International Journal of Accounting and Finance Studies
Vol. 4, No. 2, 2021
www.scholink.org/ojs/index.php/ijafs
ISSN 2576-2001 (Print) ISSN 2576-201X (Online)

Original Paper

An Empirical Analysis on the Volatility of Return of CSI 300 Index

Liang Jinling[1] & Deng Guangming[1,2]

[1] College of Science, Guilin University of Technology, Guilin Guangxi, China

[2] Institute of Applied Statistics, Guilin University of Technology, Guilin Guangxi, China

Received: September 20, 2020 Accepted: October 7, 2020 Online Published: October 18, 2021

doi:10.22158/ijafs.v4n2p1 URL: http://dx.doi.org/10.22158/ijafs.v4n2p1

Abstract

In order to better observe the trend of the stock market, this paper selects the daily closing price data of CSI 300 index from April 12, 2016 to September 30, 2021, and makes an empirical analysis on the logarithmic return of CSI 300 index. It is found that: (1) the return series of the CSI 300 index shows the statistical characteristics of peak, thick tail, bias, asymmetry and persistence. The ARMA (2,3) model can effectively fit the yield series and predict the future trend to a certain extent. (2) The residuals of ARMA model show obvious cluster effect and ARCH *effect (conditional heteroscedasticity). GARCH (1,1) model can better fit the conditional heteroscedasticity, so as to eliminate the* ARCH *effect. (3) By constructing GARCH (1,1) model, it is found that the sum of* ARCH *term coefficient and GARCH term coefficient is very close to 1, indicating that GARCH process is wide and stable, the impact on conditional variance is lasting, and the market risk is large, that is, the impact plays an important role in all future forecasts.*

Keywords

CSI 300 index, Logarithmic rate of return, ARMA model, GARCH model

1. Introduction

Volatility aggregation often occurs in financial time series such as stock and futures return series. Volatility aggregation can well describe the volatility of financial data in this time period and the next time period. Risk management and control through the volatility aggregation modeling of stock and futures has become a research hotspot in the financial field (Campbell, Huisman, & Koedijk, 2001; Hu & Ge, 2021; Zhang & Zeng, 2021, Luo & Zou, 2020).

In recent years, many scholars have modeled and analyzed the CSI 300 index. Xu Jing (Xu, 2020) took the CSI 300 index as an example to study the stock volatility based on GARCH model. Firstly, the

ARMA model is fitted to the daily log return series of the CSI 300 index, and then the GARCH model is fitted to the conditional heteroscedasticity. It is found that the CSI 300 index stock market has a significant leverage effect. Under the information impact of the same intensity, the impact of negative news on the abnormal fluctuation of the stock market is greater than that of positive news. Li Haohua, Zhang Xiaoqiang and Chen Ying (Li, Zhang, & Chen, 2018) studied the impact of stock index futures on stock trading behavior and further studied the impact of stock index futures on stock market volatility by analyzing the amount of information between Shanghai and Shenzhen 300 index, spot index and constituent stocks of Shanghai and Shenzhen 300 index. Based on the CSI 300 index, Cao Sen and Zhang Yulong (Cao & Zhang, 2012) established the GARCH family model, analyzed the volatility of the return of the CSI 300 index, and reached the relevant conclusions on the impact of the CSI 300 index on the spot market. Zheng Ping (Zheng, 2014) selected the CSI 300 index as the representative of China's stock market risk, and made an empirical analysis under t distribution and CED distribution by using GARCH, TGARCH and EGARCH models. By observing the logarithmic rate of return and VaR valuation of CSI 300 index, it can be found that the VaR values simulated and calculated by the five GARCH models through the back-test test have a large fit, and the mean value can accurately predict the risk of China's stock market represented by CSI 300 index. Wu Liuhong, Zhang Xuedong and Wang Leilei (Wu, Zhang, & Wang, 2012) Based on the empirical analysis of CSI 300 stock index futures. Based on the TGARCH and EGARCH model test, this paper makes an empirical analysis on the two sub sample data respectively, and comes to the conclusion that the introduction of stock index futures increases the asymmetric effect of the spot market. Ma Guoteng and Zhao Yanyun (Ma & Zhao, 2010) used the tarch model of asymmetric effect to analyze the variation characteristics of the yield fluctuation of Shanghai and Shenzhen 300 index. The results show that the factors affecting the strong variation of the yield of Shanghai and Shenzhen 300 index are the lag of orders 4 and 15; There is a strong conditional heteroscedasticity in the yield of CSI 300 index, and tarch model can better eliminate the conditional heteroscedasticity. Compared with the good news, the bad news has a greater impact on the CSI 300 index, and there is an obvious leverage effect on the whole.

The above literature provides feasible ideas for studying the trend of CSI 300 index and stock market. This paper selects the relevant data of CSI 300 index from April 12, 2016 to September 30, 2021 for empirical analysis to analyze and describe the fluctuation and future trend of stock market.

