Research on the Price of Stock Index Futures with ARIMA Model

Shan-shan CONG, Jia-peng LIU, Jie LIU and Xiao-li ZHAO
School of Economics and Management, China Jiliang University, Hang Zhou, China
*Corresponding author

Keywords: Stock index futures, ARIMA model, Price forecast, Simulation trading.

Abstract. On April 16, 2010 Chinese first stock index futures listed in China Financial Futures Exchange Traded. It means that, Chinese financial markets ushered in the era of short. Stock index futures have become the focus of the vast majority investors. This paper focus on the price fluctuation of stock index futures since it appears on the market. Through the collection of the CSI 300 index futures closing prices, with data processing and analysis to establishment of ARIMA model. According to the model, the price of stock index futures is forecasted, and then simulating trading based on the predicted results, Simulation results show the accuracy of ARIMA model prediction. The results provide a reference for investors to provide a reference for investment activities.

Introduction

In recent years, the stock index futures have already developed into the important financial derivatives of the global capital markets which provide an indispensable and hedging instrument for investors. China has launched the Shanghai and Shenzhen 300 stock index futures contracts on April 16, 2010, which indicated the financial futures market of China entered into a new stage and was an historic milestone on the path to a market-driven economy.

Since CSI 300 Futures listing, the market development is very rapid, from the initial 1000 points up to now more than 3000 points, the futures market showing a thriving, and prosperous development of the situation. Because CSI 300 stock index futures and stock markets are closely linked, therefore the development of the futures market means that the stock market is also developed rapidly, but as we all known lots of investors are lack of confidence in the Chinese stock market. On the one hand stock index futures can do more or short; in the delivery of cash to settle accounts; so investors can avoid the volatility of the market funds, to a certain extent, CSI Futures play an important role in stabilizing the market. On the other hand the futures market also provides a place where transfers stock market prices risk to the futures market, risk from hedgers shifted to speculators. Because of the existence of the speculator who wants to pursuit profit Stock index futures market can smoothly realize the transfer of risk.

There are many researches on stock index futures at present. Box-Jenkins (1976) ARMA model is one of the most commonly used modeling and forecasting techniques. It is generally referred to as autoregressive moving average (ARMA) model. This methodology assumes that changes of time series data are related to its own past value. It creates autocorrelation regression model. KUMAR Manoj and ANAND Madhu using the ARIMA model to forecast the sugarcane production in India. The prediction results show the accuracy of using the ARIMA model to do forecast 2014. Zheng M and Miao J build ARMA and ARMA-GARCH models to forecast the returns of shanghai stock exchange composite index in 2012. Dongmei Xue established ARIAM (3, 1, 2) model to analysis on the total social investment in fixed assets 2010. It should be noted that, when we employing ARMA model, the basic assumption is that the future mode of a time series repeats its past pattern. It was also studied that the hypothesis can only be met for short-term, while forecasting over longer the
accuracy tends to deteriorate. So the Autoregressive Conditional Heteroskedasticity Model (ARCH) by Engle 1982, and the more generalized GARCH model by Bollerslev 1986’s, was proposed to explain the conditional variance. In China many studies have been conducted to explore and model the price behaviors of Futures markets, Yu Wei used volatility forecasting models to studied CSI 300 index futures. The results show that GARCH and its extended model are the weakest in the volatility of the Shanghai and Shenzhen 300 stock index futures. Shan-yin Bai was committed to forecast and analysis of Shanghai Stock Index Based on ARIMA model, and results show using the ARIMA model was effective. As for this paper we establish the ARIMA model of the CSI 300 index futures prices to predict.

Data and Methodology
The financial data used in this research is the daily closing price of the Shanghai and Shenzhen 300 index futures which sources Great Wisdom 365 Clients. The entire data set covers the period from 16th April 2010 to 31th March 2016. There are 1447 days of observations. The data set is divided into two periods: the first data period is from 16st April 2010 to 31st Dec 2015 (1388 days of observations) while the second period is from 1st Jun 2016 to 31st March 2016 (59 days of observations). The first period, assigned to in-sample estimation, is used to determine the specifications of the models and to estimate their parameters. The second period is reserved for out-of-sample evaluation and simulated transaction. The Shanghai and Shenzhen 300 stock index futures series recorded as $P$.

