^[2-5]. Where, the values of p and q are determined by guess in Pan Feng's paper^[5]. On the basis of those papers, this paper summarizes several methods for preparation of stable data. In the paper the values of p and q are determined by the A-Information Criterion, not by guess. It is given that the values of p and q are more theoretical and targeted. Using the principle of model's precision to build ARIMA model. It's found that the forecast values are closer to the real value.

II. MODELING STEPS OF ARIMA MODEL

- According to the sequence diagram of the original sequence and autocorrelation figure, or through unit root (DF) test to determine sequence of stability.
- Stabilizing the original sequence. If the sequence is non-stationary, make it stationary by difference, and examine its stability.
- Using A-Information Criterion to identify models and determine the values of p and q, then build the ARIMA model.
- Estimated the parameters of the model test its significance, ensure the precision of the model.
- Test the residual in order to ensure the rationality of the model(if the residual sequence can pass the white noise test, it's indicated that the sequence of information has extracted intact and there is no need to analyze. if not, adjust the model until passing the white noise test).
- Using the established model to forecast.

III. USING THE ARIMA MODEL TO ANALYZE AND FORECAST THE CHINA MERCHANTS BANK SHARES

A. Data selection

Due to the sequence model often requires large sample, the paper select data of China merchants bank shares at the opening price (2013/01/04-2013/10/18) as the raw data. It can meet the needs of ARIMA model.

B. Analysis of the original sequence

The ARIMA model is established by analyzing the original sequence (China merchants bank(600036) shares of day opening price (2013/01/04-2013/10/18)). Most of the time series we usually encountered is not smooth. Here are several ways to judge the stationary of original sequence. With the help of SAS, we draw the sequence diagram of the original sequence. As is shown in figure 1:

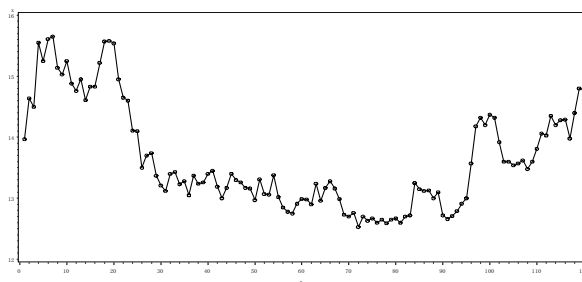


Figure1.The sequence diagram of the original data

Observe the figure 1, we find that figure is not fluctuate around a constant(the mean value function of stationary time series is constant and the covariance function is only related to the time interval, so the figure should be around a constant ups and downs), so this sequence is not smooth. Though inspecting in the data graph is intuitive, but not precise enough, we can implement stationary through the autocorrelation function test. Use SAS to draw the autocorrelation of the sequence. As is shown in figure 2:



Figure2.Autocorrelations of the original data

Observe the figure 2, we find that the autocorrelation function in the slow decay (A zero mean stationary series of autocorrelation function is either truncated, or trailing), so this sequence is not smooth. Sometimes we could not determine if it is smooth or not by observing the correlation diagram (figure2), thus we can use the unit root test.

In the SAS system, it provides a “%DFTEST” macro to process the Dickey-Fuller unit root test (the DF test), the null hypothesis and alternative hypothesis are as follows:

H_0 : Time series has a unit root

H_1 : Time series is stationary series (there is no unit root)

After running the program, you can get the p value of the original data in the LOG window by the unit root test, the $P(x) = 0.5413406209$, thus it can be seen that the x variables of unit root test is not significant. $P(x)$ values is very large, there is no good reason can reject the null hypothesis. The original sequence has a unit root and it is a smooth sequence.

Because of the original sequence is not stable. We need to make it into a stationary series by difference. Using SAS for make first order difference of the original sequence. By observing the difference after the sequence diagram, autocorrelation figure and through the unit root test, we demonstrate that the difference of data sequence is smooth (sequence diagrams and autocorrelation figure as shown in figure 3).

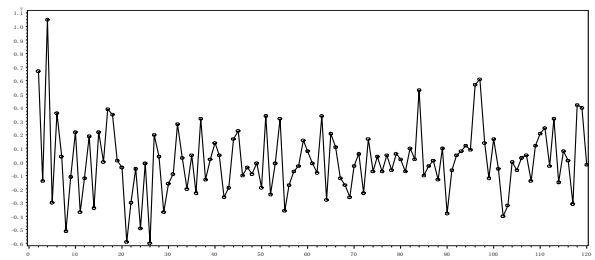


Figure 3. The difference after the sequence diagram, autocorrelation figure

At this point, this paper completed the original sequence of smooth processing,

C. Model identification

We must solve the concrete numerical value of p and q for build the ARIMA model. To determine the specific numerical process of p and q is the identification process of the model (in terms of model identification is in view of the stationary series).