2. ARMA Model

The model with the following structure is called autoregressive moving average model, abbreviated as ARMA(p，q) (Wang, 2015);

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (1)$$

If $\varphi_0=0$, the model is called centralized ARMA(p, q) model. By default, the centralized ARMA(p, q) model can be abbreviated as

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

The delay operator is introduced, and the ARMA (P, q) model is abbreviated as:

$$\Phi(B) x_t = \theta(B) \varepsilon_t \quad (3)$$

Where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is an autoregressive coefficient polynomial of order p, $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ is a moving average coefficient polynomial of order q.

When q = 0, ARMA(p, q)ARMA (P, q) model degenerates into AR(p) model; When p = 0, ARMA (p, q) model degenerates into MA (q) model. Therefore, AR (p) model and MA (q) model are actually special cases of ARMA (p, q) model, which are collectively referred to as ARMA model. The statistical properties of ARMA (p, q) model are the organic combination of AR (p) model and MA (q) model.

2.1 Stationary Test

(1) When selecting the data of time series, the stationarity test should be carried out. The test methods usually include: time series diagram, autocorrelation diagram, ADF test, PP test and kpss test. The first two are to judge whether it is stable from the intuitive feeling, while ADF test, PP test and kpss test (i.e. unit root test) are more accurate judgments from the perspective of statistical theory, That is, the T statistics under a given significance level are greater than the T statistics of ADF test and PP test, that is, when the p value is less than 0.05, the sequence is stable, otherwise it is non-stationary. Kpss on the contrary, its original assumption is that the sequence is a stationary sequence, and the T statistics at a given significance level are greater than kpss test statistics, that is, when the p value is greater than 0.05, the sequence is stationary.

(2) When the sequence is non-stationary, that is, it has a certain trend or cycle, it should be processed by difference to eliminate its non-stationary.

$$\text{First order difference: } \nabla x_t = x_t - x_{t-1} \quad (4)$$

$$\text{Second order difference: } \nabla^2 x_t = \nabla x_t - \nabla x_{t-1} \quad (5)$$

$$P-\text{order difference: } \nabla^P x_t = \nabla^{P-1} x_t - \nabla^{P-1} x_{t-1} \quad (6)$$

2.2 Pure Randomness Test

Pure randomness test is also called white noise test. Its purpose is to test whether the sequence is a pure random sequence. When a sequence is a white noise sequence, strictly speaking, there is no correlation between its sequence values, but due to the influence of space and other factors, the sample autocorrelation coefficient is not significant, which is 0, which is statistically expressed as:

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0, \forall m \geq 1$$

$$H_1 : \text{ at least one } \rho_k \neq 0, \forall m \geq 1, k \leq m$$

The test statistics of pure random sequences include *Q* statistics and *LB* statistics. In practical application, *Q* statistics has a good test effect in the case of large samples (where n is large), but it is not accurate in the case of small samples. *LB* statistics is a modification of *Q* statistics. They are often collectively referred to as Q statistics, which are recorded as *QBP* Statistics (*Q* statistics of box and price) and *QLB* Statistics (*Q* statistics of box and Ljung). When the *Q* statistic is greater than the quantile or the p value of the statistic is less than α When, the confidence level can be 1- α Reject the original hypothesis and consider the sequence as a non white noise sequence; Otherwise, the original hypothesis cannot be rejected and the sequence is considered as a pure random sequence.

2.3 Model Identification

After calculating the values of sample autocorrelation coefficient and partial autocorrelation coefficient, it is necessary to select an appropriate ARMA model to fit the observed value sequence according to their properties. In fact, this process is to estimate the autocorrelation order and moving average order according to the properties of sample autocorrelation coefficient $\hat{\rho}$ and partial autocorrelation coefficient $\hat{\phi}$. Therefore, the process of model identification is also called model order determination process.

The basic principles of ARMA model order determination are shown in the table below:

Table 1. Summary of Model Order Determination

$\hat{\rho}_k$	$\hat{\phi}_{kk}$	Model order determination
Trailing	*p*-order truncation	AR(*p*)model
q-order truncation	Trailing	MA(*q*)model
Trailing	Trailing	ARMA(*p*， *q*)model

However, in practice, this order determination principle has certain difficulties in operation. Due to the randomness of the sample, the correlation coefficient of the sample will not show the perfect situation of theoretical truncation. The autocorrelation coefficient or partial autocorrelation coefficient of the sample that should be truncated will still show small value oscillation. At the same time, because stationary time series usually have short-term correlation, with the delay order k→∞, and will decay to near zero for small value fluctuation. Therefore, the order determination of the model largely depends on the subjective experience of analysts. However, the approximate distribution of sample autocorrelation coefficient and partial autocorrelation coefficient can help inexperienced analysts make reasonable judgment as much as possible.

