Analysis of CSI 300 Stock Index Futures Price Trend Based on ARIMA Model

Ya-ting ZHANG¹,* and Bo SUN²
¹Guangdong University of Foreign Studies, Guangzhou, 510420, China
²Research Institute of International Services Outsourcing, Guangdong University of Foreign Studies, Guangzhou, 510420, China

*Corresponding author

Keywords: ARIMA model, Index futures, Forecast.

Abstract. Stock index futures has become the major tools for investors to avoid risk or make speculative arbitrage. However, the leverage of stock index futures is large. If they cannot make a correct judgment on the trend of stock index futures, the loss will be great. This paper is based on this background and engages in intensive study. Using traditional and modern econometric methods respectively, the CSI 300 index futures were fitted. In addition, RMSE (root mean square error) was used to find the optimal model. After finding the model, we can put the trading strategy forward.

Introduction

For the stock market, we usually use fundamental analysis, technical analysis and other methods for investment analysis. But for stock index futures, we cannot get its rich basic information. In the past times, some scholars used M-GARCH model to analyze the information concluding in the futures market [1] (Lien D, Shrestha K., 2009). And some scholars studied the relationship between the growth of economic and stock price [2] (Fama, 1992). Besides, some scholars predicted stock index futures using regression model based on SVM [3] (Tong S, Koller D., 2002). But not many of them studied CSI 300 stock index futures using different time series analysis methods and margin debt for comparison. Here, the paper will use different kinds of time series analysis methods and SVM to analyze the future trend of CSI 300 stock index futures. After the analysis, we can put the trading strategy forward.

The paper’s possible innovation point: Firstly, the paper does not take the daily closing price series of CSI 300 stock index futures as the sample set. It selects the recent IF1703 daily closing price sequence as the sample set with highlighting the timeliness and practicality. Secondly, as for the methods, through comparing traditional and modern methods, the paper will get a better fitting model. After that, the paper can put forward an investment strategy. Thirdly, there is not many related researches on predicting CSI 300 stock index futures combining margin debt. Margin debt’s leverage does play a great role in the stock market. So the paper will make a further study using margin debt.

Empirical Research

Data Selection and Processing

(1) Data Explanation of Stock Index Futures

The paper studies the short-term trend of CSI 300 stock index futures and its trading strategies. This paper selects the daily closing price series of CSI 300 stock index futures (IF1703) from July 19, 2016 to January 12, 2017. 120 data are used as training sets to establish a model of time series. The test set is CSI 300 stock index futures closing price series data from January 13, 2017 to February 9, 2017. The paper will use data from test set to test the model. Besides, CSI 300 stock index futures will be called IFO sequence for short.

(2) Data Explanation of Margin Debt
Margin trading has a huge impact on the stock market. This paper makes a deep analysis of it from the perspective of M2. According to the prediction of M2 (referring to private equity fund: Lingfeng Capital), M2's growth will rise exponentially, as shown in figure 1. The growth of M2 in 2014 and 2015 looked high on the surface, which was actually below the level of the exponential growth should have. This shows that the growth of M2 in the future is huge. In addition to the normal supply, M2 also derive from credit expansion. To get enough M2, it requires support from leveraged margin funds. CSI 300 stock index futures is a standard contract on the basis of CSI 300 index. Margin trading has an important impact on the stock market, so the combination of margin debt and CSI 300 stock index futures is feasible.

As for the index of margin trade, we choose the margin debt. A refinancing deal means that investors can get leveraged funds as long as the payment of margin. Because the domestic mechanism does not allow securities to sell short. And the amount of margin is limited. The impact of margin trading on the stock market is less than the amount of financing. In order to reflect the impact of margin trading on the stock market better, this paper chooses the margin debt to establish the prediction of CSI 300 stock index futures.

