Industrial Engineering and Innovation Management (2019) 2: 1-6 DOI: 10.23977/ieim.2019.21001
Clausius Scientific Press, Canada ISSN 2522-6924

Prediction Research on the Return Rate of CSI 300 index

Li Xinlei[a,*], Xin Shibo

Beijing Technology and Business University, Beijing, China
a. lixinlei9404@163.com
**corresponding author*

Keywords: GARCH model, CSI 300 index, return rate

Abstract: CSI (China Securities Index) 300 index is an indicator reflecting the overall market situation of China's stock market. The accurate prediction of CSI 300 index's yield is of great significance for investment analysis. With the deepening of the academic research on the CSI 300 index forecast, how to analyze and describe stock volatility and stock future yield has become one of the hottest topics. Scholars have found that the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model has good results in terms of financial yield volatility. Therefore, based on the GARCH model, this paper predicts the yield of CSI 300 index, and divides the collected data into two parts, one for the intra-sample prediction, the reasonable mean equation and the variance equation, and the other for the out-of-sample prediction. The GARCH(1, 1) model has been proven to be better for short-term forecasting of CSI 300 index yields.

1. Introduction

The report of the 19th National Congress emphasized deepening the reform of the financial system and promoting the healthy development of multi-level capital markets. The stock market is an important part of the capital market and is fully and effectively regulated according to law to protect the legitimate rights and interests of small and medium-sized investors. There are two major stock exchanges in the Chinese mainland, the Shanghai Stock Exchange and the Shenzhen Stock Exchange. With nearly 200 million investors and nearly 3,000 stocks, stock index volatility will affect stockholders' assets, corporate finance and economic growth. CSI 300 index is a cross-market index that measures the overall trend of the Shanghai and Shenzhen markets. It has an important role in characterizing the stock price fluctuations of the Shanghai and Shenzhen stock markets and reflecting the overall trend of the market. It is an important reference indicator for investors' investment. In recent years, with the continuous growth of China's stock market and the continuous improvement of the system, the number of investors has increased year by year. The volatility of CSI 300 index has attracted more and more attention from investors in China and around the world.

Stock index volatility seems to be erratic. The volatility and long memory of stock index yields prove the existence of stock index changes. With the establishment of ARCH (Autoregressive Conditional Heteroskedasticity model) and GARCH models in the 1980s, people gradually realized that the volatility of financial markets can be predicted and analyzed. Early scholars such as Yan and Li (2008)[1], Yang, Zhang and Li (2010)[2] mainly studied whether the yield difference series of stock indexes can apply the GARCH model and compare the predicted value with the actual value, and the short-term prediction effect is better than the long-term prediction. In recent years, such as Huang (2018)[3], Liu (2018)[4] and other scholars began to investigate the sampling frequency or use Monte Carlo simulation to further verify the model prediction effect.

The method used in the CSI 300 Index Forecasting Study transitions from the traditional capital asset pricing theory and the grey system model to the time series model comparison analysis. The application of the GARCH family prediction model is still less than other methods. The choice of evaluation models has gradually received the attention of scholars[5-7]. It is the current research trend to establish accurate mean and variance equations for CSI 300 index returns. Therefore, this paper considers the partial ARMA model of the mean equation, and compares the variance equations with the normal distribution, the ARCH family and the GARCH family model, aiming to make a more accurate prediction of CSI 300 index yield.

2. Establishment of the Volatility Model

ARCH or GARCH models are generally used for yield series with sharp and thick tail features. The establishment process includes steps such as acquiring data, drawing time series, stationarity test, pure randomness test, and information criterion ordering. The model is built according to the construction steps of the GARCH model, and part of the data is reserved for the test of the model fitting effect to ensure the accurate prediction ability of the model.

2.1. Data Acquisition

Download a total of 2000 closing indices from April 12, 2011 to July 28, 2019. The index sequence is denoted as $\{X_t\}$, the yield series is recorded as $\{R_t\}$, and the rate of return is obtained from the log yield formula.

$$R_t = \ln X_t - \ln X_{t-1}$$

Based on these data, the last 20 data (nearly one month) are listed as comparative data. The Epsis6.0 software is used to analyze the time series of the previous 1980 index yield series, and the GARCH model is used to predict the last 20 data. Compared with the actual value, the evaluation of the fitting effect is given.

2.2. Timing Diagram and Normality Test of the Yield Series

The timing chart of the yield series is shown in Fig.1. It can be seen from the timing diagram that the Shanghai-Shenzhen 300 index yield series fluctuates around the mean (about 0), and fluctuates within a certain range except for a few extreme values. There is no trend about time, but there is volatility clustering, sometimes concentrated fluctuations are intense, and sometimes concentrated fluctuations are flat. The timing diagram meets the volatility establishment requirements. In order to investigate the distribution characteristics of the rate of return, a normality test is performed. According to the JB value of 2413, the skewness value is approximately 0, and the kurtosis value is much larger than 3, the yield is not subject to the normal distribution, and there is a sharp peak-tail symmetry feature.