2.4 Parameter Estimation

After selecting the fitting model, the next step is to determine the caliber of the model by using the observations of the sequence, that is, to estimate the value of the unknown parameters in the mode $x_t = \mu \frac{\theta_q(B)}{\Phi_p(B)} \varepsilon_t$, $\varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$, $\Theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, $\Phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, There are $p+q+2$ parameters to be estimated： $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \mu, \sigma_\varepsilon^2$.

Parameter μ Is the sequence mean. Generally, the moment estimation method is used to estimate the overall mean with the sample mean to obtain its estimated value:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{7}$$

For the centralization of the original sequence, there are

$$y_t = x_t - \overline{x} \tag{8}$$

The original $p+q+2$ parameters to be estimated are reduced to 为 $p+q+1$： $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_\varepsilon^2$, there are three estimation methods for these $p+q+1$ unknown parameters: moment estimation, maximum likelihood estimation and least square estimation.

2.5 Model Significance Test

After the caliber of the fitting model is determined, the fitting model must be tested.

Model test is mainly divided into model significance test and parameter significance test.

The significance test of the model is mainly to test the effectiveness of the model. Whether a model is significantly effective mainly depends on whether the information it extracts is sufficient. A good fitting model should be able to extract almost all the sample related information in the observed value series. In other words, the fitting residual term will no longer contain any relevant information. That is, the residual sequence should be a white noise sequence. Such a model is called a significantly effective model.

On the contrary, if the residual sequence is a non white noise sequence, it means that the relevant information in the residual sequence has not been extracted, which indicates that the fitting model is not effective enough, and it is usually necessary to select other models for re fitting.

Therefore, the significance test of the model is the white noise test of the residual sequence. The original and alternative assumptions are:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0, \quad \forall m \geq 1$$
$$H_1: \text{at least one } \rho_k \neq 0, \quad \forall m \geq 1, k \leq m$$

The test statistic is LB (Ljung-Box) test statistics：

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \sim \chi^2(m), \quad \forall m > 0 \tag{9}$$

If the original hypothesis is rejected, it means that there is still relevant information in the residual sequence, and the fitting model is not significant. If the original hypothesis cannot be rejected, the fitting model is considered to be significantly effective.

The significance test of parameters is to test whether each unknown parameter is significant non-zero. The purpose of this test is to simplify the model.

If a parameter is not significant, it means that the influence of the independent variable corresponding to the parameter on the dependent variable is not obvious, and the independent variable can be eliminated from the fitting model. The final model will be represented by a series of independent variables with significantly non-zero parameters.

3. GARCH Model

3.1 Cluster Effect

In the macroeconomic and financial fields, we can often see time series with the following characteristics: after eliminating the influence of deterministic non-stationary factors, the fluctuation of residual series is stable in most periods, but it will continue to be large in some periods and small in some periods, showing a cluster effect (Wu & Liu, 2014).

People usually use variance to describe the fluctuation of the sequence. Cluster effect means that the variance of the sequence is basically homogeneous in the whole observation period of the sequence, but the variance is significantly different from the expected variance in a certain period or several periods. At this time, we need to introduce conditional heteroscedasticity model.

3.2 ARCH Test

To fit the ARCH model, ARCH test is needed first. ARCH test is a special heteroscedasticity test. It not only requires the sequence to have heteroscedasticity, but also requires that this heteroscedasticity is caused by some autocorrelation, which can be fitted by the autoregressive model of residual sequence. The two commonly used statistical methods of ARCH test are portmanteau c test and LM Test.

1). Portmanteau Q Test

In 1983, mold and l proposed portmanteau Q statistical method to test the autocorrelation of the square sequence of residuals. Now it is the statistical method of ARCH test. The construction idea of this test method is that if the variance of the residual sequence is non-homogeneous and has cluster effect, the square sequence of residuals usually has autocorrelation. Therefore, the variance non-homogeneous test can be transformed into the autocorrelation test of the square sequence of residuals.

The assumption of portmanteau Q test as

H_0: Residual square sequence pure random (Homogeneity of variance)

H_1: Residual square sequence autocorrelation (Variance heterogeneity)

Portmanteau Q Test statistic is actually LB statistic of $\{\varepsilon_t^2\}$

$$Q(q) = n(n+2)\sum_{i=1}^{q}\frac{\rho_i^2}{n-i} \tag{10}$$

When the P value of $Q(q)$ test statistic is less than the significance level α The original hypothesis is rejected and the variance of the sequence is considered to be non-homogeneous and autocorrelation.

2). Lagrange Multiplier Test (LM Test)

In 1982, Engle proposed an important ARCH test method: Lagrange multiplier test, abbreviated as LM Test.

The construction idea of Lagrange multiplier test is: if the variance of residual sequence is non-homogeneous and has cluster effect, then the square residual sequence usually has autocorrelation, then we can try to use autoregressive model (ARCH (q) model) or GARCH model to fit the residual sequence.

