# Research on China Market Index Based on ARMA-GARCH-COPULA

**Minhao Liu***

School of Economics, Shanghai University, Shanghai, China

*Corresponding author: lmhstcscr@shu.edu.cn

**Keywords:** ARMA, GARCH -Copula, Correlation

**Abstract:** This paper attempts to describe the joint and marginal distribution performance of China's market index through time series model combined with Copula. Combined with the stationary rate of the processing method and the indicators of the best time series model after stationary processing, the stationary method suitable for the domestic market is discussed. It is found that the fitting degree of t-Copula is the highest among the most basic Copula types.

## 1. Introduction

With the standardization and development of China's securities market, the stock plays an increasingly important role in the development of the whole national economy. It raises idle funds for the country and enterprises, optimizes the allocation of resources, and to some extent, the stock market is also a "barometer" of social and economic operation (2012)[1]. In recent years, forecasting methods based on time series analysis have been widely used in various fields. The assumption of normality in the original traditional model was too strong. This results in that the correlation structure between multiple assets is not a form of Gaussian multi-connection in the traditional sense. The Colupa function gradually enters people's field of vision because of its excellent ability to describe the correlation of the tail. Zhang et al. (2014) [2]and Laih et al. (2014) [3]used the Copula function to describe the relationship between multiple assets and individual assets in the Chinese market and the European market, respectively, which can better fit the real results. This paper attempts to model Shanghai stock index and Shenzhen stock index through a complete set of process, and describe their marginal trend and overall connection form from two aspects of yield and volatility.

## 2. Return modeling

### 2.1 Stability treatment method

In general research, the method used to evaluate the quality of the model mainly uses the Bayesian information criterion (BIC) value of the model fitting. But this method puts too much emphasis on the fitting effect in-sample. The financial market has changed drastically, and the out-of-sample nature must be considered. Due to the T+1 characteristics of the Chinese market, the difference in every other day is also a way to deal with it. And, as a promotion of the alternate day difference method, an additional processing method of the next day return rate is proposed, combined with the logic of the next day difference, to further stabilized the value of the time series in the form of the return rate. In order to simplify the research logic and steps, the research method of the best order BIC here only considers the Autoregressive (AR) model, and the order is determined according to the BIC value.

Table 1. Optimal order determination results of different difference methods

| | Norm-p | Bic | In-sample corr | Out-sample corr | Best order |
|---|---|---|---|---|---|
| Return | 1.876e-06 | -8.981 | 0.5598 | 0.1283 | 1 |
| Log-price | 2.955e-06 | -8.982 | 0.9741 | -0.684 | 1 |
| Divid-by-day | 0.003588 | -4.497 | 0.5223 | 0.2097 | 6 |
| Return-divid-by-day | 0.002089 | -8.782 | 0.5170 | 0.1931 | 6 |

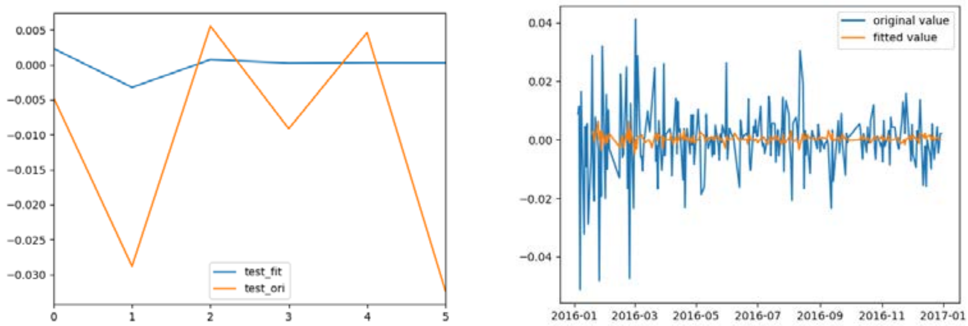


Figure 1. The best prediction performance of the fixed order sample after the rate of return processing