Types of ARMA Models
Autoregressive model methodology

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \cdots + \phi_p r_{t-p} + a_t, \quad t=1,2,3\cdots T$$

It’s called $P$ order autoregressive model, $\phi_r$ is regression coefficient, and $P$ is the autoregressive order, $\{a_t\}$ is white noise with mean 0 and variance $\delta^2$. $r_t$ is time series, in this paper it stands for the Shanghai and Shenzhen 300 index futures day closing prices.

Moving average model methodology

$$r_t = c_0 + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} + \cdots + \theta_q a_{t-q} + a_t, \quad t=1,2,3\cdots T$$

It’s called $q$ order moving average model, $\theta_q$ are the model coefficients, and $c_0$, $q$ are constants

Autoregressive moving-average model methodology

$$r_t = \phi_0 + \sum_{i=1}^{P} \phi_i r_{t-i} + a_t + \sum_{i=1}^{Q} \theta_i a_{t-i}, \quad t=1,2,3\cdots T$$

According to expression of ARMA (p, q) model, we know that if p=0, ARMA model is simplified as the q order moving average model MA (q); in contrast, if q=0, ARMA model is simplified as the P order autoregressive model AR (p).

Model Identification
The model identification is mainly determined what kind of models may be used and the ARIMA
model in this study follows the Box-Jenkins 1976 three-stage model identification strategy. According to the characteristic of partial autocorrelation function and autocorrelation function can make specific judgment.

<table>
<thead>
<tr>
<th>Model</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(p)</td>
<td>Tailing</td>
<td>p order truncation</td>
</tr>
<tr>
<td>MA(q)</td>
<td>q order truncation</td>
<td>Tailing</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>Tailing</td>
<td>Tailing</td>
</tr>
</tbody>
</table>

Table 1. Criteria for the identification of models.

From this table we can estimate the model and determine the order by observing the image of autocorrelation and partial autocorrelation coefficient. As we know from different sample the ACF and PACF are different, and the estimated value of the coefficient can only be in agreement with the theoretical trend. In the process of the actual identification model, the specific model form should be based on the information given by the autocorrelation and partial autocorrelation coefficient, after repeated experiments and tests, the final choice of the indicators are in line with the requirements of the model form.

There are two commonly used information criterion Akaike information criterion (AIC), Schwarz Bayesian information criterion (BIC), the expressions are as follow:

\[
AIC = -2 \ln(L) + 2k
\]

\[
BIC = -2 \ln(L) + k \cdot \ln(n)
\]

In formula (4) and (5) where, \(L\) is the maximum likelihood function, \(n\) is the sample size, and \(k\) is the number of parameters to be estimated in the model. With the Information criteria, we prefer to choose the smallest information criteria model.

In addition, other commonly used statistic index to measure the goodness of fit of stationary model is \(R^2\). The value of \(R^2\) is between (0, 1), the value larger means the model fit better. As far as possible to select the model with high goodness of fit, and information criterion is smallest. The specific forms of the model are determined by these indexes, and then the model parameters are estimated, getting the residual sequence. Testing the residual sequence of the model is white noise sequence or not to judge the model’s effectiveness. Ljung-Box statistic as follows:

\[
Q_m = T(T+2) \sum_{m} \frac{\rho^2}{T-s} \chi^2(m)
\]

The hypothesis is that the residuals series \(\epsilon_t\) is a white noise process. Statistic \(Q\) asymptotically obeys the chi square distribution. At a given significance level, if \(Q_m < \chi^2(m)\), we accept the hypothesis, and model checking is achieved. On the contrary, it is necessary to modify the model.

Financial time series forecasting is based on the premises that the financial market is valid and prices fluctuation history will repeat. So in this paper we assumed that the financial market is effective, publicly available information about the stock has been properly reacted in the price. We can analysis the historical data to find out the law. We can use this rule that base on the analysis the historical data to predict the financial market price. The main idea of making prediction is to minimize the variance of the forecast error, because the prediction error is a random variable, therefore the expectation of the prediction error is minimized:
\[ \text{min } E \left[ e^{\frac{T}{2}} \right] = \text{min} \left\{ \left( y_{T+L} - \tilde{y}_{T+L} \right)^2 \right\} \]  

(7)

On the above formula where \( T \) (L) is in the period of \( T \) to the next \( L \) step size of the predictive value, we can prove that \( \tilde{y}_{T+L} \) is the conditional expectation of the \( y_{T+L} \).

**Empirical Findings**

First of all, we observe the Shanghai and Shenzhen 300 index futures closing price as a whole trend, and the observation time is from 16st April 2010 to 31st Dec 2015.