The hypothesis of Lagrange multiplier test is:

H_0: Residual square sequence pure random;

H_1: The square sequence of residuals has autocorrelation

LM test statistic is:

$$LM(q) = \frac{(SST - SSE)/q}{SSE/(T-2q-1)} \tag{11}$$

Among, $SST = \sum_{t=q+1}^{T} \varepsilon_t^2$, the degree of freedom is T-q-1; $SSE = \sum_{t=q+1}^{T} e_t^2$, the degree of freedom is T-2q-1.

When the P value of LM (q) test statistic is less than the significance level α The original hypothesis is rejected, the variance of the sequence is considered to be non-homogeneous, and the autocorrelation in the square sequence of residuals can be fitted by q-order autoregressive model.

3.3 GARCH Model

The essence of ARCH model is to use the q-order moving average of residual square sequence to fit the current heteroscedasticity function value. Because the moving average model has the q-order truncation of autocorrelation coefficient, ARCH model is actually only applicable to the short-term autocorrelation process of heteroscedasticity function (Wang, 2015).

However, in practice, the heteroscedasticity function of some residual series has long-term autocorrelation. At this time, if the ARCH model is used to fit the heteroscedasticity function, it will produce a high moving average order, increase the difficulty of parameter estimation and finally affect the fitting accuracy of the ARCH model. In order to correct this problem, bollerslov proposed the generalized autoregressive conditional heteroscedasticity in 1985 (generalized autoregressive conditional heteroskedastic) model, its structure is as follows:

$$\begin{cases} x_t = f\left(t, x_{t-1}, x_{t-2}, \ldots\right) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t}\, v_t \\ h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 \end{cases} \tag{12}$$

Where $x_t = f\left(t, x_{t-1}, x_{t-2}, \ldots\right) + \varepsilon_t$ is the deterministic information fitting model of $\{x_t\}$, $e_t \overset{\text{i.i.d}}{\sim} N\left(0, \sigma^2\right)$.

This model is abbreviated as GARCH(p,q).

ARCH model is actually formed by considering the p-order autocorrelation of heteroscedasticity function on the basis of ARCH model; It can effectively fit the heteroscedasticity function with long-term memory. Obviously, ARCH model is a special case of GARCH model. ARCH (q) model is actually GARCH (p, q) model with p=0.

4. Empirical Analysis

4.1 Data Selection and Source

The empirical analysis part selects the daily closing price data of CSI 300 index, and the sample range is from April 12, 2016 to September 30, 2021. Excluding the influence of asynchronous transactions, holidays and other factors, 1335 trading day data are obtained. The data is from the official website of NetEase Finance and Economics. The analysis in this paper is carried out in Rstudio software.

Because this paper mainly studies the yield fluctuation of CSI 300 index, before starting the analysis, the data needs to be transformed into logarithmic yield series, and the transformation formula is as follows: $r = \ln P_t - \ln P_{t-1}$.

4.2 Descriptive Analysis

In order to understand the fluctuation characteristics of the yield series of CSI 300 index, descriptive statistics are made on the yield series and the sequence distribution diagram is drawn as follows.

Table 2 shows the descriptive statistical analysis results of daily logarithmic return of CSI 300 index. It can be seen that the average value of this group of data is very small, indicating that the average return of CSI 300 index is close to 0. Skewness=-1.0680810<0, kurtosis=6.454693>3, indicating that the yield of CSI 300 index is not a standard normal distribution, and this group of data has the distribution characteristics of left deviation and peak. This feature can also be seen in Figure 1. Figure 1 shows the characteristics of peak and thick tail, which shows that the stock price of Shanghai and Shenzhen 300 is easy to fluctuate.

Table 2. Results of Descriptive Statistical Analysis of Rate of Return

Mean	Minimum	Maximum	Median	Variance	Kurtosis	Skewness
0.00009	-0.09154	0.06499	0.00066	0.00024	6.454693	-1.06808

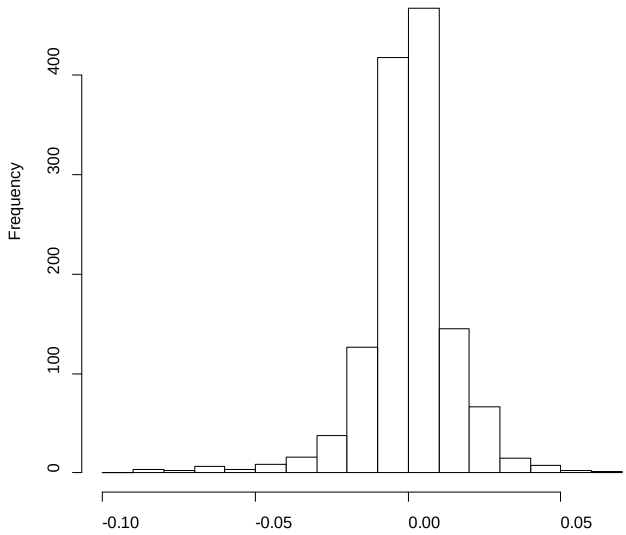

Figure 1. Histogram of Yield Series

Table 2 shows the descriptive statistical analysis results of daily logarithmic return of CSI 300 index. It can be seen that the average value of this group of data is very small, indicating that the average return of CSI 300 index is close to 0. Skewness=-1.0680810<0, kurtosis=6.454693>3, indicating that the yield of CSI 300 index is not a standard normal distribution, and this group of data has the distribution characteristics of left deviation and peak. This feature can also be seen in Figure 1. Figure 1 shows the characteristics of peak and thick tail, which shows that the stock price of Shanghai and Shenzhen 300 is easy to fluctuate.

