Analysis and Forecast on the Price Change of Shanghai Stock Index

Qi Qi

***Abstract*—This paper selects the closing price data of Shanghai stock index from January 4, 2000 to June 2020 as the research data. Through ADF test, white noise test and ARCH effect test, it is found that the closing price sequence of Shanghai stock index has the characteristics of non-stationary and autocorrelation. Its first-order difference sequence is a stationary white noise sequence, and has the characteristics of peak thick tail and conditional heteroscedasticity, based on this, ARMA model and GARCH model are selected to model the data, and it is found that the closing price sequence of Shanghai stock index has leverage effect. At the same time, the model is further used to make short-term forecast, including dynamic forecast and static forecast for the time series data, and the conclusion is that GARCH(1,2) is the more favorable model, and the shortcomings of this analysis are pointed out.**

***Index Terms*—Time series, ARMA model, GARCH model, Shanghai stock index, non-white noise series.**

I. Introduction

With the globalization of market opening and financial innovation, the research on the periodicity and law of stock market volatility has become one of the most important issues in China and even the international financial market. [1] For one thing, investors want to find the hidden economic laws in the stock market. For another, people are also exploring more accurate methods and tools to predict the stock market. There is no doubt that financial time series model analysis occupies a place in this field. In this paper, ADF test, white noise test and ARCH effect test are used to carry out the first-order difference, and ARMA model and GARCH model are selected to model and predict the data. Hence, managers understand the dynamics of the stock market, so as to make relevant decisions, and at the same time, investors can get higher returns. [2] However, the stock price series is a very complex nonlinear dynamic system. The traditional time series forecasting methods reflect the characteristics and changes of linear dynamic system through the statistical relationship of time series, so as to reveal the inherent change rules. Therefore, in order to predict stock price completely and accurately for such a nonlinear dynamic system as stock, the time series model plays an important role in the analysis and prediction of Shanghai stock index.

Manuscript received July 7, 2021; revised October 2, 2021.
The author is with Shihezi University, Shihezi, Xinjiang Uygur Autonomous Region, 832000, China (e-mail: q17755517009@163.com).

II. Related Research Methods and Theoretical Basis of Model

A. ADF Test

In order to ensure the accuracy of the model, this paper makes a stationary test on the series. The author choose ADF test, and the number of lag periods is determined by AIC and SC minimum criteria.

B. Processing of Nonstationary Time Series

White noise test is the most important part of the time series. The rationality of the selected model can be judged by the white noise test of differential residuals.

C. Establishing ARMA Model

AR model is to use the previous observation value and the current interference value through a certain linear combination to predict and analyze. MA model is based on the previous interference value and the current interference value through a certain linear combination to predict. ARMA model is composed of AR model and MA model, which is mainly used to describe stationary stochastic process.

D. Establishing GARCH Model

ARCH model is called "autoregressive conditional heteroscedasticity" model, which describes the heteroscedasticity of residual items in financial time series model. The core idea of arch model is that the variance of the residual term at time t depends on the square of the residual term at time t-1. [3]

In order to get a good fitting effect, ARCH model needs to set a large lag coefficient p, which inevitably needs to estimate a lot of parameters, which increases the amount of calculation. GARCH model solves this problem well. GARCH model introduces lag term in conditional variance equation of arch model and obtains GARCH (p, q) [4].

III. Empirical Analysis

A. Sample Data Selection

In order to make sure the selected stock index reflect the situation of China's stock market as much as possible [5], the Shanghai stock index is selected as the representative stock of China, and the closing price data of Shanghai stock index from January 4, 2000 to June, 2020 is selected as the research object to study the periodicity of China's stock market. Remove some problematic data, compare and process the remaining closing price data of Shanghai stock index in Excel on each working day, and then get the final experimental data after deleting the non-overlapping dates.

doi: 10.18178/joebm.2022.10.1.676

At the same time, the time series were further analyzed in Eviews.

B. Tests and Treatment of Time Series

1) Stability test and treatment

Before analyzing time series, in order to eliminate the phenomenon of pseudo regression, it is necessary to carry out unit root test on time series. [6]

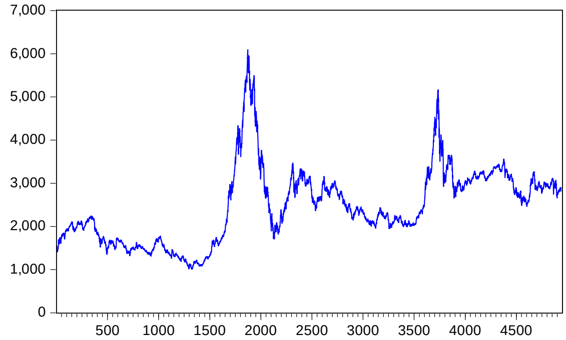

Fig. 1. Time series diagram of original time series.

