

Analysis and Forecast of monthly time Series of Shanghai Stock Index based on ARIMA Model

Yu Zhu

Beijing-Dublin International College at BJUT, Beijing, China

Keywords: The Shanghai Composite Index; ARIMA; One step forward static prediction; B-J methodology

Abstract: The complexity of the influence factors of stock price index, which affect stock price in our country at present stage are the main areas of the bank, bond market and futures market, real estate, currency, etc., from the current trend of the development of the finance and the vast majority of investors in the stock market of numerous financial tools pressing needs, through the establishment of appropriate time series model can achieve rough projections for overall stock prices. This article selects the from China's accession to the WTO in December 2011 to 2014, the Shanghai composite index since July monthly data, through the establishment of ARIMA model adopts the method of static prediction step forward for China's stock market in August 2014, the Shanghai composite index is forecasted, in China 2014 years ago by two quarters showed a trend of rising stock market as a whole. The innovation of this paper is to take the logarithm of the sample data, so as to eliminate the autocorrelation and heteroscedasticity in the time series, and at the same time make the predicted value close to the actual value, the effect is good, hoping to provide a reference for the majority of shareholders.

1. Introduction

Stock prices are "barometer" of the national economy, its formation and fluctuation affected by domestic and foreign various political and economic, in order to better study the operation of the stock market, we can use the stock index, stock index is to describe the stock market index of change in the overall price level, the Shanghai composite index by the Shanghai stock exchange to use their business knowledge and familiar with the advantages of the market, and released, has a certain authority. Investors can then test the effectiveness of their investments and use them to predict the movements of the stock market. At the same time, the press, company bosses and even political leaders take this as a reference index to observe and predict the social, political and economic development situation. Due to the unpredictable stock prices, the traditional time series is difficult to describe and time-consuming [1]. Therefore, we can adopt ARIMA model, which is a time-series short-term prediction method with high accuracy. By studying the Shanghai Composite Index through this model, we can have a more essential understanding of its structure and characteristics, so as to achieve the optimal prediction in the sense of minimum variance [2].

2. Theoretical interpretation and modeling steps of ARIMA model

2.1. The theoretical connotation of ARIMA model

The so-called ARIMA model is a model established by converting non-stationary time series into stationary time series, and then regression of the dependent variable only on its lag value and the present value and lag value of the random error term. If the time series:

Then the autoregressive moving average model of the order of the time series is called ARMA. Most of the financial time series are unstable, and we transform them into stationary time series by one or more difference methods. If the series $\{Y_t\}$ gets a stationary time series after d differences, and uses the ARMA (pMagneQ) process pair to establish the model, it is called (pQuery d). Q) order autoregressive monolithic moving average process, abbreviated as ARIMA.

2.2. The B-J methodology was used to model ARIMA

2.2.1. Model identification

First, the model is tested for its stationarity. If it is not stationary, the model is made differential, and $d=n$ for n times of difference. Then, the auto-correlator and partial autocorrelator of the stationary sequence are observed and their trailing and truncation are analyzed to determine the auto-regression order P and the moving average order Q .

2.2.2. Model to estimate

After the above appropriate D , P and Q have been confirmed, OLS or maximum likelihood estimation method will be used to further estimate and move average coefficients

2.2.3. Model test

The suitability of the model should be tested after the model is estimated. In order to determine whether the sequence of residuals is random, tests can be used. If it passes the test, the estimated model is accepted, otherwise it needs to start again from the first step. At the same time, in order to better fit the data, the lag term was added to the model, and then the judgment was made according to AIC and SC principles.

3. ARIMA model in Shanghai Composite Index empirical analysis and forecast

This article in the Shanghai composite index monthly closing price as the research object, from December 11, 2001 by the China's entry into WTO on July 14, 2014 monthly closing price (excluding the transaction date) for a total of 152 data, to meet the needs of the large sample stock index research, using ARIMA model modeling, data from input after the big wisdom website EVIEWS6.0 modeling analysis [3].

3.1. Identify the stationarity of time series data

In order to eliminate or reduce autocorrelation and heteroscedasticity in time series without affecting the results of model analysis, logarithms are taken for all data. Set December 2001 as the closing price of the first month, January 2002 as the closing price of the second month, and so on. Record the closing price of the T month of the Shanghai Composite Index as. Figure 1 is the monthly data trend chart since December 2001 after taking logarithm.