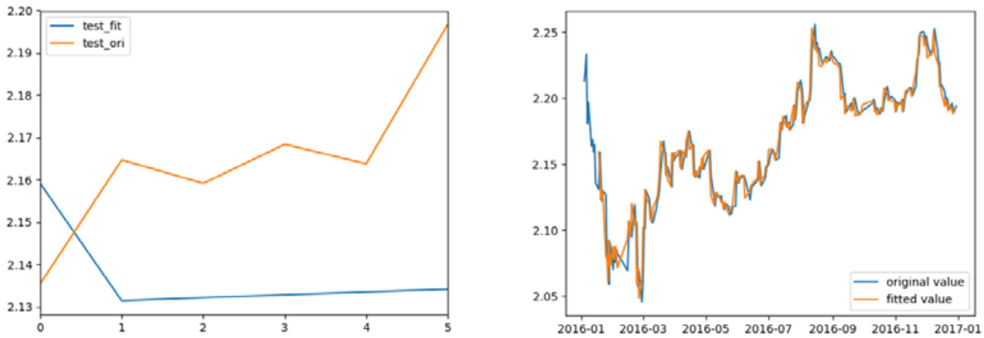


Figure 2. The best order sample after logarithmic price processing

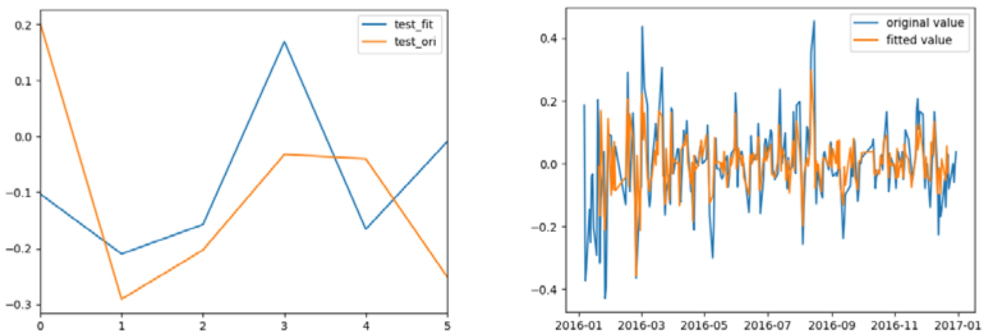


Figure 3. The performance of the best order sample in and out prediction after the difference processing every other day

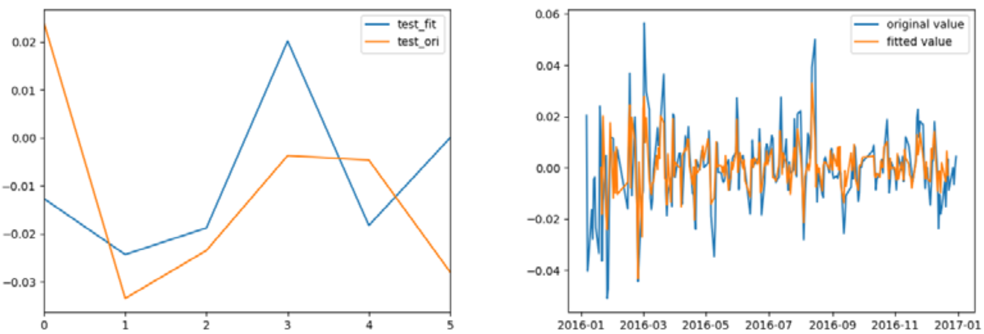


Figure 4. The best prediction performance of the best order sample after the differential treatment of the next day yield