From the time series, it can be seen from Fig. 1 that the time series has not obvious periodicity, and direct observation cannot directly determine whether it is stable. [7] At this time, we can further judge whether the series is stable with the help of autocorrelation graph.

Date: 06/09/20 Time: 12:52
Sample: 1 4949
Included observations: 4949

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.998	0.998	4937.1	0.000
		2	0.997	-0.030	9859.5	0.000
		3	0.995	0.046	14769.	0.000
		4	0.994	-0.044	19663.	0.000
		5	0.992	-0.061	24541.	0.000
		6	0.990	-0.008	29403.	0.000
		7	0.989	0.048	34249.	0.000
		8	0.987	-0.026	39080.	0.000
		9	0.985	-0.006	43895.	0.000
		10	0.983	0.003	48693.	0.000
		11	0.982	0.009	53477.	0.000
		12	0.980	-0.021	58244.	0.000
		13	0.978	-0.031	62994.	0.000
		14	0.976	-0.056	67725.	0.000
		15	0.974	0.034	72439.	0.000
		16	0.972	-0.048	77134.	0.000
		17	0.970	0.001	81810.	0.000
		18	0.968	0.001	86467.	0.000
		19	0.966	-0.015	91105.	0.000
		20	0.964	0.050	95725.	0.000
		21	0.962	-0.012	100326	0.000
		22	0.960	-0.009	104909	0.000
		23	0.958	-0.010	109474	0.000
		24	0.956	0.025	114021	0.000
		25	0.954	-0.014	118549	0.000
		26	0.952	-0.003	123060	0.000
		27	0.950	0.011	127552	0.000
		28	0.948	-0.000	132027	0.000
		29	0.946	-0.035	136482	0.000
		30	0.944	0.002	140919	0.000
		31	0.942	0.021	145338	0.000
		32	0.940	0.014	149739	0.000
		33	0.938	0.040	154123	0.000
		34	0.936	0.006	158491	0.000
		35	0.934	-0.041	162841	0.000
		36	0.932	-0.018	167175	0.000

Fig. 2. Autocorrelation graph of original time series.

According to Fig. 2, it can be seen that the ACF is always positive above the zero axis as the number of delayed periods increases, and the attenuation rate is very slow. Therefore, it can be judged that the original time series is non-stationary.

Meanwhile, the results show that the P value of LB test statistic is far less than 0.05 under each order delay. Therefore, it can be determined that the original sequence belongs to non-white noise sequence, which has the significance of further research.

2) Model order determination

Combined with Fig. 3, it can be seen from Fig. 4 that the value of ADF statistic is - 60.86428, which is less than the corresponding critical values of 1%, 5% and 10% (respectively - 3.959894, - 3.410714, - 3.127144), and P value is 0, which means that the original hypothesis is rejected, the first-order difference sequence has no unit root and the sequence is stable.

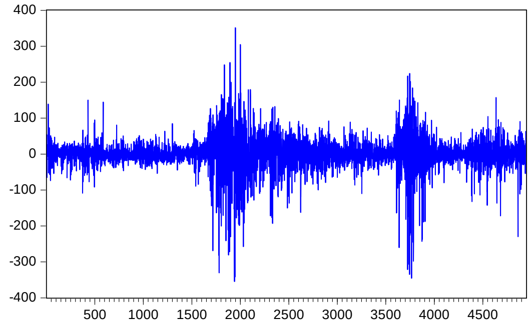

Fig. 3. First order differential time series diagram.

Null Hypothesis: D(DX) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-60.86428	0.0000
Test critical values:	1% level	-3.959894	
	5% level	-3.410714	
	10% level	-3.127144	

*MacKinnon (1996) one-sided p-values.

Fig. 4. Unit root test after first order difference.