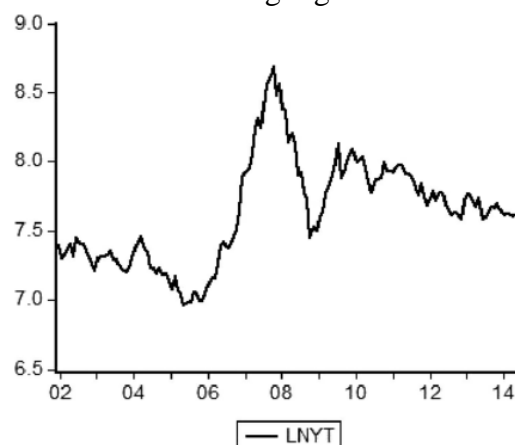


Fig.1 Monthly chart of Shanghai Composite Index

It can be roughly seen from Figure 1 that the time series obviously does not conform to the assumption of zero mean covariance, so ADF test is carried out on it. As can be seen from FIG. 2, the test T-statistic is -2.3389, which is greater than the critical value of 10% significance level. Therefore, the sequence has unit roots and is non-stationary, so the first-order difference is carried out. As can be seen from FIG. 3, the sequence after the first-order difference is determined as $d=1$.

3.2. The order is determined according to the stationary time series after the first difference

The B-J methodology believes that the corresponding ARMA model can be identified according to the autocorrelation function and partial autocorrelation function graph of the time series model. If the autocorrelation function (AC) and partial autocorrelation function (PAC) of a sequence are trailing, the sequence can be set as the PROCESS of ARMA (P, Q) [4]. Now let's observe the autocorrelation and partial autocorrelation graphs of sequences:

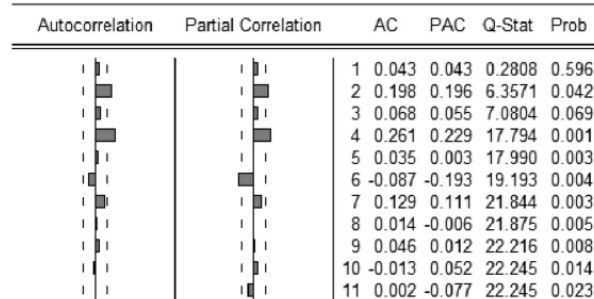


Fig.2 Autocorrelation and partial autocorrelation of sSE index sequence after first difference

It is clear from Figure 2 that both the autocorrelation and partial autocorrelation functions are trailing, and that they both fall dramatically from the second and fourth orders. So we could say p is equal to q is equal to 2 or p is equal to q is equal to 4. Through comparison, it can be seen that the probability of 0.1788 corresponding to the estimated coefficient value of AR (1) in ARIMA (2,1,2) model cannot pass the test at the significance level of 10%, so it is omitted. However, the parameters of ARIMA (4,1,4) model are completely significant at 1% level, which exactly conforms to the stationary condition of ARMA(p , q) process. Meanwhile, $AIC = -2.2032$ in ARIMA (4,1,4) < ARIMA(2,1,2) is $AIC = -2.1526$, which also conforms to ARIMA information criterion. In conclusion, the model is finally set as ARIMA (4, 1, 4).

3.3. Test for residuals

Next, the residual sequence of the ARIMA (4,1,4) model is tested for white noise. The test criteria for the test include Q statistics and the corresponding probability. The test results are shown in Figure 3. Can be seen from the figure 3 nearly all the absolute value of autocorrelation coefficients (AC) is less than 0.1, under the null hypothesis, the Q statistic subject distribution, which m is the largest lag, because the sample size $n = 152$, 152 material can make $m = 12$, the $Q - Stat = 12.956$, when $p = Q = 4$, when the significance level = 0.1, check the distribution table, $(12-4-4) = > Q - Stat = 12.956$ 13.277, will not be refused to residual sequence is independent of the original hypothesis, test through, accept the null hypothesis, namely the residual sequence is purely random white noise process.

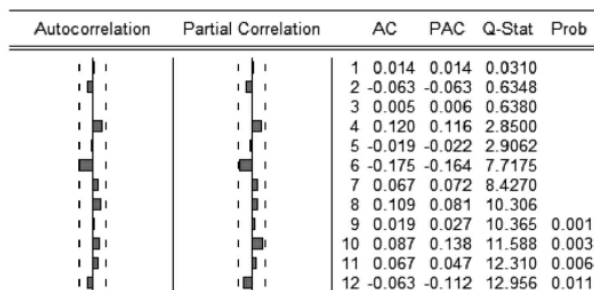


Fig.3 Output result of ARIMA (4,1,4) residual sequence test

3.4. Model fitting and prediction

The prediction of this model is to establish ARIMA model to predict the data of a period of time in the future based on the historical data of time series. The reason why ARIMA modeling method is widely used in various fields now is largely due to its success in prediction, especially in short-term prediction. In this paper, one-step forward static prediction is adopted, that is, each prediction is performed once, the real value is replaced by the real value, and the estimation interval is added to

carry out further prediction .Since the monthly data used in this paper is up to July 2014, the one-step forward static forecast can only predict the value of the first period, namely the closing price of August.

