Analysis and Forecast of CSI All Share Health Care Index Based on the ARMA-GARCH Model

Jialin Xu[1,a]

[1]*School of Economics, Shanghai University, Chengzhong Road, Jiading Town, China*
a. xjl0209@126.com

Keywords: ARMA-GARCH, forecasting performance, CSI All Share Health Care Index.

Abstract: In this paper, by using closing prices of CSI All Share Health Care Index on a daily basis from January 4, 2005 to November 29, 2019, constituting 3624 daily in-sample data points, the ARMA-GARCH model under different error distributions are employed to analyze and predict the index returns and prices. The empirical results demonstrate that the return rate of CSI All Share Health Care Index is volatility clustering, with non-normal characteristics of high kurtosis and fat tail, while no strong evidence of the existence of leverage effect has been found. Contrary to the expectation, although the ARMA(1,1)-GARCH(1,1) model under student's t distribution shows the best fitting effect, it presents a poor short-term forecasting performance for the index. Based on the result, a recommendation can be made that it is necessary to further consider relevant factors and explore the combination of nonlinear models so as to improve the prediction accuracy.

1. Introduction

With the rapid development of China's capital market, how to extract useful information from highly dynamic and complex financial time series for prediction has become important for market participants to make decisions on financial investment and risk management, which has also raised scholars' interest in studying the characteristics and changing laws of returns and risks of Chinese securities markets for the last two decades. As compared to other domestic studies, this paper has selected CSI All Share Health Care Index as the experimental object rather than broad-based market indexes mostly used by Chinese scholars. The health care sector in China has demonstrating a good investment value, while the recent Chinese policy of Purchase with Volume has triggered large fluctuations in the sector, arousing much attention from investors. Thus, it is meaningful to analyze the features of returns and its volatility of the sector index.

To capture features of financial time series, a growing number of empirical studies have attempted to establish more effective statistical models, modifying some unrealistic assumptions of traditional econometric models. In 1982, the ARCH model was first put forward by Engle to estimate the variance of inflation in the UK, which successfully depicted the volatility clustering effect of financial time series[1]. Based on the ARCH model, Bollerslev(1986) proposed the GARCH model, which is essentially a return-based model and has proven to be widely-used for modeling the time-varying conditional volatility[2]. Curto et al.(2009) first used the ARMA-

GARCH model based on different error distributions to conduct an empirical study on the daily return rate of stock market indexes[3].

Inspired by western statistical theories, many domestic scholars have conducted similar experiments on the securities market in China. For instance, Qiming Tang and Jian Chen (2001) and Wanbo Lu (2006) both took SSE(Shanghai Stock Exchange) composite index and SZSE(Shenzhen Stock Exchange) component index as empirical objects, showing that there was an obvious ARCH effect of the stock market in China. Additionally, the former found that there was a weak negative correlation between the returns and volatility of the stock market in China[4], and the latter pointed out that the non-parametric GARCH(1,1) model would significantly improve prediction accuracy of the volatility of index returns[5]. In 2012, Bo Wang showed that ARMA-GARCH model under student's t distribution could make the optimal fitting effect on SSE composite index returns during the period of September 30, 2004 to September 30, 2011[6]. Qi Yang and Xianbing Cao(2016) used the ARMA-GARCH model to predict stock prices and supported the model effectiveness on short-term forecasting[7].

From the perspective of numerous empirical findings, the most representative and classic model applied to analyze the conditional expectation and heteroscedasticity of univariate returns series is ARMA-GARCH. Thus, this paper aims to employ the ARMA-GARCH model to analyze characteristics of the index returns and its volatility of health care sector in Chinese stock market as well as test short-term forecasting performance of the model. The daily data comes from closing prices of CSI All Share Health Care Index from January 4, 2005 to November 29, 2019.

The remainder of the paper is organized as follows. Section 2 provides the econometric models and methodology for this research. Section 3 describes the data collection and the statistical characteristics of sample. Section 4 shows the empirical analysis and results, which includes the stationarity test, autocorrelation test, model estimation and forecasting. Section 5 concludes.