Date: 06/09/20 Time: 21:07
Sample: 1/04/2000 6/08/2020
Included observations: 4947

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.495	-0.495	1214.0	0.000
		2	-0.056	-0.399	1229.4	0.000
		3	0.039	-0.311	1236.9	0.000
		4	0.049	-0.187	1249.0	0.000
		5	-0.011	-0.105	1249.6	0.000
		6	-0.066	-0.154	1271.3	0.000
		7	0.041	-0.133	1279.9	0.000
		8	0.013	-0.107	1280.7	0.000
		9	-0.005	-0.076	1280.8	0.000
		10	-0.033	-0.103	1286.3	0.000
		11	0.021	-0.099	1288.5	0.000
		12	-0.002	-0.114	1288.5	0.000
		13	0.030	-0.054	1293.1	0.000
		14	-0.044	-0.076	1302.8	0.000
		15	0.017	-0.068	1304.2	0.000
		16	0.019	-0.039	1306.0	0.000
		17	-0.028	-0.059	1309.9	0.000
		18	0.029	-0.012	1314.1	0.000
		19	-0.045	-0.059	1324.0	0.000
		20	0.040	-0.035	1332.1	0.000
		21	-0.012	-0.033	1332.8	0.000
		22	-0.001	-0.022	1332.8	0.000
		23	-0.014	-0.041	1333.8	0.000
		24	0.025	-0.016	1336.9	0.000
		25	-0.024	-0.047	1339.9	0.000
		26	0.003	-0.059	1339.9	0.000
		27	0.025	-0.025	1343.0	0.000
		28	-0.025	-0.043	1346.2	0.000
		29	0.027	-0.005	1349.9	0.000

Fig. 5. Correlation coefficient graph after first order difference.

As shown in Fig. 5, autocorrelation is truncated and partial autocorrelation is trailing. From the autocorrelation graph, the following analysis can be made:

a) The autocorrelation coefficient decreases rapidly to 0 after k=1, so the MA (1) model can be fitted.

b) The partial autocorrelation coefficient decreases to 0 after k=1 or k=2, so AR (1) model or AR (2) model can be fitted.

c) At the same time, ARMA (1,1) and ARMA (2,1) models can be considered.

IV. Modeling

A. ARMA Model

1) Modeling

- **Establishing MA (1) model**

As can be seen from Fig. 6, the coefficient P value is less than the significance level 0.05, so it is considered that the coefficient passes the test, and the MA (1) model can be established.

Dependent Variable: DX
Method: Least Squares
Date: 06/10/20 Time: 15:01
Sample (adjusted): 1/05/2000 6/08/2020
Included observations: 5329 after adjustments
Convergence achieved after 6 iterations
MA Backcast: 1/04/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.287420	0.638207	0.450356	0.6525
MA(1)	0.035725	0.013692	2.609094	0.0091

R-squared	0.001152	Mean dependent var	0.287371
Adjusted R-squared	0.000964	S.D. dependent var	45.00412
S.E. of regression	44.98241	Akaike info criterion	10.45080
Sum squared resid	10778746	Schwarz criterion	10.45327
Log likelihood	-27844.14	Hannan-Quinn criter.	10.45166
F-statistic	6.142814	Durbin-Watson stat	2.003442
Prob(F-statistic)	0.013225		

Inverted MA Roots -.04

Fig. 6. Parameter estimation of MA(1) model.

- **Establishing AR (1) model**

Dependent Variable: DX
Method: Least Squares
Date: 06/10/20 Time: 15:04
Sample (adjusted): 1/06/2000 6/08/2020
Included observations: 5328 after adjustments
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.286827	0.636920	0.450334	0.6525
AR(1)	0.032303	0.013695	2.358675	0.0184

R-squared	0.001043	Mean dependent var	0.286804
Adjusted R-squared	0.000856	S.D. dependent var	45.00832
S.E. of regression	44.98906	Akaike info criterion	10.45109
Sum squared resid	10779906	Schwarz criterion	10.45356
Log likelihood	-27839.71	Hannan-Quinn criter.	10.45195
F-statistic	5.563325	Durbin-Watson stat	1.996802
Prob(F-statistic)	0.018376		

Inverted AR Roots .03

Fig. 7. Parameter estimation of AR(1) model.

As shown in Fig. 7 above, the coefficient P is less than 0.05, it is considered that the coefficient passes the test and AR (1) model can be established.

- **Establishing AR (2) model**

Dependent Variable: DX
Method: Least Squares
Date: 06/10/20 Time: 15:08
Sample (adjusted): 1/07/2000 6/08/2020
Included observations: 5327 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.276707	0.608502	0.454734	0.6493
AR(1)	0.033765	0.013689	2.466605	0.0137
AR(2)	-0.045744	0.013689	-3.341693	0.0008

R-squared	0.003134	Mean dependent var	0.276672
Adjusted R-squared	0.002759	S.D. dependent var	45.00647
S.E. of regression	44.94434	Akaike info criterion	10.44929
Sum squared resid	10754446	Schwarz criterion	10.45300
Log likelihood	-27828.68	Hannan-Quinn criter.	10.45058
F-statistic	8.368051	Durbin-Watson stat	1.995124
Prob(F-statistic)	0.000235		

Inverted AR Roots .02+.21i .02-.21i

Fig. 8. Parameter estimation of AR(2) model.