4.3 Stationary Test

The time series diagram of the yield of CSI 300 index made is as follows:

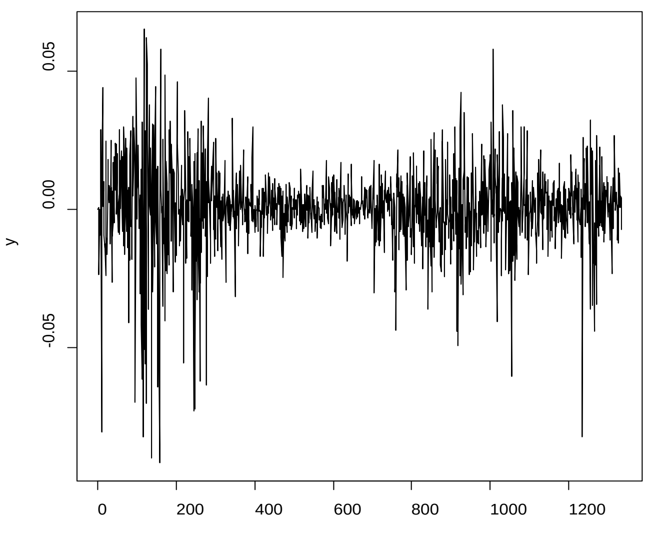

Figure 2. Time Series Diagram of Yield of CSI 300 Index

It can be seen from Figure 2 that the sequence mainly fluctuates around a certain value, without obvious trend or cycle, and can be basically regarded as a stationary sequence. In order to further verify the stability of the sequence, ADF test, PP test and kpss test are also carried out to verify whether the sequence is stable. The results are as follows:

Table 3. Stability Test Results

	Test tatistics	5% critical value / P value	conclusion（α=0.05）
ADF test	-27.193	-1.95/—	stable
PP test	-31.244	-2.864/—	stable
KPSS test	—	—/0.1	stable

The results of the three tests in Table 3 show that the sequence is stable. Combined with the time series diagram, it can be considered that the logarithmic return series is a stationary series. (Note: the original hypothesis of ADF and PP test is that the sequence is unstable, and the original hypothesis of kpss test is that the sequence is stable.)

4.4 Pure Randomness Test

In order to determine whether the data still has extractable information, a pure randomness test is carried out below, and the results are shown in the table below:

Table 4. Pure Randomness Test Results

Delay order	LB Test Statistics	P value
6	20.842	0.002
12	41.162	0.000

It can be seen from the Table 4 that under the condition of significance level of 0.05, the original hypothesis is rejected and it is considered that the yield series of CSI 300 index still has relevant information that can be extracted and cannot be regarded as white noise series.

4.5 Model Identification

After a stationary non white noise sequence is obtained, we start to model the sequence. Next, select the appropriate model by observing the autocorrelation diagram and partial autocorrelation diagram of the sequence.

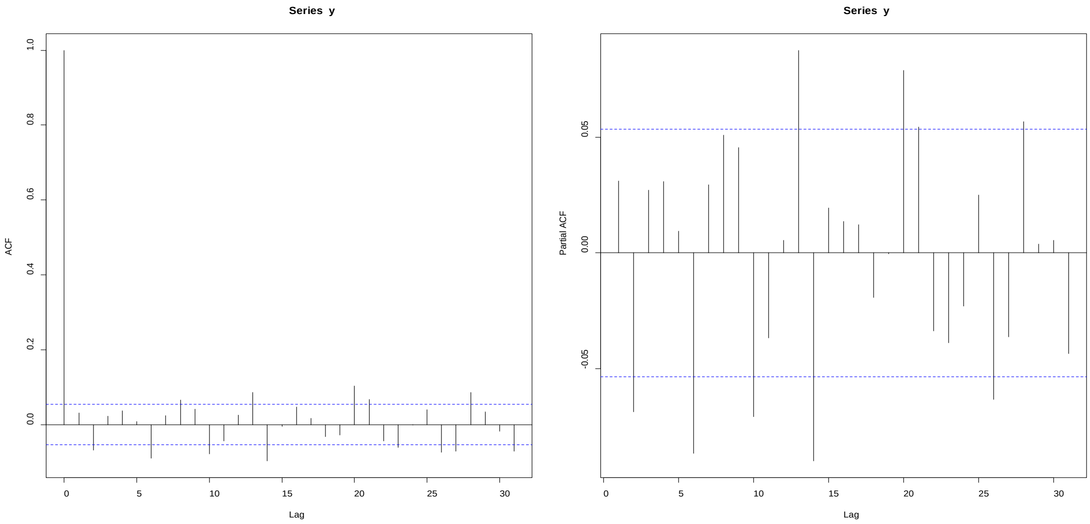

Figure 3. Autocorrelation Coefficient Diagram (Left) and Partial Autocorrelation Coefficient Diagram (Right)