2. Econometric Methodology

2.1. ARMA Model

In the 1970s, Box and Jenkins, two statisticians, first proposed the ARMA(Autoregressive and Moving Average) model, which has been commonly used in modeling stationary time series with complex autocorrelation behavior. The process is jointly characterized by the p-order autoregressive model and the q-order moving average model, reflecting the dependence of the present observed value on its own lag value (the historical observed value) and the lag value of white noise (the present and past impacts of various random factors). The general expression is as follows.

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \qquad (1)$$

Where, Y_t is the observed value, c is the mean of $\{Y_t\}$, p and q are the lag order, φ_i (i=1,2,…p) and θ_j (j=1,2,…q) are the parameter of AR process and MA process, respectively, ε_t obeys white noise process with zero-mean.

2.2. GARCH Model

In the field of macroeconomics and finance, it is common to discover that many time sequences have the characteristic of volatility clustering, which will cause the forecast to vary over time. Therefore, depicting volatility plays an important role. The ARCH model, a basic model to describe the volatility proposed by Engle(1982), however, is not applicable to heteroscedasticity function with long-term autocorrelation in practice. In order to better deal with high order autoregression, Bollerslev(1982) brought forward the GARCH(Generalized ARCH) model, setting the conditional variance as a linear function of historical market impact and past conditional variance. The general expression is as follows.

$$\begin{cases} \sigma_t^2 = \omega + \sum_{j=1}^{r} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \\ \varepsilon_t = z_t \sigma_t \end{cases} \tag{2}$$

Where, r and s are the lag order, $\omega, \alpha_j, \beta_i$ are respectively the coefficients of intercept term, ARCH term and GARCH term of variance equation. To guarantee positive conditional variance of return rate and stationary process, this model has several essential conditions: $\omega>0$, $\alpha_j \geq 0$, $\beta_i \geq 0$, and the sum value of α_j and β_i should be less than one. The closer the sum value approaches to one, the higher volatility persistence is reflected by the returns series.

Here, the error term z_t in a standard GARCH model is often assumed to obey a normal distribution. However, due to non-normal characteristics showed by many financial time series, other non-normal distributions should be given priority to consider. Moreover, when the conditional mean equation is ARMA, the expression of ARMA(p,q)-GARCH(r,s) model can be a combination of Equation (1) and (2), used to describe the condition mean and the condition variance of the returns series.

2.3. EGARCH Model

Since the market impact in the GARCH model is devised in the form of squares, it is impossible to judge the impact differences between negative and positive shocks on volatility. Thus, Nelson(1991) proposed the EGARCH(Exponential GARCH) model, which can explain the asymmetric impact of market shocks on volatility.The expression can be written as:

$$In(\sigma_t^2) = \omega + \sum_{j=1}^{q} (\gamma_j \varepsilon_{t-j} + \alpha_j (|\varepsilon_{t-j}| + E|\varepsilon_{t-j}|)) + \sum_{i=1}^{p} \beta_i In(\sigma_{t-i}^2) \tag{3}$$

Where, the parameter γ_j measures the asymmetry effect. If the value of γ_j is significantly negative, it suggests bad news have a more significant influence on conditional variance. In other words, returns on assets are more volatile when prices fall, which is usually called leverage effect.

3. Data

The experimental object adopted in this article, CSI All Share Health Care Index, consists of stocks with high liquidity and good market representation in the health care sector of Shanghai and Shenzhen stock markets, which is a better reflection of overall performance of the listed companies in this particular sector. To ensure a sufficient sample size, daily closing prices of CSI All Share Health Care Index are collected, regarding the period ranging from January 4, 2005 to December 31, 2019, making up a total of 3646 observations. Then, the first 3624 data from January 4, 2005 to

November 29, 2019 have been adopted as the estimation sample, while the last 22 data have been used as out-of-sample period for forecasting. The total daily data set used in this article is supplied by the WIND database, and the calculation results are realized by R software.