First of all, we observe the Shanghai and Shenzhen 300 index futures closing price as a whole trend, and the observation time is from 16st April 2010 to 31st Dec 2015.

**Stationary Test**

First of all, we made a test on the stability of CSI 300 Futures daily closing data, as we all known the ADF test is applied frequently; the inspection results are as follow:

<table>
<thead>
<tr>
<th>Test critical values</th>
<th>1% level</th>
<th>5% level</th>
<th>10% level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey-Fuller test statistic</td>
<td>-1.590740</td>
<td>-3.434879</td>
<td>-2.863427</td>
</tr>
<tr>
<td>probability</td>
<td>0.4871</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the table we can see that the \( T \) test statistic value is -1.590740, the critical values were greater than the significance level of 1%, 5% and 10%. We have accepted the null hypothesis, that the original sequence is the unit root process. Because of the original time series are non-stationary, therefore, we should carry out the unit root test to the first order difference sequence of the CSI 300 Futures. The test results are as follow:

Figure 1. CSI 300 Futures daily time series.

From the observation of the line chart, we can see the overall trend of the CSI 300 Futures. The straight line is the mean of CSI 300 Futures, we can see that the fluctuation range is different, which shows that the variance of the sequence changes with time \( t \). Therefore, the preliminary determination of the CSI 300 Futures sequence is not smooth.
Table 3. CSI 300 Futures daily time series of first-order differential ADF test results.

<table>
<thead>
<tr>
<th></th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey-Fuller test statistic</td>
<td>-17.64352</td>
<td>0.0000</td>
</tr>
<tr>
<td>Test critical values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% level</td>
<td>-3.434879</td>
<td></td>
</tr>
<tr>
<td>5% level</td>
<td>-2.863427</td>
<td></td>
</tr>
<tr>
<td>10% level</td>
<td>-2.567824</td>
<td></td>
</tr>
</tbody>
</table>

As it can be seen from the above, the value of the T test statistic is -17.64352, the confidence level were less than the critical value of 1%, 5%, and 10%. Therefore, we have rejected the original hypothesis, the first order difference series of the CSI 300 Futures does not have the unit root and the sequence is stable.

**Model Set**

Since the original sequence of first-order differential is steady, so the d=1 in the model. The ARIMA (p.1.q) model can be established for first order differential sequence. The p and q value through the autocorrelation and partial autocorrelation to preliminary determination. By observing the autocorrelation and partial autocorrelation, the p model of the ARMA model was 4, and q was 3. After adjustment of different independent variable lag order p and random disturbance term lag order q, the parameters of the model are both significant. In addition to the 4 order of the independent variables, the coefficients of the model are significant at the 21% significant level. For this result may be explained by the ARIMA model’s order, the larger lag order the greater impact on the current increasingly. On the other word, with the increasing of the number of lags, the influence degree is more and smaller for the model. Through the model selection criteria determined the model for ARIMA (4, 1, 3) to predict the CSI 300 stock index futures price trend.

ARIMA (4, 1, 3) model’s parameters were estimated and the residual sequence should be tested to determine whether model is correct or not. The residual error of the estimated ARIAM (4, 1, 3) model is tested by the autocorrelation test.

After testing the residual series autocorrelation function of this model is in the 95% confidence interval, from the 1 order to 8 order self-correlation function of the corresponding probability are greater than the significance level 0.05. Therefore, we can’t refuse the original hypothesis indicates that the residual sequence is a white noise process, so it is suitable to determine the ARIMA (1, 4, 3) model to fit the CSI 300 stock index futures series.

**Forecast Analysis**

On the basis of ARIMA (4, 1, 3) model to predict the CSI 300 futures index prices. In this paper, we applied out of sample prediction and the period was from 1st Jun 2016 to 31st March 2016, reserving for out-of-sample evaluation and simulated transaction. We applied the ARIMA model to obtain the predicted value; according to the predicated value we calculated the absolute and relative error. Absolute error represents the absolute deviation between the predicted and the actual, and the relative error indicates the degree of deviation from the true value of the predicted value.

The following table is some statistical indicators that predicted by the ARIMA model. In this form hit rate refers to proportion of the forecast trend are same with the actual trend. And mean error rate indicates that the relative proportions of the average deviation of the predicted value and the actual value; and the ρ is representative of correlation coefficient between predictive value and true value.
Table 4. Correlation statistics index between real and predictive value.