As can be seen from Figure 2, there is no obvious truncation trend in autocorrelation and partial autocorrelation, so we can try to fit the sequence with ARMA (*p*，*q*) model. Because the trend is not obvious, auto. ARIMA () is used to automatically identify the model order and obtain the fitted ARIMA (2,0,3), i.e. ARMA (2,3) model. In order to ensure that the selected model is optimal, several models are compared below.

Table 5. Optimal Selection of Model

Model	Parameter	Estimate	P value	Significance	AIC
ARMA(1,1)	AR(1)	-0.8738	0.0000	Significant	-7334.26
	MA(1)	0.9223	0.0000	Significant	
ARMA(1,2)	AR(1)	-0.314	0.3899	Not significant	-77327.19
	MA(1)	0.3492	0.3775	Not significant	
	MA(2)	-0.056	0.1778	Not significant	
ARMA(1,3)	AR(1)	0.2752	0.280	Not significant	-7325.84
	MA(1)	-0.2413	0.3045	Not significant	
	MA(2)	-0.0749	0.0097	Significant	
	MA(3)	0.0487	0.1054	Not significant	
ARMA(2,1)	AR(1)	-0.1409	0.3287	Not significant	-7327.19
	AR(2)	-0.0649	0.0154	Significant	

	MA(1)	0.176	0.2893	Not significant	
ARMA(2,2)	AR(1)	0.1075	0.0005	Significant	-7359.98
	AR(2)	-0.9595	0	Significant	
	MA(1	-0.1244	0.0058	Significant	
	MA(1)	0.9123	0.0000	Significant	
ARMA(2,3)	AR(1)	0.1414	0.0000	Significant	-7363.22
	AR(2)	-0.9592	0	Significant	
	MA(1)	-0.1102	0.0015	Significant	
	MA(2)	0.9147	0.0000	Significant	
	MA(3)	0.0694	0.0080	Significant	

It can be seen from Table 5 that only all parameters of ARMA (1,1), ARMA (2,2) and ARMA (2,3) models have passed the significance test. Then, according to the AIC criterion, their AIC value and BIC value are compared, and it is found that the AIC value of ARMA (2,3) model is the smallest. To sum up, ARMA (2,3) model is selected as the relatively optimal model. The model is as follows:

$$x_t = 0.1414x_{t-1} - 0.9592x_{t-2} + \varepsilon_t - 0.1102\varepsilon_{t-1} + +0.9147\varepsilon_{t-2} + 0.0694\varepsilon_{t-3} \quad (13)$$

4.6 Model Significance Test

It can be seen from Table 5 that the parameters of the model have passed the t-test, so the parameters of the model are significant. The white noise test (pure randomness test) of residual sequence is carried out below, and the test results are as follows:

Table 6. White Noise Test of Residual Sequence

Delay order	LB Test Statistics	P value
6	3.6387	0.7254
12	10.494	0.5727
18	21.887	0.237

It can be seen from Table 6 that under the significance level of 0.05, the P values of LB statistics of delay order 6, 12 and 18 are significantly greater than 0.05. It can be considered that the residual sequence of the fitting model belongs to white noise sequence, that is, the model is significantly effective.

4.7 Cluster Effect Test

In order to determine whether there is cluster effect in the model residuals, the time series diagram and distribution diagram of the model residuals in equation (13) are given below.

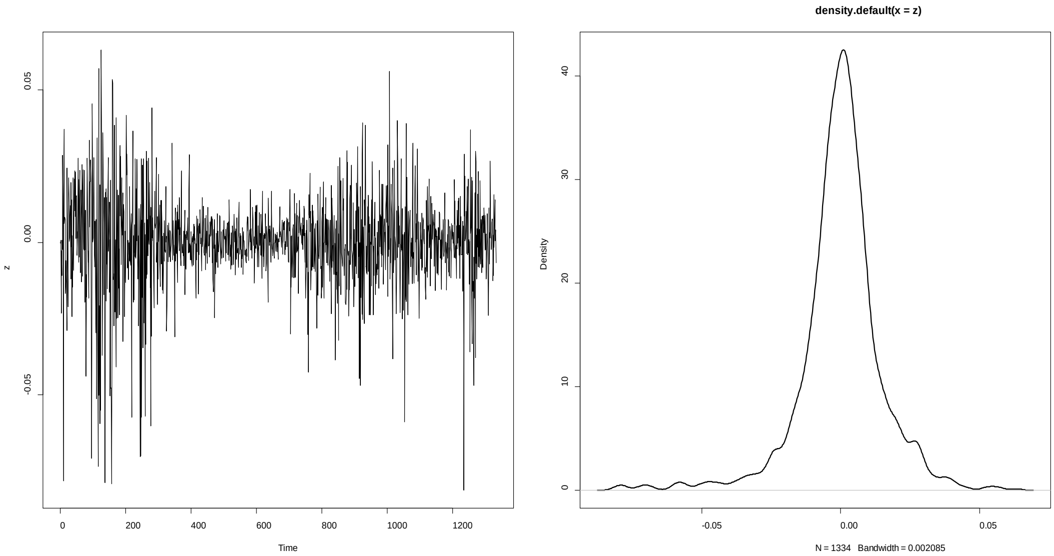

Figure 4. Residual Sequence Diagram (Left) and Distribution Diagram (Right)