Compared with index prices, index returns better reflect investment opportunities of assets as well as have statistical characteristics easy to deal with. Therefore, the sample data is preprocessed with logarithm and first-order difference to obtain daily index returns which are defined by $R_t=\ln(P_t)-\ln(P_{t-1})$. Where, P_t and P_{t-1} represents the index closing price at time t and time t-1, respectively. R_t is the daily index returns in a logarithmic form, and R_t series will be used as the modeling object in this paper.

Figure 1, below, illustrates that the daily returns of CSI All Share Health Care Index are moving around the approximately zero with time-varying clustering fluctuations. This suggests the trend variation has been basically eliminated, and the small(or large) fluctuations of returns tend to be followed by small(or large) fluctuations, leading to a preliminary confirmation on stationarity of the R_t series and existence of heteroscedasticity.

Table 1, below, reports the descriptive statistics of sample. The daily return rate of has a mean of about 0.0006 and a standard deviation of about 1.86%. The skewness value is negative. The kurtosis value is greater than 3, showing a fat tail. In addition, the test value of Jarque-Bera statistic is 1558.2, which is large enough to reject the null hypothesis of normal distribution. Hence, non-normal error distributions are preferred to consider in the process of subsequent modeling.

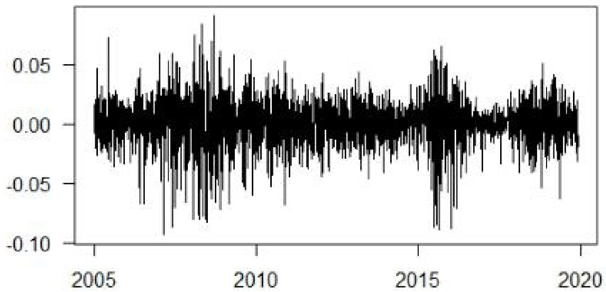

Figure 1: Daily returns series of CSI All Share Health Care Index from January 4, 2005 to November 29, 2019.

Table 1: Descriptive Statistics.

Observations	Mean	Std. Dev.	Skewness	Kurtosis	JB statistic	Probability
3623	0.00063	0.01855	-0.58932	5.98703	1558.2	0.00000

4. Empirical Analysis and Results

4.1. Stationarity Test and Autocorrelation Test

Before establishing ARMA models, it requires to take a stationarity test and an autocorrelation test so as to determine whether the relevant model is suitable to analyze the series.

For the stationarity test, a widely-used method is ADF (Augmented Dickey-Fuller) Test. Table 2 has listed the testing result. It shows that in the process of successively removing the trend term and drift term, the t statistic of ADF test has always been obviously lower than the test critical value at 1% significant level. Thus, the null hypothesis of unit root should be rejected, indicating the R_t series meets the stationarity requirement of modeling. Besides, it is found that in the cases involving trend term or drift term, the coefficient of drift term is marked at the significance level of 10%, which suggests that the R_t series might be a stationary process with intercept term.

For the autocorrelation test, two methods, Ljung-Box test (LB statistic) and Box-Pierce test (Q statistic), are adopted in this paper. Table 3 has given the results of 8 lags as well as the values of AC and PAC (autocorrelation and partial autocorrelation function). Under each order of lag, the associated probability values corresponding to all of the LB statistics and Q statistics are approaching to zero, rejecting the null hypothesis of pure random. Hence, the R_t series have the autocorrelation and can be extracted more useful information by establishing the mean equation like the ARMA model.

Table 2: Stationarity test results.

Coefficient	Trend type	Drift type	None type
ADF test term	-0.9675^{***} (-42.827)	-0.9665^{***} (-42.798)	-0.9644^{***} (-42.736)
Drift term	0.001358^{*} (2.206)	0.000607^{*} (1.974)	
Trend term	-4.140e-07 (-1.408)		

Notes: t-statistic in parentheses indicate the coefficients' significance at the 10%, 5% and 1% levels, respectively.