Here using the A-Information Criterion for model identification. In the SAS system, it can compute all the possible BIC information index of ARMA(p,q) models. And it based on the system calculates the BIC index is the smallest model as a recognition based on. Use the A-Information Criterion to model identification in the SAS system. it needs to determine the lower limit values of p and q. On the determination of p and q when the lower limit value, it can be based on autocorrelation and partial correlation diagram, and no special provisions. Because of the economic variables are generally for ARIMA model of order 1 or 2 order, so the regulation of p and q minimum value is 0 and a maximum of 4. After running the program,

they get the minimum information value table of all sorts of model. Figure 4 shows:

Minimum Information Criterion					
Lags	MA 0	MA 1	MA 2	MA 3	MA 4
AR 0	-2.88329	-2.88474	-2.86831	-2.83341	-2.81017
AR 1	-2.88431	-2.85268	-2.83572	-2.80053	-2.77697
AR 2	-2.8597	-2.82748	-2.80983	-2.77443	-2.75263
AR 3	-2.82483	-2.79383	-2.77567	-2.75066	-2.72725
AR 4	-2.80052	-2.76836	-2.74983	-2.7246	-2.70114

Error series model: AR(8)
Minimum Table Value: BIC(0,1) = -2.88474

Figure 4. the A-Information Criterion

In figure 4 output of various models of BIC information index, the SAS system in the last row of the table “Minimum Table Value” automatically prompt “BIC (0,1) = -2.88474”. The smallest BIC index is the ARIMA(0,1,1) model. The system has elected the BIC minimum model, but the parameters of the model is significant needs to take a test. There is the parameter estimation of the model.

D. The parameter estimation of the model

After identifying the model of time series and determining the specific form of the model, it's necessary to estimate the parameter and test on the significance of parameters of the model. The purpose of the screening test is in order to make the most compact model. After inspection, the parameter is not significant variables was removed from the model, thus the modeling again to get the most accurate model. There are three kinds of parameters estimation method of the model: Torque estimation, maximum likelihood estimation and least squares estimate. Using least squares estimation of model parameters were estimated by SAS. Running the SAS program, it will get the following results. Fig.5 shows:

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	-0.01014	0.02076	-0.87	0.3838	0
AR1,1	0.01503	0.08552	0.18	0.8608	1
Constant Estimate			-0.01786		
Variance Estimate			0.058115		
Std Error Estimate			0.24107		
AIC			0.948682		
SBC			6.817629		
Number of Residuals			139		

* AIC and SBC do not include log determinant.

Figure 5. ARIMA(0,1,1)model parameters estimation ,AIC and SBC values

It can be seen the ARIMA(0,1,1,)model parameters estimation in “Pr>|t|” value were greater than 0.05. The results show that the parameters of the model was not significant, we need to adjust the model of p and q value to make the parameters of the model significantly and ensure that the model is most accurate. According to the adjusted the ARIMA(2,1,2) model parameters greatly. Fig. 6 shows:

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	-0.01799	0.02130	-0.84	0.3999	0
MA1,1	-0.85348	0.18674	-4.57	<.0001	1
MA1,2	-0.83581	0.18630	-4.49	<.0001	2
AR1,1	-0.83885	0.22741	-3.69	0.0003	1
AR1,2	-0.74324	0.22818	-3.26	0.0014	2
Constant Estimate			-0.04845		
Variance Estimate			0.058171		
Std Error Estimate			0.241188		
AIC			4.004923		
SBC			18.87729		
Number of Residuals			139		

* AIC and SBC do not include log determinant.

Figure 6. ARIMA(2,1,2)model parameters estimation ,AIC and SBC values

The parameter of this model is significant by observing the figure 6. We find that the AIC value and SBC value of the ARIMA(0,1,1,) than the AIC value and SBC value of the ARIMA(2,1,2) is small, but its parameters are not significant. We need to create a simplified model to forecast, no significant parameters of model prediction results are often not satisfied. Only by examining the model parameters are significant to ensure the precision of the model. So the ARIMA(2,1,2) is a simplified model. Its expression is:

$$X_t = -0.01799 - 0.83885X_{t-1} - 0.74324X_{t-2} + \varepsilon_t + 0.85348\varepsilon_{t-1} + 0.83581\varepsilon_{t-2}$$

E. Residual test of the model

Residual test refers to an ARIMA model has completely or mostly explains the correlation of the data. Check the SAS software is used to analyze the model residuals. Residual inspection is whether white noise test of residual sequence (figure 7).

The ARIMA Procedure									
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	3.75	2	0.1537	0.009	0.025	0.035	-0.015	-0.039	-0.117
12	9.68	8	0.2897	-0.057	0.146	0.077	0.023	-0.030	0.010
18	17.68	14	0.2216	0.061	-0.108	-0.042	-0.017	0.052	-0.173
24	27.54	20	0.1207	0.065	-0.061	0.056	-0.215	0.006	0.040

Figure 7. White noise test of residual sequence

It can be seen from the figure 7 in “Pr>ChiSq” the value is greater than 0.05. The remnants of difference as a white noise sequence. That means no residue in the residual sequence information, and fitting model is obvious.

F. model prediction

The ARIMA model is used to predict the next five days (2013/10/21-2013/10/25) stock opening price data. Prediction results are as follows:

Date	True value	Predicted value
2013/10/21	10.63	10.6210
2013/10/22	10.74	10.5844
2013/10/23	10.68	10.5753
2013/10/24	10.68	10.5637
2013/10/25	10.60	10.5338

By contrast, we find that the built model predictive value compared with the real value, the error is small and it is close to the real value. It shows that the built model fitting is very good. It showed the ARIMA model to short-term forecast the price of the stock.

IV. CONCLUSION

From what has been discussed above, the paper introduced modeling ideas of the ARIMA model and analysis the shares of China Merchants Bank. It further perfect and improve the ARIMA model, so that the prediction results are more accurate. In the entire modeling process, using the SAS software is a good way to analysis and forecast data.

Therefore, this article can get the build ARIMA model to analysis and forecast the stock by SAS. It can make the investors and managers to maximize interests.

REFERENCES

- [1] Anderson Paul L, Meerschaert Mark M, Zhang Kai. Forecasting with prediction intervals for periodic ARMA models[J]. Journal of time series analysis/a journal sponsored by the Bernoulli Society for Mathematical Statistics and Probability, 2013, 34(2): 187-193.
- [2] Shao Li-na. The stock price prediction based on ARMA model[J]. The agricultural economy and technology, 2007(12): 105-107.
- [3] Quan Fu-sheng, Peng Bai-yu. The application of the ARMA model in Chinese stock market[J]. Journal of Hengyang Normal University, 2009, 30(3): 26-28.
- [4] Deng Jun, Yang Xuan, Wang Wei, et al. Using ARMA model for stock price forecasting of empirical research[J]. Guide to Business, 2010(6): 266-267.
- [5] Feng Pan, Cao Xian-bing. An empirical study on the stock price analysis and prediction based on ARMA model[J]. Mathematics in Practice and Theory, 2011, 41(22): 84-90.
- [6] Ping-Feng Pai. A hybrid ARIMA and support vector machines model in stock price forecasting[J]. Department of Industrial Engineering and Technology Management, 2005, 33: 497-505.
- [7] Sangyeol Lee, Siyun Park, Koichi Maekawa, et al. Test for parameter change in ARIMA models[J]. Communication in Statistics-Simulation and Computation, 2006, 35(2).
- [8] Simon Stevenson. A comparison of the forecasting ability of ARIMA models[J]. Journal of Property Investment & Finance, 2007, 25(3).
- [9] O. Valenzuela, I. Rojas, F. Rojas, et al. Hybridization of intelligent techniques and ARIMA models for time series prediction[J]. Fuzzy Sets and Systems, 2007, 159(7): 821-845.
- [10] Fong-Lin Chu. Forecasting tourism demand with ARMA-based methods[J]. Tourism Management, 2008, 30(5): 740-751.
- [11] Zhang Hui, Liu Meng. Forecasting of China resident's consumption level based on ARMA model[J]. International Business and Management, 2011, 2(2): 62-65.
- [12] Juan J. Flores, Mario Graff, Hector Rodriguez. Evolutive design of ARMA and ANN models for time series forecasting[J]. Renewable Energy, 2012, 44: 225-230.