As shown in Fig. 8, similarly, the coefficient P < 0.05 indicates that the AR (2) model can be established after the coefficient passes the test.

- **Establishing ARMA(1,1) model**

Dependent Variable: DX
Method: Least Squares
Date: 06/10/20 Time: 15:11
Sample (adjusted): 1/06/2000 6/08/2020
Included observations: 5328 after adjustments
Convergence achieved after 10 iterations
MA Backcast: 1/05/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.286772	0.626850	0.457480	0.6473
AR(1)	-0.852904	0.048010	-17.76530	0.0000
MA(1)	0.887473	0.042340	20.96074	0.0000

R-squared	0.004367	Mean dependent var	0.286804
Adjusted R-squared	0.003993	S.D. dependent var	45.00832
S.E. of regression	44.91838	Akaike info criterion	10.44813
Sum squared resid	10744044	Schwarz criterion	10.45184
Log likelihood	-27830.83	Hannan-Quinn criter.	10.44943
F-statistic	11.67731	Durbin-Watson stat	1.996675
Prob(F-statistic)	0.000009		

Inverted AR Roots -.85
Inverted MA Roots -.89

Fig. 9. Parameter estimation of ARMA(1,1) model.

As Fig. 9 shows, the P values of coefficients AR (2) and MA (1) are 0, which can be detected and ARMA (1,1) model can be established.

- **Establishing ARMA(1,2) model**

Dependent Variable: DX
Method: Least Squares
Date: 06/10/20 Time: 15:18
Sample (adjusted): 1/06/2000 6/08/2020
Included observations: 5328 after adjustments
Convergence achieved after 7 iterations
MA Backcast: 1/04/2000 1/05/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.287467	0.616814	0.466052	0.6412
AR(1)	-0.387575	0.221172	-1.752366	0.0798
MA(1)	0.423541	0.221173	1.914978	0.0555
MA(2)	-0.033081	0.019171	-1.725555	0.0845

R-squared	0.003988	Mean dependent var	0.286804
Adjusted R-squared	0.003427	S.D. dependent var	45.00832
S.E. of regression	44.93114	Akaike info criterion	10.44889
Sum squared resid	10748132	Schwarz criterion	10.45383
Log likelihood	-27831.84	Hannan-Quinn criter.	10.45062
F-statistic	7.105560	Durbin-Wa	2.000729
Prob(F-statistic)	0.000092		

Inverted AR Roots -.39
Inverted MA Roots .07 -.49

Fig. 10. Parameter estimation of ARMA(1,2) model.

Observing P values in Fig. 10, they were all greater than the significance level of 0.05, the coefficients could not pass the test, and the ARMA (1,2) model could not be established.

After that, the model needs to be selected. Among the five models mentioned above, it is not difficult to find that as long as AR (1), MA (1) and ARMA (1,1) models can pass the test, but the optimal model still needs to be determined. Therefore, AIC criterion and SC criterion can be used to further judge the advantages of the models, and the data values of the two criteria of each model are analyzed. The smaller the value, the better. The values are shown in Table I below:

TABLE I: AIC and SC Criterion

Model	AIC criteria	SC criteria
AR(1)	10.451	10.454
MA(1)	10.450	10.453
ARMA(1,1)	10.448	10.451

It can be seen from Table I that ARMA (1,1) meets the requirements of AIC and SC, and the minimum order is the optimal order. Therefore, it is thought that ARMA (1,1) is more suitable as the mean value equation of Shanghai stock index.

2) Model test (residual test)

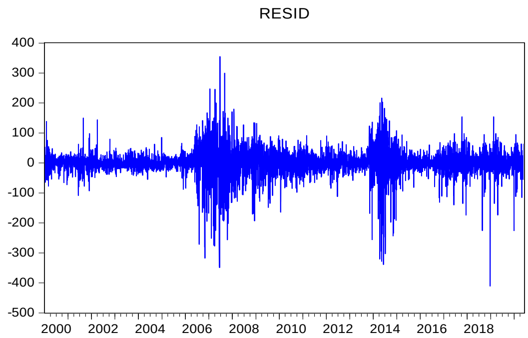

Fig. 11. Residual time series diagram.