Table 3: Autocorrelation Test Results.

Lag Period	AC	PAC	LB statistic	Q statistic	Probability
1	0.076	0.076	20.719	20.702	0.000
2	-0.040	-0.046	26.434	26.410	0.000
3	0.043	0.050	33.287	33.254	0.000
4	0.018	0.009	34.462	34.427	0.000
5	-0.023	-0.021	36.350	36.312	0.000
6	-0.031	-0.028	39.785	39.739	0.000
7	0.042	0.044	46.064	46.002	0.000
8	0.016	0.009	47.035	46.970	0.000

4.2. Model Estimation

First, to better describe the mean of the R_t series, AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(2,1) and ARMA(1,2) models have been adopted, respectively. Table 4 has given the results of parameter estimation, where, c is the intercept parameter of the mean equation, ar1 and ar2 represent the coefficients of AR process with the lag order of 1 and 2, and ma1 and ma2 represent the coefficients of MA process with the lag order of 1 and 2.

As can be seen in Table 4, the intercept parameters of each model prove to be marked at the significance level of 5%. Among the estimations of AR(1), AR(2), MA(1), MA(2) and ARMA(1,1), all of which show the significantly non-zero coefficients of the lag terms with associated probability values of t statistics smaller than the significance level of 1%, the three information criteria (AIC, BIC, LL) are best fit to be selected via the ARMA(1,1) process according to the maximum value of

LL as well as the minimum value of AIC and BIC. Thus, the ARMA(1,1) model is selected as the optimal fitting model to describe the mean of the R_t series. The mean equation is as follows:

$$R_t = 0.0006 - 0.5751R_{t-1} + 0.6549\varepsilon_{t-1} + \varepsilon_t \quad (4)$$

Moreover, the Portmanteau Q method and Lagrange Multiplier(LM) method are both employed to test ARCH effect of the residuals sequence based on the ARMA(1,1) model with lag orders from one to ten. Table 4 mainly exhibits the testing results of ARCH(10). And it is found that, under each test, associated probability values of the higher-order statistics are approaching to zero with the rejection of null hypothesis at the 1% significance level. Therefore, the residuals sequence turns out to have an high-order ARCH effect, and a conditional heteroscedasticity model need to be further introduced to fit the volatility of the R_t series.

Table 4: Estimation results of the ARMA model.

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(2,1)	ARMA(1,2)
Parameter estimation							
c	0.0006** (2.00)	0.0006** (2.00)	0.0006** (2.00)	0.0006** (2.00)	0.0006** (2.00)	0.0006** (2.00)	0.0006** (2.00)
ar1	0.0756*** (4.55)	0.0791*** (4.77)			-0.5751*** (-3.33)	-0.4706 (-1.13)	-0.4595 (-1.07)
ar2		-0.0457*** (-2.75)				-0.0087 (-0.18)	
ma1			0.0836*** (4.78)	0.0834*** (5.02)	0.6549*** (4.09)	0.5521* (1.32)	0.5410 (1.26)
ma2				-0.0452*** (-2.77)			-0.0097 (-0.20)
σ_t^2	0.0003423	0.0003417	0.0003421	0.0003415	0.0003411	0.0003412	0.0003412
Model optimization							
LL	9315.69	9319.48	9316.76	9320.53	9322.39	9322.28	9322.28
AIC	-18625.38	-18630.95	-18627.51	-18633.17	-18636.77	-18634.56	-18634.57
BIC	-18606.80	-18606.17	-18608.93	-18608.39	-18611.99	-18603.59	-18603.59
ARCH(10)							
LM	149.59***	149.69***	149.04***	149.44***	148.63***	148.81***	148.82***
Portmanteau Q	1047.8***	1040.1***	1045.7***	1044.9***	1074.3***	1066.0***	1066.4***

Notes: t-statistic in parentheses. *,**,*** indicate the coefficients' significance at the 10%, 5% and 1% levels, respectively. ARCH(10) is the test for ARCH effects with 10 lags by using squared standardized residuals.