It can be seen from the residual sequence diagram in Figure 4 that the variance of the sequence is basically homogeneous in the whole sequence observation period, but there are some periods with large fluctuations, showing a cluster effect. In the distribution diagram, it can be seen that the residual sequence presents the characteristic of "peak and thick tail", which, like the cluster effect, can be used as the indication principle of GARCH model.

4.8 ARCH Effect Test

To fit ARCH or GARCH model, ARCH effect test must be carried out first. The LM Test (Lagrange multiplier test) is used to verify whether there is ARCH effect in the residual sequence. The results are as follows:

Table 7. ARCH **Effect Test Results**

Lag order	LM statistics	P value	Portmanteau *Q* statistics	P value
1	63.736	< 0.0000	63.916	< 0.0000
2	114.3	< 0.0000	139.74	< 0.0000
3	158.53	< 0.0000	229.84	< 0.0000
4	168.18	< 0.0000	281.99	< 0.0000

5	177.93	< 0.0000	338.77	< 0.0000
6	177.86	< 0.0000	359.47	< 0.0000

It can be seen from Table 7 that portmanteau Q test and LM test show that the significant variance of the sequence is non-homogeneous, and there is ARCH effect. The square sequence of residuals has significant autocorrelation. Let's start fitting the residual sequence.

4.9 Fitting GARCH Model

The previous paper verified that there is ARCH effect in the residual sequence of the model, and because the heteroscedasticity function of the residual sequence of the return rate often has long-term autocorrelation, if the ARCH model is used to fit the heteroscedasticity function, it will produce a high moving average order, increase the difficulty of parameter estimation and finally affect the fitting accuracy of the ARCH model, so we choose to fit the GARCH model here.

GARCH model generally does not need to be too high. Most people regard GARCH (1,1) model as a "standard" model, so GARCH (1,1) model is also selected here to fit the residual sequence. The fitting parameter results are as follows:

Table 8. GARCH (1,1) Model Fitting Results

Parameter	Estimate	T statistics	P value
α_0	0.000	3.061	0.0022
α_1	0.0806	14.481	0.0000
β_1	0.9192	187.372	0.0000

It can be seen from Table 8 that at the significance level of 0.05, the results are significant, and the model parameters have passed the t-test. Next, ARCH effect test is performed on the residual residual of the model, and the results are as follows:

Table 9. Arch Effect Test Results

Lag order	LM statistics	P value	Portmanteau Q statistics	P value
1	4.6977	0.0302	4.711	0.02997
2	4.8656	0.08779	4.8056	0.09046
3	8.7141	0.03334	8.5044	0.03666
4	9.0553	0.05973	8.6502	0.07046
5	9.2796	0.09842	8.8101	0.1169
6	10.087	0.121	9.3741	0.1536

It can be seen from Table 9 that at the significance level of 0.01, portmanteau Q test and LM test show that there is no ARCH effect in this sequence.

In conclusion, it can be considered that the fitting effect of GARCH (1,1) model for residual sequence is relatively good. The expression of some GARCH (1,1) models is as follows:

$$\begin{cases} \varepsilon_t = \sqrt{h_t} v_t \\ h_t = \omega + 0.0806\varepsilon^2{}_{t-1} + 0.9345h_{t-1} \end{cases} \quad (14)$$

The 95% confidence interval of fluctuation is given below, as shown in the figure:

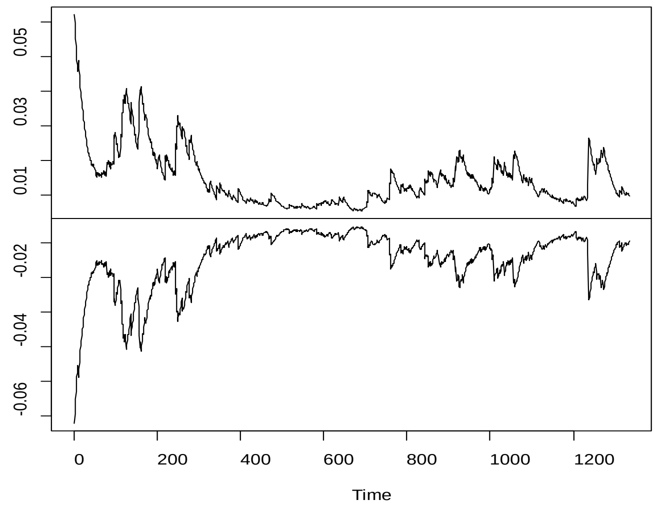

Figure 5. Confidence Interval of Residual Sequence

The fluctuation of the two solid lines in the figure is the 95% confidence interval obtained from the conditional variance fitted by GARCH (1,1) model.