By observing the sequence diagram of the residuals in Fig. 11, the author find that the data may be stable. It can be seen that the residuals show large and small fluctuations, and the fluctuation phenomenon has the following property, which indicates that the residual fluctuation has the aggregation, and further shows that the residual distribution is asymmetric and there may be conditional heteroscedasticity [8]. In order to prove this conjecture, the conditional heteroscedasticity test is carried out on this basis, that is, the ARCH effect test is carried out on the residuals. This test is shown in B 1)）

Date: 06/10/20 Time: 15:47
Sample: 1/04/2000 6/08/2020
Included observations: 5328

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.000	-0.000	0.0012	0.972
		2	0.003	0.003	0.0429	0.979
		3	0.034	0.034	6.2210	0.101
		4	0.072	0.072	33.666	0.000
		5	0.000	0.000	33.667	0.000
		6	-0.059	-0.061	52.004	0.000
		7	0.027	0.022	55.851	0.000
		8	0.010	0.006	56.408	0.000
		9	-0.003	0.001	56.443	0.000
		10	-0.018	-0.011	58.134	0.000
		11	0.031	0.027	63.125	0.000
		12	0.033	0.028	68.768	0.000
		13	0.056	0.060	85.298	0.000
		14	-0.038	-0.038	93.013	0.000
		15	0.048	0.041	105.23	0.000
		16	0.014	0.005	106.30	0.000
		17	-0.005	-0.007	106.41	0.000
		18	0.015	0.019	107.55	0.000
		19	-0.034	-0.037	113.82	0.000
		20	0.014	0.005	114.87	0.000
		21	0.004	0.012	114.97	0.000
		22	0.000	-0.000	114.97	0.000
		23	-0.025	-0.023	118.28	0.000
		24	0.014	0.010	119.38	0.000
		25	0.010	0.004	119.93	0.000
		26	-0.016	-0.015	121.22	0.000
		27	-0.004	-0.000	121.33	0.000
		28	0.042	0.034	130.89	0.000
		29	0.006	0.005	131.11	0.000

Fig. 12. Autocorrelation graph of residuals.

Null Hypothesis: F has a unit root
Exogenous: Constant
Lag Length: 5 (Automatic - based on SIC, maxlag=32)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-29.52152	0.0000
Test critical values:	1% level	-3.431398	
	5% level	-2.861888	
	10% level	-2.566998	

*MacKinnon (1996) one-sided p-values.

Fig. 13. Unit root test chart of residuals.

The author test the unit root of residuals and draw autocorrelation graphs. It can be found in Fig. 12 that the residuals have passed the unit root test, and the values of autocorrelation coefficient and partial autocorrelation coefficient in the autocorrelation graph are also very low. Therefore, we can believe that the residual has reached the level of white noise, which indicates that the AR (1) model has a good fitting degree and the model information has been fully utilized.

Redefining the residuals, the author test the unit root of the newly generated sequence as can be seen in Fig. 13.

B. GARCH Model

1) The selection of GARCH model

When GARCH model is used, the data dx of the first order difference is also used. Before fitting the GARCH model, it is necessary to observe whether the data has arch effect. [9] First, de average the sequence dx, that is, define w=dx-mean, and then conduct correlation test on the square of w to make the residual square correlation graph, as shown in Fig. 6-Fig. 10. Through the residual square correlation graph, we can find that P value is less than the significance level, the sequence has autocorrelation, not white noise time series, so there is ARCH effect. Because the unit root test results are stationary data, it can fit GARCH model.

2) Model testing

Before establishing GARCH model, ARCH effect should be tested first. In actual financial data processing, autoregressive conditional heteroscedasticity model is widely used, abbreviated as ARCH(q).

Investigating the variance homogeneity of the original time series, the residual time series after square processing obviously shows the characteristics of heteroscedasticity. At this time, it needs to be further processed. [10]

According to the above analysis results, the residual time series of AR model has heteroscedasticity, asymmetry and non-normality. Most of the previous studies focused on the assumption that the random error term obeys the normal distribution. In this paper, the author assume that the random perturbation term of the mean equation obeys the generalized error distribution (GED) distribution, and on the basis of this analysis, the author choose to add the wave term to the mean equation, that is, the GARCH-M model proposed by Engle et al. These models can not only describe the autoregressive conditional heteroscedasticity process, but also introduce the volatility into the corresponding regression equation. In addition to describing some other factors affecting the return of financial assets, they can also reflect the impact of return volatility on the return of financial assets.