Second, since the simplest form of the GARCH(r,s) model can well describe the volatility of most series, this article selects the GARCH(1,1) model to help further capture conditional heteroscedasticity of the R_t series based on the ARMA(1,1) model. Besides, due to the non-normal characteristics of the R_t series, ARMA(1,1)-GARCH(1,1) models under various error distribution conditions, such as student's t(std), skew-student's t(sstd), generalized error(ged) , skew-generalized error(sged) and normal(norm) distributions, will be respectively constructed and estimated in the following. The estimation results are shown in Table 5, where, omega, alpha1 and beta1 are successively the coefficients of intercept term, ARCH term and GARCH term.

In Table 5, it is found that, under each distribution condition, the parameter omega doesn't show significance, while the parameters ar1 and ma1 of the mean equation and the parameters alpha1, beta1, skew and shape of the variance equation are marked at the confidence level of 99%, indicating that the selected model is correct. Through plenty of regression testing and comparisons, the models with the intercept term of the mean equation prove to have better fitting effect under three information criteria (AIC, BIC, LL). Then, the cases within normal(norm), skew-student's t(sstd) and skew-generalized error(sged) distributions are excluded from consideration due to their non-significant or less significant estimates of intercept parameter c. By comparison, values of AIC and BIC in the case of student's t distribution are smaller than the values in the case of generalized error distribution. Thus, the ARMA(1,1)-GARCH(1,1) model based on student's t distribution has been determined to be the optimal model. The equation is as follows.

$$\begin{cases} R_t = 0.000933 - 0.6599 R_{t-1} + 0.7266 \varepsilon_{t-1} + \varepsilon_t \\ \sigma_t^2 = 0.0645 \varepsilon_{t-1}^2 + 0.9345 \sigma_{t-1}^2 \end{cases} \tag{5}$$

In Eq. (5), both of alpha1 and beta1 in the variance equation are larger than zero, and the sum value of these coefficients equals about 0.999(close to one but less than one), which meets the constraint conditions of stationary process and reflects higher volatility persistence. According to the results of ARCH(10) using LM test statistics, as indicated in Table 5, associated probability values of the higher-order statistics are much higher than the significance level of 10%, accepting the null hypothesis of no remaining ARCH effects. Thus, it can conclude that the ARMA(1,1)-GARCH (1,1) model under student's t distribution does a better job to capture the volatility clustering of the R_t series.

Table 5: Estimation results of the ARMA-GARCH model.

	ARMA(1,1) -GARCH(1,1) -norm	ARMA(1,1) -GARCH(1,1) -std	ARMA(1,1) -GARCH(1,1) -sstd	ARMA(1,1) -GARCH(1,1) -ged	ARMA(1,1) -GARCH(1,1) -sged
Parameter estimation					
c	0.000485* (2.003)	0.000933*** (4.109)	0.000485* (2.037)	0.000884*** (3.867)	0.000395 (1.620)
ar1	-0.6480*** (-6.002)	-0.6599*** (-6.037)	-0.6952*** (-8.710)	-0.6519*** (-4.574)	-0.7076*** (-16.967)
ma1	0.7186*** (7.283)	0.7266*** (7.258)	0.7598*** (10.537)	0.7154*** (5.433)	0.7701*** (20.395)
omega	0.000001 (0.585)	0.000001 (1.035)	0.000001 (1.026)	0.000001 (0.932)	0.000001 (0.961)
alpha1	0.0524*** (3.163)	0.0645*** (6.373)	0.0654*** (6.147)	0.0589*** (5.326)	0.0609*** (5.230)
beta1	0.9445*** (55.692)	0.9345*** (91.880)	0.9336*** (87.306)	0.9389*** (81.961)	0.9366*** (77.353)
shape		6.656*** (10.874)	7.0164*** (10.344)	1.380*** (36.618)	1.4136*** (33.905)
skew			0.8660*** (42.952)		0.8758*** (44.568)
Model optimization					
LL	9792.423	9863.032	9882.468	9860.169	9880.054
AIC	-5.402	-5.441	-5.451	-5.439	-5.449
BIC	-5.392	-5.429	-5.437	-5.427	-5.436
ARCH(10)					
LM	6.187 (p=0.799)	4.324 (p=0.9316)	4.087 (p=0.9433)	5.247 (p=0.8741)	4.765 (p=0.9063)