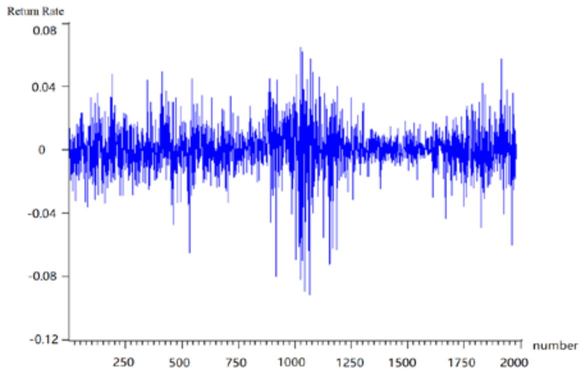

Figure 1: CSI 300 index yield timing chart.

2.3. Stationarity Test

In order to determine that the CSI 300 index yield series does not have a random trend or determine the time trend, to avoid the pseudo-regression problem, a stationarity test is needed. Using the ADF test, as shown in Table 1, the ADF test was passed without the constant term and the trend term, and the Shanghai-Shenzhen 300 index yield sequence was stable. In addition, according to the pure randomness test, the ACF has a tailing phenomenon with decreasing order, and there is also a tailing phenomenon in PACF. It can be known that the yield sequence may be an ARMA type time series.

Table 1: ADF test value and each variable t test under the NONE form.

Augmented Dickey-Fuller test statistic			t-Statistic	Prob
			-43.23977	0.0001
Variable	Coefficient	Std. Error	t-Statistic	Prob
R(-1)	-0.971935	0.022478	-43.23977	0.0000

2.4. Establish ARMA(1, 1) Model for the Mean Equation

The mean equation part ARMA(p, q) is determined according to the information criterion, wherein the ARMA(1, 1) model is simple and the AIC value is small, and the regression results are better as shown in Table 2. So, establish ARMA(1, 1) for off-sample comparison and ARCH and GARCH model was established.

Table 2: ARMA model regression results of yield.

Variable	Coefficient	Std. Error	t-Statistic	Prob
AR(1)	0.1240	0.0165	7.5295	0.0000
MA(1)	-0.1366	0.0223	-6.1285	0.0000

2.5. Examination of the ARCH Effect

The ARCH effect of the residual of the mean equation is tested. First, the pure randomness test of the residual square is shown in Table 3. The Q test results show that there is a significant autocorrelation of the residual square. Then the LM test is shown in Table 4. Both prove the existence of the ARCH effect, and the ARCH or GARCH family model can be established.

Table 3: Pure randomness test of residual squared.

Order	AC	PAC	Q-Stat	Prob
1	0.206	0.206	84.047	0.000
2	0.233	0.199	191.56	0.000
3	0.241	0.176	307.07	0.000
4	0.206	0.111	391.32	0.000
5	0.178	0.066	454.24	0.000
6	0.131	0.011	488.49	0.000

Table 4: LM test of residual squared.

Variable	Coefficient	Std. Error	t-Statistic
C	0.000220	2.65E-05	8.279134
AR(1)	0.110387	0.022396	4.928946
AR(2)	0.151076	0.022237	6.793994
AR(3)	0.161792	0.022236	7.276061
AR(4)	0.110974	0.022395	4.955312
AIC	-12.17542	BIC	-12.16126

2.6. Establishment of the Volatility Model

After the above test, it is proved that the Shanghai-Shenzhen 300 index yield can be described by the ARCH or GARCH model. Using EVIEWS6.0 software, the mean equation is set to the ARMA(1, 1) variance part, and the error distribution is based on the normal distribution and the t distribution to try to establish the ARCH and GARCH models, as shown in Table 5. According to the parameter t-test and information criterion, the regression results of ARCH(3) and GARCH(1, 1) based on t-distribution are ideal. The GARCH(1, 1) regression results are shown in Table 6. The R-square is small, indicating that the volatility information extraction is sufficient.

Table 5: ARCH and GARCH model regression results.

Model	Error distribution	Parameter test	AIC
ARCH(1)	normal	pass	-5.652
ARCH(2)	normal	no	——
ARCH(3)	normal	pass	-5.732
ARCH(4)	normal	pass	-5.752
GARCH(1,1)	normal	no	——
GARCH(1,2)	normal	no	——
ARCH(1)	student's t	pass	-5.830
ARCH(2)	student's t	pass	-5.851
ARCH(3)	student's t	pass	-5.870
ARCH(4)	student's t	no	——

GARCH(1,1)	student's t	pass	-5.950
GARCH(1,2)	student's t	no	——

Table 6: GARCH(1, 1) regression results.