Combining the horizontal model and the fluctuation model, the complete fitting model is as follows:

$$\begin{cases} x_t = 0.1414x_{t-1} - 0.9592x_{t-2} + \varepsilon_t - 0.1102\varepsilon_{t-1} + +0.9147\varepsilon_{t-2} + 0.0694\varepsilon_{t-3} \\ \varepsilon_t = \sqrt{h_t} v_t \\ h_t = \omega + 0.08059\varepsilon^2{}_{t-1} + 0.9192h_{t-1} \end{cases} \quad (15)$$

From the GARCH partial model in equation (15), $\alpha_1+\beta_1=0.9979<1$, very close to 1, indicating that GARCH process is wide and stable, the impact on conditional variance is long-lasting, and the market risk is great, that is, the impact plays an important role in all future forecasts.

5. Conclusion

Based on the daily closing price trading data of CSI 300 index, this paper constructs the logarithmic return series and carries out modeling analysis. The conclusions are as follows:

(1) The yield series of CSI 300 index shows the statistical characteristics of peak, thick tail and bias, as well as asymmetry and persistence. By constructing ARMA (2,3) model, the yield series can be fitted effectively and accurately, and the future trend can be predicted to a certain extent.

(2) The residuals of ARMA model show obvious cluster effect and ARCH effect (i.e. conditional heteroscedasticity). GARCH (1,1) model can better fit the conditional heteroscedasticity, so as to eliminate the ARCH effect.

(3) By constructing GARCH (1,1) model, it is found that the sum of ARCH term coefficient and GARCH term coefficient is less than 1 but very close to 1, indicating that GARCH process is wide and stable, the impact on conditional variance is lasting, and the market risk is large, that is, the impact plays an important role in all future forecasts.

Reference

Campbell, R., Huisman, R., & Koedijk, K. (2001). Optimal portfolio selection in a Value-at-Risk framework. *Journal of Banking & Finance*, *25*(9), 1789-1804. https://doi.org/10.1016/S0378-4266(00)00160-6

Hu, K., & Ge, R. Y. (2021). Price forecast of CSI 300 index based on ARMA mode. *Economy and Management Digest*, *18*, 169-171.

Zhang, F., & Zeng, Q. Z. (2021). Relationship between Investor Sentiment and Stock Returns from the Perspective of Margin Trading—Evidence from Shanghai-Shenzhen 300 Stock Index. *Price: Theory & Practice*, *25*(9), 1789-1804.

Luo, H. J., & Zou, Y. M. (2020). Volatility clustering of CsI 300 stock index and futures based on Markov regime-switching SC copula model. *Journal of Shandong University of Science and Technology (Natural Science)*, *40*(1), 1-11.

Xu, J. (2020). Research on Stock Volatility Based on GARCH Model—Taking Shanghai and Shenzhen 300 index as an example. *Industrial Innovation*, 2, 21-37.

Li, H. Y., Zhang, X. Q., & Chen, Y. (2018). The impact of Shanghai and Shenzhen 300 stock index futures on the volatility and trading behavior of China's stock market. *Statistics & Decision*, *2*, 154-158. http://doi.org/10.13546/j.cnki.tjyjc.2018.06.037

Cao, S., & Zhang, Y. L. (2012). An empirical study on the impact of Shanghai and Shenzhen 300 stock index futures on the spot market. *Statistics & Decision*, *10*, 155-158. http://doi.org/10.13546/j.cnki.tjyjc.2012.10.041

Zheng, P. (2014). Simulation and estimation of VaR value model of Shanghai and Shenzhen 300 index by GARCH model. *Statistics & Decision*, *14*, 154-158. http://doi.org/10.13546/j.cnki.tjyjc.2014.14.046

Wu, L. H., Zhang, X. D., & Wang, L. L. (2012). An empirical study on the impact of stock index futures on Stock Market Volatility—An Empirical Analysis Based on Shanghai and Shenzhen 300 stock index futures. *Statistics & Decision*, *3*, 98-100.

Ma, G. T., & Zhao, Y. (2010). An empirical analysis of the return volatility of Shanghai and Shenzhen 300 stock index—Based on tarch model. *Statistics & Decision*, *3*, 169-171.

Wang, Y. (2015). *Applied time series analysis* (4th Edition). Renmin University of China Press.

Wu, X. Z., & Liu, M. (2014). *Application time series analysis* (accompanied by R software). Machinery Industry Press.

Wang, Y. (2015). *Time series analysis—Based on* R. Renmin University of China Press.