Notes: t-statistic in parentheses. *,**,*** indicate the coefficients' significance at the 10%, 5% and 1% levels, respectively. ARCH(10) is the test for ARCH effects with 10 lags by using squared standardized residuals.

Finally, in order to explore whether leverage effect exists or not, the ARMA(1,1)-EGARCH(1,1) model is estimated under various error distributions. Table 6 shows two example cases within student's t(std) and skew-student's t(sstd) distributions, where, the parameter gamma1 measures the asymmetry effect. As can be seen, although the parameter gamma1 is negative, it is not significant at all, demonstrating that there is no obvious asymmetry or leverage effect existing in the volatility of CSI All Share Health Care Index during the sample period. Furthermore, the ARMA(1,1)-EGARCH(1,1) model hasn't eliminated the heteroscedasticity of the original residuals sequence.

Table 6: Estimation results of leverage effect.

	ARMA(1,1)-EGARCH(1,1)-std			ARMA(1,1)-EGARCH(1,1)-sstd		
	Coeffcient	t value	Pr(>|t|)	Coeffcient	t value	Pr(>|t|)
mean	0.00092***	5.078	0.00000	0.00044*	1.855	0.06365
ar1	-0.6696***	-26.526	0.00000	-0.7026***	-15.129	0.00000
ma1	0.7315***	31.269	0.00000	0.7633***	18.152	0.00000
omega	-0.0607**	-10.653	0.00000	-0.0677**	-2.408	0.01603
alpha1	0.1445***	13.010	0.00000	0.1479***	2.588	0.00965
beta1	0.9926***	1271.685	0.00000	0.9917***	269.322	0.00000
gamma1	-0.00016	-0.019	0.98487	-0.0022	-0.339	0.73461
shape	6.6108***	9.812	0.00000	7.0041***	5.637	0.00000
skew				0.8591***	42.453	0.00000
LL	9867.836			9888.994		
AIC	-5.443			-5.454		
ARCH (3)	6.123** （p=0.01334）			5.408** （p=0.02004）		

Notes: *,**,*** indicate the coefficients' significance at the 10%, 5% and 1% levels, respectively. ARCH(3) is the LM test for ARCH effects with 3 lags by using squared standardized residuals.

4.3. Model Forecasting

To test the forecasting performance of the ARMA(1,1)-GARCH(1,1) model on CSI All Share Health Care Index, the static prediction method is used to make short-term predictions for the logarithmic returns of the index in the 22 trading days after November 29, 2019. To be compared with the out-of-sample data, the predicted values of index returns have been manually restored to the predicted values of index prices. As shown in Table 7, two popular measures, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), are calculated to evaluate and compare prediction accuracy of the ARMA(1,1)-GARCH(1,1) model under four error distributions. Among the four cases, the ARMA(1,1)-GARCH(1,1) model under student's t distribution has the minimum value of MAE and RMSE, demonstrating again that the optimal model has the best imitative effect and forecasting power on the R_t series.

Table 7: Forecasting performance under four conditions.

	ARMA-GARCH-std	ARMA-GARCH-ged	ARMA-GARCH-norm	ARMA-GARCH-sstd
RMSE	90.44594	90.50483	90.8428	90.94757
MAE	75.18617	75.20373	75.5488	75.65510

Table 8, below, has specifically listed the forecasting errors by using the optimal model. Unexpectedly, there is a certain degree of deviations existing between the predicted values and the actual values with the maximum absolute value of 2.14%, the minimum absolute value of 0.08%, and the average deviation value of 22 days reaching 0.77%. As a matter of fact, the fluctuation of index returns is usually ranging between -2% and 2% over a single trading day. Thus, in a strict sense, the forecasting errors are relatively large, which illustrates that although the ARMA(1,1)-GARCH(1,1) model with student's t distribution can better describe the features of the index returns, it is still lacking in prediction accuracy to some extent.