	Variable	Coefficient	Prob
Mean equation	AR(1)	-0.980136	0
	MA(1)	0.989605	0
Variance Equation	C	1.05E-06	0.029
	RESID(-1)^2	0.058195	0
	GARCH(-1)	0.942278	0

3. Sample Out-Of-Sample Test

Based on the ARCH(3) and GARCH(1, 1) models, the data of the last 20 days are estimated and compared with the actual values. As shown in Table 7, the deviations between the predicted values and the true values of the first 5 steps of the two models are shown. Square case. Table 8 evaluates the fitting effects of the two models from the mean of the square of the deviation (ie, the mean square error). According to the results of Table 8, the GARCH(1, 1) model is smaller than the average deviation square of the ARCH(3) model within 10 days, and it is more reliable for CSI 300 index yield trend prediction. On the other hand, with the extension of time, the deviation of the correctness of the prediction direction of the GARCH model becomes larger, and the prediction of the model will be interfered by various variables in the long-term and produce a non-negligible error, so the short-term result will be better.

Table 7: Comparison of predicted and true values in the first 5 days.

Actual rate of return	ARCH(3)		GARCH(1,1)	
	Forecasted rate of return	Square of deviation	Forecasted rate of return	Square of deviation
-3.13E-03	7.03E-05	4.94E-09	1.29E-05	1.67E-10
6.13E-04	6.98E-05	4.87E-09	-1.27E-05	1.61E-10
-9.24E-03	6.94E-05	4.81E-09	1.24E-05	1.54E-10
-3.78E-04	6.89E-05	4.75E-09	-1.22E-05	1.48E-10
-9.01E-03	6.85E-05	4.69E-09	1.19E-05	1.43E-10

Table 8: Evaluation of fitting effect.

Mean square error	ARCH(3)	GARCH(1,1)
2days	4.971E-06	4.741E-06
5days	3.887E-05	3.842E-05
10days	1.579E-04	1.590E-04
15days	1.987E-04	2.000E-04
20days	1.609E-04	1.619E-04

4. Conclusions and Recommendations

Through the analysis and prediction of CSI 300 index, the research shows that the GARCH(1, 1) model can predict its future short-term rate of return to a certain extent, but if the model is used to predict the long-term trend, the prediction effect will be greatly reduced. This is because for long-term trends, it will be affected by a combination of factors, such as industry changes, corporate information, and policies. These factors may have a large impact on long-term forecasts based on the original time series model. In the short-term forecast, these factors have not changed much, so the short-term forecasting effect is better than the long-term forecast.

Through the above research, in the current stock market with multiple factors, it is necessary to find a reasonable model that conforms to the current market and keep updating, in order to use the time series model to guide investment decisions. It should be noted that the current reasonable forecasting model may generate large errors when the market conditions change in the future. Therefore, it is necessary to adjust the forecasting model according to the market environment in time to match the current market environment to obtain a more reliable short-term forecast result. At the same time, when using the GARCH(1, 1) model to predict CSI 300 index, the short-term forecast results will be significantly better than the long-term forecast results. Therefore, investors need to be more cautious if they use the long-term prediction results of the model as a reference.

Acknowledgment

This work was supported in part by a grant from Research on Dynamic Prediction Mechanism of Internet Public Opinion in the Background of Big Data (18YJZH203).

References

[1] YAN Dingqi and LI Yufeng. (2008) Prediction of Volatility of CSI 300 index Based on GARCH Family Model. Journal of Lanzhou Jiaotong University, 92-95.

[2] Yang Erpeng, Zhang Desheng and Li Wenjing. (2010) Study on the Statistical Prediction of the Prosperity of CSI 300 index Based on GARCH Model. Journal of Yangtze University (Natural Science Edition), 7, 151-153+388.

[3] Huang Xuan and Zhang Qinglong. (2018) Analysis and Prediction of Volatility of CSI 300 index Based on ARMA-GARCH Model. China Price, 44-46.

[4] Liu Haoyu. (2018) Stock Market Volatility Forecast Based on GARCH Model——Taking CSI 300 index as an Example. Journal of Wuhan Commercial College, 32, 53-59.

[5] Sun Rong and Wu Tianwei. (2011) Using Grey System Model to Predict the Change of CSI 300 index. Finance and Accounting Monthly, 65-67.

[6] Liu Dongmei and Cai Yanping. (2016) Prediction of Shanghai and Shenzhen Composite Index Based on Markov Chain. Shang, 183.

[7] Basel M A, Awartani and Valentina Corradi. (2015) Predicting the volatility of the S&P-500 stock index via GARCH Models: The role of asymmetries. International Journal of Forecasting, 167-183.