① Fitting GARCH model

The commonly used GARCH models are GARCH (1,1), GARCH (1,2), GARCH (2,1), GARCH (2,2), therefore use each model to establish model:

a) GRACH(1,1) modeling

It can be seen from Fig. 14 that the coefficient P values are all 0, which can pass the test and the model can be established. GARCH (1,1) model is the most common model in this kind of analysis. Although it may not be the optimal model here, it is necessary to analyze it. It can be seen from GARCH model that a [1] is 0.055990, β [1] is 0.945578, and the sum of them is 0.9989, which is close to 1. At this time, it can be considered that the volatility shows high persistence. In addition, from the β [1] of 0.945578, it can be seen that the volatility has a very slow decay rate, which shows that volatility has obvious clustering characteristics. This can also show that during this period, when the fluctuation of Shanghai stock index is impacted, the fluctuation will be large and continuous.

Dependent Variable: W
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 06/10/20 Time: 16:43
Sample (adjusted): 1/05/2000 6/08/2020
Included observations: 5329 after adjustments
Convergence achieved after 50 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	3.551901	0.588425	6.036286	0.0000
RESID(-1)^2	0.055990	0.001757	31.86161	0.0000
GARCH(-1)	0.945578	0.001061	891.1664	0.0000

R-squared	-0.000041	Mean dependent var	0.287115
Adjusted R-squared	0.000147	S.D. dependent var	45.00412
S.E. of regression	45.00081	Akaike info criterion	9.786158
Sum squared resid	10791615	Schwarz criterion	9.789863
Log likelihood	-26072.22	Hannan-Quinn criter.	9.787452
Durbin-Watson stat	1.935311		

Fig. 14. Parameter estimation of GARCH(1,1) model.

b) GARCH(1,2) modeling

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	2.547698	0.453581	5.616850	0.0000
RESID(-1)^2	0.113718	0.006090	18.67361	0.0000
RESID(-2)^2	-0.069289	0.006653	-10.41474	0.0000
GARCH(-1)	0.956600	0.001261	758.7901	0.0000

R-squared	-0.000041	Mean dependent var	0.287115
Adjusted R-squared	0.000147	S.D. dependent var	45.00412
S.E. of regression	45.00081	Akaike info criterion	9.782602
Sum squared resid	10791615	Schwarz criterion	9.787541
Log likelihood	-26061.74	Hannan-Quinn criter.	9.784327
Durbin-Watson stat	1.935311		

Fig. 15. Parameter estimation of GARCH(1,2) model.

It can be seen that in Fig. 15 the coefficient P is 0, so the model can be established.

c) GARCH(2,1) modeling

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	4.746952	0.846762	5.606007	0.0000
RESID(-1)^2	0.076679	0.003169	24.19760	0.0000
GARCH(-1)	0.476881	0.078674	6.061508	0.0000
GARCH(-2)	0.448318	0.075993	5.899437	0.0000

R-squared	-0.000041	Mean dependent var	0.287115
Adjusted R-squared	0.000147	S.D. dependent var	45.00412
S.E. of regression	45.00081	Akaike info criterion	9.784513
Sum squared resid	10791615	Schwarz criterion	9.789453
Log likelihood	-26066.84	Hannan-Quinn criter.	9.786239
Durbin-Watson stat	1.935311		

Fig. 16. Parameter estimation of GARCH(2,1) model.

It can be seen in Fig. 16 that the coefficient P is 0, so the model can be established.

d) GARCH(2,2) modeling

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	8.531191	1.428245	5.973198	0.0000
RESID(-1)^2	0.055775	0.002272	24.55311	0.0000
RESID(-2)^2	0.077329	0.005672	13.63450	0.0000
GARCH(-1)	-0.004827	0.021876	-0.220645	0.8254
GARCH(-2)	0.877443	0.022512	38.97746	0.0000

R-squared	-0.000041	Mean dependent var	0.287115
Adjusted R-squared	0.000147	S.D. dependent var	45.00412
S.E. of regression	45.00081	Akaike info criterion	9.785909
Sum squared resid	10791615	Schwarz criterion	9.792084
Log likelihood	-26069.56	Hannan-Quinn criter.	9.788066
Durbin-Watson stat	1.935311		

Fig. 17. Parameter estimation of GARCH(2,2) model.

It can be seen in Fig. 17 that the p value of G1 is greater than the confidence level of 0.05, which cannot pass the test and the model cannot be established.

According to the test results, the sum of *α[i]* and *β[i]* in GARCH (1,1), GARCH (1,2), GARCH (2,1), GARCH (2,2) are 0.9846,1.0351,0.9851,0.9670 respectively, the difference is not significant, and there is no obvious sign of exceeding the expectation.

Nevertheless, only GARCH (1,1), GARCH (1,2), GARCH (2,1) coefficients all passed the test, and the AIC of the three were 9.7862, 9.7826, 9.7845, and SC were 9.7898, 9.7875 and 9.7894 respectively. Judging from the method of judging the minimum value, GARCH (1,2) model is better, so the model can be better fitted.