The observation results found that if only the BIC criterion is used as the evaluation index, then the result of the selection should be the rate of return as the difference method. But this is obviously wrong, and its predictive ability is too weak. Consider using in-sample observations and fitting prediction correlations as evaluation indicators, using logarithmic prices for judgment is simply a perfect choice. Both the graphical representation and the specific values fully reflect the accuracy of this method. However, its predictive ability is not stable, and its out-of-sample performance shows the powerlessness of the AR process with the best order of order 1 obtained by this method. In comparison, the latter two methods in table 1 can better balance the properties of the inside and outside of the sample. In addition, the residuals obtained by almost all methods have not passed the normality test, indicating that it is difficult to completely extract the factors except for the use of AR modeling for time series. However, because the comparison method is more inclined to find a specific indicator as the criterion, Ljung-Box is used for the residual to judge whether there is autocorrelation. Although it is a common and effective method to check the fitting results, it is not considered due to too many variables. In consideration of the range, besides, it is difficult to have autocorrelation for the residual error obtained by a model selected by the order. Therefore, we will focus on the several indicators mentioned in table 1. Due to the excellent in-sample performance of logarithmic prices, we believe that using logarithmic prices is a better choice, and its out-of-sample performance can be modified by extending it to the Autoregressive Moving Average (ARMA) model.

### 2.2 Model order determination method

Model order determination is another very important link, which determines the explanatory power and predictability of a model. In this section, we are not limited to the AR model, but to the ARMA model to discuss how to determine the order of the model.

If BIC only contains the information in-sample, then the correlation coefficient outside-of-sample, as described in the previous section, is a good description method, which can be considered and introduced as a part of the discrimination. Therefore, the final candidate indexes for model order determination are AIC, BIC, correlation coefficient inside and outside of the sample and mean square deviation inside and outside of the sample.

Based on the data of Shanghai stock index from January 1, 2016 to December 31, 2016, the maximum order p and q are 7. Observe the index of each order of the model after the treatment of every other day yield.

Table 2. Evaluation indexes of ARMA model under different orders

| | In-sample corr | Out-sample corr | p | q | Aic | Bic | In-sample mse | Out-sample mse |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.529038 | 0.818264 | 0 | 1 | -1429.3 | -1418.96 | 0.166133 | 0.042316 |
| 1 | 0.526207 | 0.858208 | 0 | 2 | -1429.98 | -1416.19 | 0.165243 | 0.040238 |
| 2 | 0.434938 | -0.59367 | 1 | 0 | -1368.75 | -1358.41 | 0.190339 | 0.05662 |
| 3 | 0.529081 | -0.55571 | 1 | 1 | -1430.47 | -1416.69 | 0.165087 | 0.06757 |
| 4 | 0.478869 | -0.05049 | 2 | 0 | -1388.77 | -1374.99 | 0.181443 | 0.054317 |
| 5 | 0.537842 | -0.39173 | 2 | 1 | -1430.43 | -1413.19 | 0.164442 | 0.057188 |
| 6 | 0.525639 | -0.16791 | 2 | 3 | -1429.19 | -1405.06 | 0.163224 | 0.059048 |
| 7 | 0.570767 | -0.01137 | 2 | 4 | -1441.38 | -1413.81 | 0.158372 | 0.066032 |
| 8 | 0.531773 | -0.3032 | 2 | 5 | -1424.34 | -1393.32 | 0.163582 | 0.054951 |
| 9 | 0.5691 | 0.084432 | 2 | 6 | -1438.88 | -1404.41 | 0.157915 | 0.068362 |
| 10 | 0.484863 | 0.669936 | 3 | 0 | -1391.03 | -1373.79 | 0.179767 | 0.054789 |
| 11 | 0.536678 | 0.543238 | 3 | 1 | -1429.19 | -1408.51 | 0.16421 | 0.057171 |
| 12 | 0.576664 | 0.747179 | 3 | 3 | -1441.18 | -1413.61 | 0.158458 | 0.062883 |

Regardless of the index inside and outside of the sample, a certain degree of consideration is need. It's easier to think about an operation here. The top ten orders of each index were selected by group selection method.