(3) Data Processing

The closing prices of CSI 300 stock index futures are related to time significantly with a certain trend. To have a further test of stationarity of the sequence, the unit root test (Here is ADF test) is needed. As a result, the test doesn’t refuse the original assumption. So IFO sequence is non-stationary. Through the first orders operation to eliminate the non-stationary, the paper does ADF test again. And the T value is -5.495777, which means the new sequence (called DIFO sequence) is stationary.

Model Establishment

(1) Standard of Comparison

RMSE\(^4\) value (root mean square error) of each model will be compared. And the model with the minimum RMSE value will be chosen as the optimal model.

\[
\text{RMSE} = \sqrt{\frac{\sum(y_{i,t}-y_{t})^2}{n}}
\]

(2) Traditional Method Fitting

1) Time Series Analysis Method

A. Moving Average Method

Using Excel, the IFO sequence is fitted by simple moving average method\(^5\). And the interval is 5. It is found that the MSE is very large. The DIFO sequence is fitted by the simple moving average method, and the intervals are 5 and 3 respectively. Finally, according to the experience, the weight of weighted moving average method is obtained, and the weighted moving average method is used to fit the DIFO sequence, in which the interval is 5 and 3. As for the criteria for the selection of
intervals, the stock index futures trading in the working day, that is, 5 days a week. So the paper choose interval for 5. Besides, in order to obtain more accurate results, this paper also chooses the interval of 3 days for comparison.

The results are as followed:

<table>
<thead>
<tr>
<th>Methods and interval</th>
<th>MSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving average method for IFO sequence (interval: 5)</td>
<td>2022.64</td>
<td>44.97</td>
</tr>
<tr>
<td>Moving average method for DIFO sequence (interval: 5)</td>
<td>1094.11</td>
<td>33.08</td>
</tr>
<tr>
<td>Moving average method for DIFO sequence (interval: 3)</td>
<td>1147.08</td>
<td>33.87</td>
</tr>
<tr>
<td>Weighted moving average method for DIFO sequence (interval: 5)</td>
<td>1187.28</td>
<td>34.46</td>
</tr>
<tr>
<td>Weighted moving average method for DIFO sequence (interval: 3)</td>
<td>1243.06</td>
<td>35.26</td>
</tr>
</tbody>
</table>

From table 1, it shows that the DIFO sequence of simple moving average method has (interval: 5) the minimum MSE, RMSE. And the fitting degree is relatively high. The fitting formula is:

\[ \text{DIFO}_{t+1} = (\text{DIFO}_t + \text{DIFO}_{t-1} + \text{DIFO}_{t-2} + \text{DIFO}_{t-3} + \text{DIFO}_{t-4}) / 5 \]  

**B. Exponential Smoothing Method**

Exponential smoothing is a method to obtain the prediction results by weighted average of the observed values. The farther the time of observations is, the number of weights decreases exponentially. As we know, the IFO sequence only has a linear trend. But as the IFO sequence selection time is short, there is no seasonal component in them. Therefore, this paper uses the non-seasonal Holt linear trend index smoothing model, with SPSS.

The Alpha value of the exponential smoothing model is 0.998 and the P value is 0.00, which indicates that the horizontal coefficient is very significant. The trend Gamma value is 0.00005, which has no significance. So it can be concluded that the sequence has almost no trend feature.

The fitting formula is as follows:

\[ L_{t+1} = 0.998 \text{IFO}_t + 0.002 (L_t + T_t) \]  
\[ T_{t+1} = T_t \]  
\[ \text{IFO}_{t+1} = L_{t+1} + T_{t+1} = 0.998 \text{IFO}_t + 0.002 L_t + 1.002 T_t \]  

Lt represents the average demand; Tt represents the trend of growth; formula (3) is the smoothness of the time series’ trend; formula (5) is a smooth form of trend increment.

**C. ARIMA Model**

In the model test, through the first orders, the DIFO sequence is stable, namely, d=1. Then it’s time to determine the number of autocorrelation coefficient and partial autocorrelation coefficient, which are p and q. The autocorrelation (ACF) and partial autocorrelation (PACF) graphs are obtained through reviews. After observation, the paper finds that autocorrelation coefficient and partial autocorrelation coefficient are near 0. So there is no truncation and smearing. The p and q value cannot be included from the graph.