V. Prediction of the Model

A. Short Term Forecast

The author use the effective model ARMA (1,1) to make short-term forecast. There are two kinds of forecasting methods: dynamic forecast and static forecast. The former is multi-step forward prediction according to a certain estimation interval selected; the latter is only rolling forward prediction, that is, every time the prediction is made, the real value is replaced by the predicted value, added to the estimation interval, and then the forward prediction is made.

B. Dynamic Forecast

The predicted values are stored in the DXF sequence. At this time, it can be observed from Fig. 18 and Fig. 19 that the dynamic relationship between the original sequence dx and dxf. At the same time, select dx and dxf, right-click, click *open/as group*, and then click *view/graph/line*, and then the following figure will appear. Dynamic prediction is almost a straight line, which shows that the dynamic prediction effect is not good.

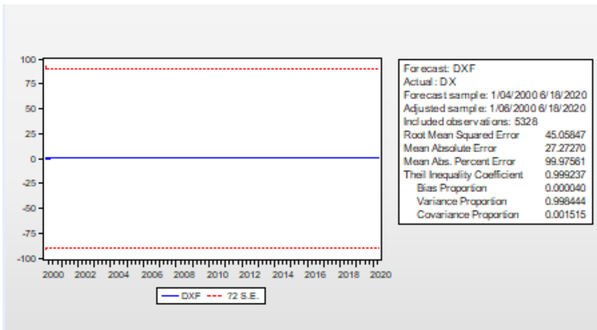

Fig. 18. Dynamic prediction chart.

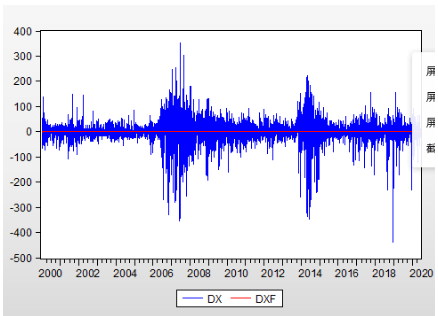

Fig. 19. Effect diagram of dynamic prediction.

C. Static Forecast

The static prediction is shown in the Fig. 20 below. The prediction is still stored in DXF. The DX and DXF charts show that the static prediction effect is good.

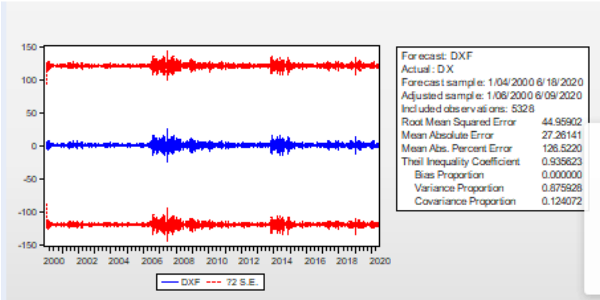

Fig. 20. Static prediction chart.

Forecast results: as shown in Fig. 21 below, the fluctuation range of static forecast price is not large in the increase of 8 days, which are all around the average value of 2901.34.

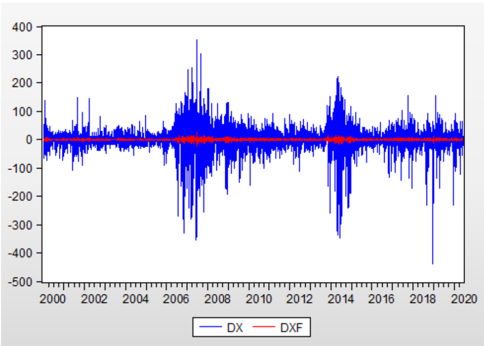

Fig. 21. Effect diagram of static prediction.

VI. Conclusion

To sum up, from the criterion of the smaller the better, it can found that GARCH (1,2) model is the best, so the author also adopt GARCH (1,2) model. In the model, it is considered that there is leverage effect (that is, leverage the larger capital with less capital to improve the yield) [11], and if a positive and negative impact of the same size can be exerted at this time, the negative impact will bring more obvious impact.

From the analysis, it is not difficult to see that the GARCH group model can well fit the stock market in China. From the model of this paper, it can be seen that the Prob. value of all index mean equation and variance equation is 0.0000. It can be said that under 1% confidence level, the Shanghai Composite Index of China can be well fitted by GARCH model, and the model is very significant. Secondly, the selected Shanghai stock index model shows the volatility state in line with the financial market. The sample data show the characteristics of peak thick tail and left deviation, as well as volatility clustering (high volatility and low volatility of the stock market tend to gather in a certain period of time, and their periods will appear alternately).