Table 3. Ranking of index orders

| | In-sample corr | Out-sample corr | AIC | BIC | In-sample MSE | Out-sample MSE |
|---|---|---|---|---|---|---|
| 1 | 6, 7 | 4, 0 | 2, 4 | 0, 1 | 6, 6 | 2, 3 |
| 2 | 6, 6 | 5, 0 | 3, 3 | 1, 1 | 6, 7 | 2, 5 |
| 3 | 7, 5 | 1, 1 | 3, 5 | 0, 2 | 7, 5 | 7, 2 |
| 4 | 5, 6 | 0, 2 | 2, 6 | 2, 4 | 3, 5 | 1, 1 |
| 5 | 4, 3 | 2, 1 | 4, 3 | 3, 3 | 7, 4 | 0, 2 |
| 6 | 3, 3 | 0, 1 | 4, 4 | 2, 1 | 2, 6 | 4, 0 |
| 7 | 3, 5 | 4, 1 | 7, 5 | 3, 1 | 5, 6 | 5, 0 |
| 8 | 7, 3 | 3, 1 | 6, 1 | 4, 3 | 7, 3 | 0, 1 |
| 9 | 7, 4 | 2, 5 | 3, 4 | 3, 5 | 4, 4 | 2, 1 |
| 10 | 2, 4 | 7, 2 | 7, 3 | 2, 3 | 2, 4 | 3, 1 |
| 11 | 4, 4 | 5, 1 | 6, 6 | 2, 6 | 3, 3 | 4, 1 |
| 12 | 2, 6 | 6, 1 | 7, 1 | 6, 1 | 4, 3 | 5, 1 |

From Table 3, observing the BIC index, ARMA (0,1) is too simple, and the fitting effect in-sample is very poor. Similarly, there is the ARMA (0,2) process, which may be the result of randomness. And if we observe the correlation coefficient index in-sample, the higher order obviously dominates. However, the out-of-sample performance is too poor, and there is obviously over-fitting. Therefore, as mentioned above, the order should be determined by combining observations of multiple variables. At the same time, the ranks that meet the most conditions are (3,5) and (2,4) which appears 4 times in total. And because the fitting performance of (3,5) is better, the order ARMA(3,5) is set. Similarly, the Shenzhen Component uses similar rate of return modeling and ordering methods, and the final order is ARMA (3,5).

## 3. Volatility modeling

### 3.1 GARCH and Copula model

The GARCH model assumes that the conditional variance obeys a normal distribution, which is inconsistent with the results of many studies. Longin and Solnik (2001) [4]have proposed that when considering extreme events, one should no longer rely on typical symmetrical distributions, such as normal distributions, to build models. Ang and BekaAlt (2002) [5]used the mechanism transformation model to describe the joint behavior of market returns, and conducted empirical analysis. These evidences show that the cross-market correlation in a bear market is significantly higher than that in a bull market.

According to the Sklar (1973) theorem [6], a multivariate distribution function can be uniquely determined by the marginal distribution function and a specific Copula function. This is the theoretical basis on which we can simulate financial market data through the Copula function. There are many types of Copula functions, and the academia often uses the Archimedes generating function family to fit real financial data. Among them, Gumbel, Frank, and Clayton three types of generating functions are the most important. They are respectively characterized by head correlation, symmetry and tail correlation, which can well represent the actual correlation in capital assets. Among them, the Clayton generating function has received the most attention for its conformity with the tail correlation widely appeared in the financial market. Low (2013)[7] established the Vine-Copula function based on the Clayton generator and achieved very good performance.

Therefore, according to the residual error of the GARCH model, further fitting is performed through different Copula functions to observe which of the different Copula is more in line with the real situation of the domestic market.

### 3.2 Test of fitting effect

It is necessary to develop an index to evaluate the fitting results of different Copulas. The traditional fitting method is m-statistics chi square test. However, this method has some limitations, including how to divide the grid, adjust the degree of freedom, test results with the sensitivity of the division method, as multiple models are more likely to bring over fitting problems. Therefore, the Cramér–von Mises criterion is considered here.