Therefore, this paper uses multiple test methods to verify the optimal model by combining AR (1), AR (2), MA (1) and MA (2) respectively. In this paper, we use the AIC criterion, that is, the minimum information criterion, to find p and q with the minimum AIC.
Table 2. The situation for AIC.

<table>
<thead>
<tr>
<th>AR lag</th>
<th>MA lag</th>
<th>AIC</th>
<th>AR lag</th>
<th>MA lag</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9.564654</td>
<td>2</td>
<td>1</td>
<td>9.673406</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>9.651116</td>
<td>1</td>
<td>1,2</td>
<td>9.687801</td>
</tr>
<tr>
<td>1,2</td>
<td>1</td>
<td>9.671641</td>
<td>2</td>
<td>1,2</td>
<td>9.637243</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>9.673245</td>
<td>1,2</td>
<td>2</td>
<td>9.615750</td>
</tr>
</tbody>
</table>

From table 2, when AR (1), AR (2), MA (1) and MA (2) are used together to regress, the AIC value is the smallest, which is 9.589112. Based on the sample data of 120, it is not suitable to make multi order lag. At the same time, the AIC of ARIMA (3, 1, 3) is 9.64. So the paper choose ARIMA (2, 1, 2) as the optimal model.

It is found that the coefficients of ARIMA (2, 1, 2) are significant, and the reciprocal roots of the lag polynomial are within the unit circle. And the modeling results of ARIMA (2, 1, 2) are obtained:

\[
DIFO_t = 0.221 DIFO_{t-1} - 0.937 DIFO_{t-2} + 0.189 e_t - 1 + 0.971 e_{t-2}
\]  \hspace{1cm} (6)

Then it’s the test of the residuals. Q statistics is large, which shows that the residual sequence is white noise series. So there is no correlation between the residuals. This model has extracted the information from the residual.

2) Regression Model

A regression model with margin debt, the error between fitting value and real value is not small. Besides, the RMSE is 54.667, which indicates the error is big. According to the number of fitting objects, it can be concluded that when the DW value is close to 2, the fitting effect is very good, and the autocorrelation of the original fitting object is small. Therefore, the generalized difference method is used to correct the model.

For the above regression, the residual sequence can be obtained. Through the first orders, make a regression equation \( E_t = 0.861698 E_{t-1} \) to make a model using the generalized difference method.

\[
IFO_t - 0.861698 IFO_{t-1} = a(1-0.861698) + b(RZYEO_t - 0.861698 RZYEO_{t-1}) + v_t
\]  \hspace{1cm} (7)

However, when the generalized difference method is used, the DW value satisfies the requirement, but the coefficient of the regression equation decreases. In order to improve the coefficient of determination to make the regression equation more accurate, the autocorrelation coefficient 0.861698 of the original model is adjusted. After debugging, the output of the adjusted generalized difference method is as follows:

![Figure 2. The result of generalized difference method.](image)
And the regression equation is:

\[ \text{IFO}_t - 0.685 \text{IFO}_{t-1} = 138.832 + 0.305 (\text{RZYEO}_t - 0.685 \text{RZYEO}_{t-1}) \]  

(8)

3) Modern Method Fitting-SVM

As the neural network model is complex and inconvenient to deal with, this paper adopts a relatively simple linear support vector machine model to fit IFO sequence. The correlation coefficient is as follows:

Table 3. The correlation coefficient.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Closing price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closing price</td>
<td>1</td>
</tr>
<tr>
<td>Opening price</td>
<td>0.969562</td>
</tr>
<tr>
<td>Highest price</td>
<td>0.988869</td>
</tr>
<tr>
<td>Lowest price</td>
<td>0.987192</td>
</tr>
<tr>
<td>Holding volume</td>
<td>-0.203352</td>
</tr>
<tr>
<td>Trading volume</td>
<td>0.062348</td>
</tr>
</tbody>
</table>

Although correlation coefficient for closing price and the change range of trading volume and holding volume is not high, these two variables indeed have an important impact on the stock index futures. So the paper still put them as independent variables. CSI 300 stock index futures’ daily opening price, high price, low price, holding volume range, trading volume range will work as independent variables. This paper uses Phyton to do model. While the fitting formula is as follows:

\[ \text{IFO sequence} = 0.0497 \times \text{opening price sequence} + 0.456 \times \text{highest price sequence} + 0.489 \times \text{lowest price sequence} - 0.0127 \times \text{holding volume range} + 0.0633 \times \text{trading volume range} + 0.0002. \]

According to the score from Phyton, \( R^2 \) is 0.91655, which means the fitting effect is good.

Model Evaluation

(1) Model Fitting and Prediction Effect

The fitting and prediction Effect for the above model is as follows:

Table 4. RMSE for the model.

<table>
<thead>
<tr>
<th>Name of model</th>
<th>RMSE (fitting effect)</th>
<th>RMSE (predict effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving average method for DIFO sequence (interval: 5)</td>
<td>33.08</td>
<td>21.839</td>
</tr>
<tr>
<td>Non-seasonal Holt linear trend index smoothing model</td>
<td>30.014</td>
<td>60.181</td>
</tr>
<tr>
<td>ARIMA(2, 1, 2)</td>
<td>28.7231</td>
<td>19.218</td>
</tr>
<tr>
<td>Generalized difference method for regression using margin</td>
<td>28.38679</td>
<td>66.25</td>
</tr>
<tr>
<td>SVR</td>
<td>0.91 (R^2)</td>
<td>77.2</td>
</tr>
</tbody>
</table>

From table 4, the fitting results show that the best model is ARIMA (2, 1, 2) and the regression using margin debt. As for prediction effect, the better model is DIFO sequence simple moving average method (phase 5) and ARIMA (2, 1, 2). In general, the optimal model is ARIMA (2, 1, 2) with relatively good fitting and prediction effect.

(2) Investment Strategy

As ARIMA (2, 1, 2) is the optimal model, the following investment strategy is put forward from it.

When the forecast value is higher than the previous day’s CSI 300 stock index futures closing price, take sell operation; otherwise, take the buy operation.

In order to facilitate the calculation, the assumption is to make an operation on each trading day, and the stock index futures will be liquidated before the end of the trading day. The specific operation is as follows:
The total profit for 15 trading days is 92.6 points. And the total amount is 92.6*300=27780 yuan. Assume that the fee is 5% (in the real life, the fee is not so high). The return rate of the investment strategy is 95%, which is much higher than that of the same period.

Conclusion

In this paper, the traditional and modern methods are used to predict the daily closing price series of CSI 300 stock index futures, and then select a relatively small RMSE model, namely the ARIMA model. According to the prediction results of the model, this paper puts forward the trading strategy. The operation strategy has a high profit, which shows the effect of the trading strategy is good.

The effect of ARIMA model is relatively optimal. This shows that traditional methods, such as ARIMA model, still have some advantages when fitting stock index futures, especially in a stable period with small fluctuation range. But there are some weaknesses in this model.

Firstly, the above model is basically linear. Although the nonlinearity of the CSI 300 stock index futures is not obvious. But this paper does not use the nonlinear method to explore the model more deeply.

Secondly, for investment strategy, it can also be discussed from the perspective of term arbitrage. But it needs to grasp more indicators, so this article does not put forward the trading strategy from more angles.

In this paper, we can make a more detailed study of stock index futures from the nonlinear method. At the same time, stock index futures can be other one. In the future, we can use different types of futures contracts to further explore.

Acknowledgments

This work was supported in part by a grant from Guangdong province science and technology planning project. (2013B040404009) and (2016B070704010), Natural Science Foundation of Guangdong Province (2017A030313432).

References