In addition, some scholars pointed out that it is a means of economic regulation for the government to adjust the price level through the stock market. The government hopes that the purpose of the rise of the stock market is to highlight direct financing, reduce the leverage of the real economy and obtain a slightly higher leverage in the capital market. In this sentence, it has been known from the GARCH (2,2) model above that there is indeed leverage effect. Therefore, comparing the analysis results with the actual situation, the author think that the results of the model are more reasonable.

Finally, the research of this paper still has some limitations, because the time series of stock price is quite complex. For the future research and analysis, the author can summarize its periodicity more accurately and forecast the stock price more accurately, so as to know the decision-making of managers and investors. In addition, only the closing price is selected as the predictor, and the sample number of stock price is not very large. The author only study ARMA model and GARCH model. These shortcomings need to be corrected in the follow-up research work to get a better prediction.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

This paper is independently completed by the author.

Acknowledgement

This paper is completed under the guidance of Professor Xian Sun and Miss Eva Chen. On the completion of this course thesis, I would like to express my gratitude to all the people who have helped me, cared for me and supported me. First of all, I would like to thank Professor Sun and Eva Chen. The profound knowledge, rigorous academic style, creative way of thinking, modest attitude and amiable personality charm of professor and teacher have had a profound impact on me and benefited me a lot. Thank you for your good intentions. Further, I would like to thank my parents and relatives who care about me. Your encouragement and support is the greatest spiritual motivation in my study.

References

[1] K. Meng and L. Li, “Empirical analysis of stock price prediction based on ARMA model,” *Journal of Hebei North University (Natural Science Edition)*, vol. 5, pp. 55-60, 2016.

[2] J. Li, “Research on the application of Shanghai stock index based on ARMA-ARCH model,” *Journal of Jilin Business and Technology College*, vol. 4, pp. 87-112, 2019.

[3] L. Wang and Z. She, “An empirical study on China's CPI fluctuation based on arch family model,” *Management and Administration*, vol. 7, pp. 119-123, 2019.

[4] D. W. Zhou, “Fitting and forecasting of stock index based on GARCH model in MATLAB--Taking Hang Seng Index as an example,” *Foreign Economic Relations & Trade*, vol. 4, pp. 77-80, 2020.

[5] L. X. Cui, “An empirical analysis of consumer finance driving economic growth in the internet era,” *Journal of Commercial Economics*, vol. 8, pp. 164-167, 2019.

[6] P. J. Zhao, Y. Hai, and Y. Yin, “Comparative study on volatility characteristics of gem index returns based on GARCH-VAR model,” *Value Engineering*, vol. 25, pp. 5-9, 2019.

[7] Y. Miao and J. M. Zhu, “Empirical analysis of Shanghai and Shenzhen 300 stock index futures arbitrage strategy based on GARCH,” *Journal of Beijing Institute of Graphic Communication*, vol. 5, pp. 77-80, 2019.

[8] G. L. Zhang, “Research on the relationship between gold and stock price index,” *Journal of Technical Economics & Management*, vol. 5, pp. 87-91, 2020.

[9] G. Q. Zeng, "Empirical analysis of Shanghai Shenzhen 300ETF fund based on GARCH-VAR model," *Times Finance*, vol. 5, pp. 178-179, 2017.
[10] Z. Q. Zhang, "Research on volatility characteristics of daily return of Shanghai and Shenzhen 300 index based on ARMA-GARCH model," *Shanxi Science and Technology*, vol. 1, pp. 43-46, 2016.
[11] X. F. Lu and H. R. Cui, "Volatility prediction of Shanghai and Shenzhen 300 index based on GARCH family hybrid model," *Yuejiang Academic Journal*, vol. 3, pp. 78-120, 2019.

Qi Qi was born in Ma'anshan, Anhui Province, China, on Jan. 17th, 1999. She is a senior undergraduate in Shihezi University, Shihezi, Xinjiang Uygur Autonomous Region (2017.09-prensent) with major of finance. She is expected for the bachelor degree in June, 2021.

The author worked as an assistant finance officer in full time in Ma'anshan Road and Bridge Engineering Co., Ltd as an intern from Jul. 20th to Aug. 14th, 2020. During summer vacation of 2019, the author lead campers from China to America for a traditional camp as a group leader. In October, 2019, the author published Survey and Analysis of Customer Satisfaction of Insurance Companies (China: Consume Guide. 2019(43): 79+81). Currently, the author's research interest focuses on fin-tech and financial engineering.