In statistics, Cramér–von Mises criterion is used to judge the goodness of fit of cumulative distribution function, that is, the gap between $F^*$ and the given empirical distribution function $F_n$. It is also used as part of other algorithms, such as minimum distance estimation. The judgement is defined as

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \, \mathrm{d}F^*(x) \tag{1}$$

Using this distance, we can perform Cramér–von Mises test:

$$T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} [\frac{2i-1}{2n} - F(x_i)]^2 \tag{2}$$

In order to avoid over fitting, this paper used the parametric bootstrap method to sample the sample points repeatedly, and constructed the bootstrap parametric test quantity according to the method proposed by pesarin (2001)[8], the p value of repeated sampling is calculated:

$$p = \frac{(0.5 + \sum(T_b \geq T, b=1,..,N))}{N+1} \tag{3}$$

Where n is the number of self-service sampling. Where T and $T_b$ represents test statistics and self-help test statistics respectively. This ensures that the approximate p value is strictly between 0 and 1, which is sometimes necessary for further processing.

### 3.3 Empirical results

By observing the fitting results, we can find that the overall distribution is more symmetrical than the tail or head distribution. Because Gumbel describes the middle correlation, the significance of fitting is the lowest in symmetric distribution. Because the tail probability of t-distribution is higher, and it is consistent with the frequency of tail risk in the marginal distribution, and it is more inclined to synchronous occurrence of tail or top events, and the fitting significance is higher. Therefore, the Copula of multivariate t-distribution is more suitable to explain the real financial series.

Table 4. Fitting results of various Copula distributions

| | PARA | T | P-Value |
|---|---|---|---|
| Clayton | 6.009 | 0.13469 | 0.002041 |
| Frank | 20.495 | 0.034457 | 0.002041 |
| Gumbel | 5.9398 | 0.012003 | 0.2388 |
| Joes | 7.84 | 0.080483 | 0.002041 |
| Norminal | 0.96482 | 0.002041 | 0.3735 |
| T | 0.9658 | 0.0091967 | 0.5939 |

## 4. Conclusion

This paper deeply discusses the different effects of different stationary processing methods on time series research. Combined with the stationarity of the difference method and the indicators of the best time series model after stationary processing, the paper further discusses the effectiveness of every other day difference method in China's stock market, and tests its robustness. In order to improve the numerical robustness, we proposed to use the next day yield instead of the next day difference method. At the same time, we find the potential of logarithmic price and will continue to analyze it in the follow-up research. Then a more comprehensive order determination scheme was proposed. By introducing the correlation coefficient inside and outside of the sample, the mean square deviation inside and outside of the sample and other indicators, we could further quantify and objectively describe the fitting effect and prediction ability of time series inside and outside of the sample.

Through sorting screening and frequency counting, the order with better comprehensive performance was selected, and then the final screening was made according to the index priority and specific value. Then we used BIC value to rank the residual in GARCH volatility model. The empirical results show that GARCH (1,1) has the best fitting degree in the region with lower order. In the end, we find that the fitting results of Copula are more asymmetric than those of Copula. Finally, it is found that t-Copula has won the best fit with its thick tail.

## References

[1] Wu Xiaoqiang, LV Wenlong. Trend prediction of stock price index: time series analysis based on Shanghai stock index data. Financial economy (Academic Edition). 2012 (1): 2-3

[2] Zhang bangzheng, Wei Yu, Yu Jiang, et al. Research on multi market correlation and portfolio selection based on EVT vine Copula [J]. Management science, 2014 (3): 133-144

[3] Laih Y W. Measuring rank correlation coefficients between financial time series: A GARCH-Copula based sequence alignment algorithm[J]. European Journal of Operational Research, 2014, 232(2):375-382.

[4] Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–676.

[5] Ang, A., Bekaert, G., 2002. International asset allocation with time-varying correlations. Rev. Financ. Stud. 15, 1137–1187.

[6] Sklar A. Random variables, joint distribution functions, and Copulas[J]. Kybernetika -Praha-, 1973, 9(6):449-460.

[7] Low R K Y, Alcock J, Brailsford T, et al. Canonical Vine Copulas in the Context of Modern Portfolio Management: Are They Worth it?[J]. Journal of Banking & Finance, 2013, 37(8):3085-3099.

[8] Pesarin, F. Multivariate Permutation Tests: With Applications in Biostatistics. Wiley. 2001