Table 8: Forecasting errors under the optimal model.

Date	Error(%)	Date	Error(%)	Date	Error(%)	Date	Error(%)	Date	Error(%)
12-02	1.16	12-09	1.66	12-16	-0.69	12-23	1.16	12-30	-0.26
12-03	0.18	12-10	-0.75	12-17	-0.83	12-24	-0.34	12-31	-2.14
12-04	-0.43	12-11	0.35	12-18	1.13	12-25	0.08		
12-05	-0.98	12-12	0.24	12-19	-0.35	12-26	-0.57		
12-06	-0.49	12-13	-1.43	12-20	0.97	12-27	0.80		

5. Conclusions

Given the recent considerable focus from investors on the health care sector of Chinese stock market, it is meaningful to analyze the features of returns and its volatility of the sector index. To this end, the ARMA-GARCH model has been applied by using daily data of CSI All Share Health Care Index from January 4, 2005 to November 29, 2019. And the following conclusions can be drawn.

First, the return rate of CSI All Share Health Care Index shows non-normal characteristics of high kurtosis and fat tail with obvious volatility clustering and conditional heteroscedasticity. And this is consistent with the previous evidence on Chinese overall stock market indexes such as SSE composite index and SZSE component index. However, according to the estimation results of the ARMA(1,1)-EGARCH(1,1) model, there is no strong evidence of the existence of leverage effect or asymmetric effect which is often observed in stock markets, in other words, the volatility of the index returns shows a similar degree of response to negative shocks and positive shocks.

Second, according to the empirical results of the ARMA(1,1)-GARCH(1,1) model under five different error distributions, although the ARMA(1,1)-GARCH(1,1) model under student's t distribution proves to be the optimal model fit for effectively characterizing the returns series of CSI All Share Health Care Index, it doesn't perform an acceptable short-term forecasting power for the index.

Here are two possible explanations. On the one hand, the parameters are uncertain in reality, so it is difficult for the ARMA-GARCH model with constant parameters to capture the instantaneous structural changes of the series through collecting data from various time in an equivalent way. On the other hand, for lack of the consideration of other potential factors or exogenous variables, the modeling of univariate time series can only dig out part of the laws, which leads to the lower prediction accuracy. For investors' part, the movement of sector index prices or returns is dynamically influenced by many uncertain factors like industrial policies and market circumstances. Thus, it is necessary to further consider related factors and explore the combination of non-linear models to improve prediction accuracy and achieve a better forecasting performance.

References

[1] Engle, R.F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. inflation. Econometrica, 50, 987-1008.

[2] Bollerslev,T. (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31, 307-327.

[3] Curto, J.D., Pinto J.C., Tavares, G.N. (2009) Modeling stock markets' volatility using GARCH models with Normal, Student' t and stable Paretian distributions. Statistical Papers, 50, 311-321.

[4] Tang Qiming, Chen Jian. (2001) The Analysis of the ARCH's Effects of Stock Markets in China. the Journal of World Economy, 24(3), 29-36.

[5] LU Wanbo. (2006) Forecasting Volatility of Chinese Stock Markets with Nonparametric GARCH Model. Application of Statistics and Management, 25(4), 455-461.

[6] Wang Bo. (2012) Empirical Analysis for the Shanghai Composite Index Based on ARMA-GARCH Model. Science Technology and Engineering, 12(5), 1219-1221,1226.

[7] Yang Qi, Cao Xianbing. (2016) Analysis and Prediction of Stock Price Based on ARMA-GARCH Model. Mathematics in Practice and Theory, 46(6), 80-86.