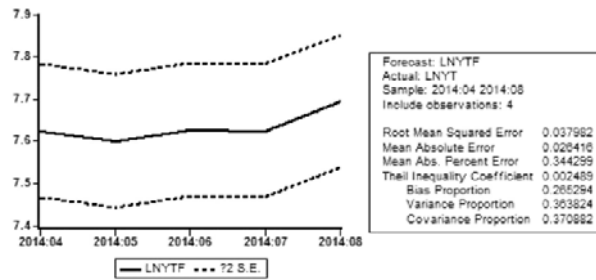


Fig.4 Forecasts from April 2014

In this model, the logarithm of the sample data was taken. In order to better observe the difference between the real value and the predicted value, we calculated the original value through scientific calculator.

4. Model analysis and outlook

4.1 Model analysis

From the above fitting and prediction, it can be concluded that the ARIMA model can make a better short-term prediction by using the static one-step forward method. from the established model, it can be seen that although the Shanghai Composite Index has risen and fallen since April 2014 to August 2014, it has shown a gradual upward trend [5]. This may benefit from the following reasons: first, China's economic fundamentals are still good. From the consumption point of view. The demand for services such as culture, education, medical care, old-age care and tourism is growing rapidly, while the development of new industries such as online shopping effectively promotes the release of consumption potential. From the perspective of investment, there is an extremely urgent demand for urban rail transit, environmental governance, urban drainage and rural infrastructure in China. secondly, the external environment tends to improve, and since the beginning of this year, the global economic recovery has been gradually strengthened in fluctuations. The improvement of the US economy will play an important role in other developed countries and even the global economy. With the strengthening of exchanges and cooperation among EU countries, the adverse effects of the European sovereign debt crisis are gradually being eliminated. Third, market expectations are getting better. Since the beginning of this year, China's inflationary pressure has continued to ease, and the overall price level (CPI) is within 3.5% of the regulation target. At present, entrepreneurs' confidence has generally rebounded and investment willingness has increased. Procurement activities accelerated.

5. Conclusion

Although ARIMA model has a good performance in the prediction of non-stationary time series, it is only feasible in the short-term prediction to a certain extent, but will show great limitations in the long-term trend, with large deviation of the prediction and distortion of the results. Moreover, this model only considers the characteristics of the time series itself, but ignores the influence of other more complex external factors, so it is difficult to accurately depict the long-term trend of the stock market with multiple changes. In general, this model can realize the short-term prediction of the market index, and then provide the basis for the majority of investors to make investment decisions.

For the integrated prediction of financial time series data, the prediction results of high frequency components, low frequency components and trend items are different in different machine learning algorithms and depth learning methods. As a whole, the prediction accuracy of different models for trend items is the highest, followed by low frequency components, and the prediction accuracy of high frequency components is relatively low. From the point of view of the integrated pre-test results, the integrated prediction accuracy is much higher than the direct prediction accuracy of CNN-GRU,

LSTM neural network and other algorithms for Shanghai Stock Index. Therefore, through the sub-quantity modeling of different wave motion characteristics, it is easier to dig the information hidden in the data, and then improve the pre-measurement precision of the financial time series data. In summary, from the theoretical and empirical point of view, this paper confirms the rationality and effectiveness of the application of deep learning algorithm in the field of financial time series data prediction, which has important theoretical and practical meaning for expanding the existing financial research methods. Therefore, it is still worthwhile for us to further explore by combining the research methods of traditional financial economics with artificial intelligence algorithms such as machine learning and deep learning to improve the cognitive ability of financial market operation rules and the pre-measuring power of financial risks.

References

- [1] Wu Xiaoqiang, Lu Wenlong. Trend forecast of Stock Price Index -- Time Series analysis based on SSE Index data [J].Financial Economy: Second Half, 2012(10):80-82.
- [2] Kuang Yun, Liang Zongjing. Forecast analysis of S&P500 Index based on ARIMA Model [J]. Modern Commerce and Trade Industry, 2012, 24(14) : 100-102.
- [3] Liu yun. ARIMA's Prediction of Shanghai Stock Index in China [J].Modern Commerce and Trade Industry, 2012, 24(16): 97-98.
- [4] Yan Min, Ho Chi Minh.An Empirical Analysis of Shanghai Composite Index based on ARIMA Model [J].Operation Manager, 2013, 21:076.
- [5] Xu Zhen, Li Xingye.Study on the prediction effect of Wavelet ARMA model and ARIMA Model on Shanghai Stock Exchange Index [J].Modern Business, 2012(30):32-33.