<table>
<thead>
<tr>
<th>Relevant index</th>
<th>Hit rate</th>
<th>Mean error rate</th>
<th>ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical value</td>
<td>54.24%</td>
<td>1.5%</td>
<td>0.907</td>
</tr>
</tbody>
</table>

As it can be seen from the above table 4, the mean error rate is 1.5%, it indicates that predicted results are close to the actual value, and the error rate is within the acceptable range. In addition, the predictive model hit rate is 54.24%, it refers to that in the 100 forecast of 54 times forecast trend are same with the actual trend. As we all known, as long as the prediction hit rate is more than 50%, and it will be accepted. And then, the correlation coefficient between the predicted value and the true value reached 0.907, which shows that the predicted value is highly correlated with the actual value, indicating that the forecasted value is very close to the true value. At the same time the Theil inequality coefficient is 0.0096, which means that the ARIMA (4, 1,3) model fitting effect is remarkable.

The next picture is CSI 300 index futures actual closing price series trend and the predicted tendency which forecasted by the ARIMA model.

In the picture ‘rp’ is real price and the ‘fp’ is the forecast price. As it can be seen from the above figure3, the tendencies of the predictive and true value are basically same. Therefore we can predict the future trend to simulated trading of CSI 300 stock index futures.

Simulated Transaction

We can simulate CSI 300 Futures transaction according to the predicted value which forecasted by the ARIMA model. We formulate the trading strategy as: when the predicted stock index futures prices fell we do short; on the contrary, if the stock index futures prices rise to do more. In order to simplify the transaction we do the following assumptions:

- Investors are speculative traders, conducting fictitious transactions;
- Not consider charges and associated transaction costs
- Investors to buy or sell only one futures contract in each transaction

The investor holding period is one day, after the stock market closing that investor didn’t have unsettled contracts.
And after the simulation transactions we found that tendency revenue prediction correct acquired much more than prediction errors the loss. In addition, according to our forecast tendency of stock index futures to conducted the fictitious transactions. We make the relative index of the revenue or loss of total transaction as $R_t$. And the definition as follows:

$$R_t = \sum_{i=1}^{K} (m\%)$$  \hspace{1cm} (8)

In the above formula $m$ is the profit or loss of each transaction which based on the prediction data to do short and more transactions. And $N$ is the prediction interval length, in this article $N$ is Fifty-nine.

After simulated trading we got the average relative yield of simulated trading is 53.33%. It means that the ARIMA model to predict the CSI 300 index futures to simulate the transaction is successful. Therefore, investors can base on the results of this study and combined with the real situation in the market to conduct transactions of CSI 300 index futures.

Conclusions

According to the empirical analysis of the article, we found that using ARIMA model to predict the price fluctuation of CSI 300 index futures is feasible. And the simulation results are very close to the market real data, the results of the forecast include most of the information in the market. To a certain extent, it represents the overall trend of stock index futures and the forecasting results have some reference value for investors.

In the prediction of the actual use of ARIMA model, using for short term predicate the accuracy will be higher but the forecasting accuracy will be reduced with the extension of predicted time. From the above analysis, we can know that the estimated coefficient is not significant with the increase of the lag order; the main reason is that with the extension of the forecast time, the more uncertain factors that include futures market volatility, stock market fluctuations, interest rate fluctuations, the country’s macroeconomic policies and foreign economic policy will affect the price volatility of stock index futures.

By simulating of the prediction results show that the model is very successful in forecasting, therefore investors can refer to the prediction model established in this paper to build their own portfolios. But we have to remind investor that in this article the prediction is established on the market is effective and the financial market history will repeat itself. But we cannot remove the history will not repeat itself. When the market is affected by national policies, impacted of domestic and foreign economic situation and other external shocks, the market will fluctuate violently. And it is conceivable that this prediction will not be effective.

In this paper, there are some shortcomings; we do not take the transaction costs and commission into consideration in the simulation trading, if the transaction costs be considered more accurate results will be obtained.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China (71473235), Zhejiang Provincial Natural Science Foundation of China (LY16G030023) and the soft science key project of Zhejiang Province (2014C25032), Zhejiang Provincial Key Research Base of Management Science and Engineering, and Zhejiang Industrial Development Policy Key Research Centre of Philosophy and Social Science of Zhejiang Province.
References


Reference to a book:


Reference to chapters in an edited book:


Reference to some chapters in an edited book: