Research on Stock Price Forecast Based on ARIMA-GARCH Model

Jie Gao[1, *]
* Jie Gao: gaojie_1997@shu.edu.cn

[1]Institute of Economics, Shanghai University, Shanghai, China

Abstract. The stock plays a vital role in economic life, and the economic development of enterprises can be measured by the development and change of stocks. In this paper, the closing price of Ping An stock in China from 2017 to 2019 is selected as the time series empirical analysis data, and the ARIMA-GARCH model is established to predict the law and trend of the stock price change. The results show that the compound model can fit the fluctuation law well, and reasonably predict the short-term fluctuation trend.

Keywords: ARIMA-GARCH model, stock price forecast, Insurance industry

1 Introduction

With the continuous development of China's stock market, buying stocks has become one of the most important investment means. Accurately forecasting the stock price will help investors make correct investment choices, reduce the risk of loss and obtain reasonable returns. However, in the stock market, the change of stock price is affected by many complicated factors, which brings some difficulties to the accurate prediction of stock price.

Sun Shaoyan (2019) drew a conclusion that the middle price fluctuation in China's foreign exchange market has ARCH effect, and the short-term fluctuation trend of RMB exchange rate is reasonably predicted [1]. Wang Xiaoquan (2018) proposed a combined model ARIMA-GARCH-M to forecast short-term traffic flow, and tested the prediction accuracy [2]. Fang Yan (2018) used ARIMA-GARCH model to forecast the media sector index, and found that the model can be extended to individual stocks with the distribution characteristics of "peak and thick tail"[3]. Feng Tieying (2016) proposed an ARIMA-GARCH modeling and simulation method based on stochastic theory and time series analysis to estimate the future value and prediction interval of fertility rate [4].

In recent years, the role and influence of insurance industry has become increasingly significant. The rising status of insurance industry has attracted more and more attention and research from scholars and policy makers. The innovation of this paper lies in applying ARIMA-GARCH model to forecast the closing price of Ping An stock in China. This paper analyzes the historical closing price of Ping An in China, builds a stock price forecasting model based on ARIMA-GARCH, and looks forward to the future development prospects of insurance companies and insurance industry. The results show that the model has a good prediction effect.

2 Quantitative Analysis

2.1 Model Introduction

C. P. Box and G. M. Jenkins first proposed the differential integration moving average autoregressive model (ARIMA model for short), Let $\{Y_t\}$ be a stationary time series with zero mean, and ARMA(p, q) formula of p-order autoregressive q-order moving average is expressed as follows:

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{1}$$

D in ARIMA (p, d, q) model is the order of difference. Time series data in financial market are generally non-stationary. Difference is one of the ways of stabilization.

GARCH is an important extension of ARCH model variance, mainly aiming at regression model of financial data and analysis and prediction of volatility, which greatly reduces the complexity of the model and makes the results look simpler and clearer. GARCH model structure is:

$$y_t = c + \sum_{i=1}^m \phi_i y_{t-i}^2 + \sum_{j=1}^n \phi_j \mu_{t-j}^2 + \mu_t \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \mu_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \tag{3}$$

2.2 Preliminary Analysis and Processing of Data

This paper selects the stock closing price of Ping An (601318) from July 3, 2017 to July 3, 2019, and the data of the next four trading days as the test data. The company is a leading enterprise in the insurance industry, and no major equity change occurred during this period.

Using Eviews software to analyze the data preliminarily, as shown in Figure 1 and Figure 2, it is found that the data are non-stationary time series. Therefore, it is necessary to use first-order or multi-order differences to stabilize the unstable original data [5, 6]. It is found that the absolute value of T statistic is larger than the absolute value of critical value under each confidence level, and the P value is 0.0000, so the data after first-order difference is a stationary time series.
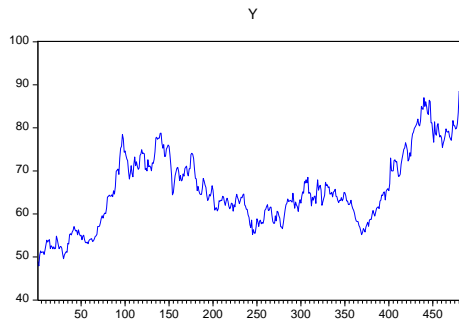

Fig 1 Original sequence timing diagram

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.569684	0.8036
Test critical values:	1% level	-3.976974	
	5% level	-3.419056	
	10% level	-3.132086	

Fig 2 ADF test of original sequence
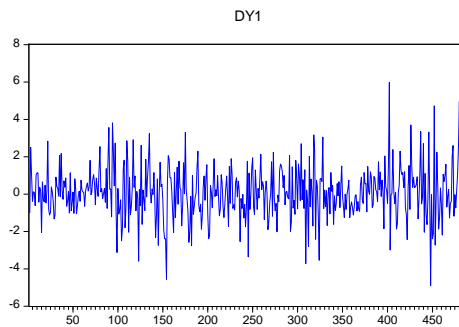
Fig 3 First-order differential sequence timing diagram

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-21.41472	0.0000
Test critical values:	1% level	-3.977013	
	5% level	-3.419076	
	10% level	-3.132097	

Fig 4 ADF test of first-order difference sequence

2.3 Selection of Models

The ARMA model is used to fit the data after the first-order difference, and the parameters are estimated according to ACF diagram and PACF diagram (Figure 5). Comparing the test statistics of each model, combining the significance of the coefficient corresponding to the maximum lag variable of each model and the minimum AIC information criterion value, it is judged that ARIMA (3,1,4) is superior to the other three models.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.028	0.028	0.3791	0.538
		2	-0.035	-0.036	0.9828	0.612
		3	0.057	0.059	2.5864	0.460
		4	-0.109	-0.115	8.5097	0.075
		5	-0.050	-0.039	9.7435	0.083
		6	-0.089	-0.100	13.636	0.034
		7	0.079	0.098	16.735	0.019
		8	-0.039	-0.063	17.491	0.025
		9	0.076	0.095	20.394	0.016
		10	0.084	0.037	23.904	0.008
		11	-0.063	-0.041	25.885	0.007
		12	0.031	0.016	26.379	0.009
		13	-0.002	0.015	26.381	0.015
		14	-0.076	-0.067	29.281	0.010
		15	0.027	0.050	29.653	0.013
		16	0.013	-0.001	29.741	0.019
		17	0.008	0.011	29.770	0.028
		18	-0.036	-0.046	30.429	0.033
		19	-0.036	-0.049	31.099	0.039
		20	0.025	0.021	31.409	0.050
		21	-0.008	0.017	31.442	0.067
		22	0.024	0.005	31.730	0.082
		23	0.022	0.025	31.980	0.101
		24	0.006	0.001	31.996	0.127
		25	-0.008	-0.021	32.029	0.157

Fig 5 ACF and PACF diagram of first-order difference sequence

Table 1 Comparison of precision indexes of various models

Model	AIC
ARIMA (1,1,1)	3.454412
ARIMA (2,1,2)	3.458117
ARIMA (4,1,3)	3.453131
ARIMA (3,1,4)	3.450976

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.082007	0.057376	1.429278	0.1536
AR(1)	0.533361	0.148490	3.591887	0.0004
AR(2)	0.062946	0.152541	0.412649	0.6800
AR(3)	-0.554211	0.142038	-3.901836	0.0001
MA(1)	-0.492449	0.149946	-3.284169	0.0011
MA(2)	-0.142222	0.136901	-1.038866	0.2994
MA(3)	0.689295	0.128193	5.377032	0.0000
MA(4)	-0.155819	0.054097	-2.880353	0.0042

Fig 6 Regression results of ARIMA (3,1,4) model

Regression results show that the estimated values of other coefficients are significant at the confidence level of 5%, so we think that the selected model is more suitable. Therefore, ARIMA (3,1,4) model is selected as the final mean equation, namely:

$$dy_t = 0.82007 + 0.533361 dy_{t-1} - 0.554211 dy_{t-3} + \varepsilon_t - 0.492449 \varepsilon_{t-1} + 0.689295 \varepsilon_{t-3} - 0.155819 \varepsilon_{t-4} \tag{4}$$

After obtaining the mean value equation, it is necessary to test the residual sequence of the equation with white noise. It can be observed from Figure. 7 that there is obvious "clustering" phenomenon in the fluctuation, and the fluctuation is very small in some long time periods, but very large in other long time periods, so it can be considered that there is a high probability of conditional heteroscedasticity in the error term.
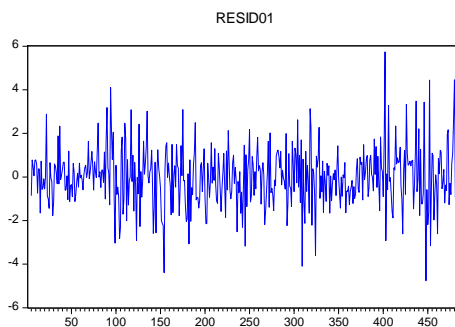

Fig 7 Residual timing diagram

2.4 Test of the ARCH Effect

We need to further test whether the residual conditional heteroscedasticity exists or not. Therefore, the mean equation of ARIMA (3, 1, 3) is tested by ARCH LM with conditional

heteroscedasticity, and the result is shown in Figure. 8. The p value here is relatively small, so the original hypothesis can be rejected, that is, the ARCH effect exists in the residual sequence of the equation.

Heteroskedasticity Test: ARCH

F-statistic	3.906151	Prob. F(1,482)	0.0487
Obs*R-squared	3.890828	Prob. Chi-Square(1)	0.0486

Fig 8 ARCH LM test results

2.5 Establishment of the ARIMA-GARCH Model

Because the residual error in ARIMA (3,1,4) model has conditional heteroscedasticity, we should establish the volatility equation for the residual error of the model and choose GARCH model to explain the conditional heteroscedasticity. From the above explanation, we set up a model combining ARIMA (3,1,4) and GARCH (1,1) to fit the closing price of Ping An in China. The fitting result is shown in Figure 9.

Variable	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.516687	0.152563	3.386721	0.0007
AR(2)	0.014222	0.162071	0.087751	0.9301
AR(3)	-0.577827	0.142550	-4.053515	0.0001
MA(1)	-0.444338	0.165530	-2.684337	0.0073
MA(2)	-0.088219	0.135490	-0.651111	0.5150
MA(3)	0.727643	0.113082	6.434649	0.0000
MA(4)	-0.110801	0.065540	-1.690593	0.0909
Variance Equation				
C	0.073411	0.033207	2.210703	0.0271
RESID(-1)^2	0.077633	0.027544	2.818548	0.0048
GARCH(-1)	0.886595	0.038631	22.95039	0.0000

Fig 9 Regression results of ARIMA-GARCH model

From the regression results, we can see that the coefficients of AR (2), MA (2) and MA (4) failed the significance test, so the final estimation model should be:

$$dy_t = 0.516687dy_{t-1} - 0.577827dy_{t-3} + \varepsilon_t - 0.444338\varepsilon_{t-1} + 0.727643\varepsilon_{t-3} \quad (5)$$

$$\varepsilon_t = \sigma_t a_t \quad (6)$$

$$\sigma_t^2 = 0.073411 + 0.077633\varepsilon_{t-1}^2 + 0.886595\sigma_{t-1}^2 \quad (7)$$

The final ARIMA-GARCH model is tested by LM to see if the residual is irrelevant. The results in Figure 10 show that the original hypothesis that the residual sequence conforms to the white noise process can't be rejected. Therefore, there is no need to modify the established model.

Heteroskedasticity Test: ARCH

F-statistic	0.127830	Prob. F(1,482)	0.7208
Obs*R-squared	0.128327	Prob. Chi-Square(1)	0.7202

Fig 10 Heteroscedastic test of the model

2.6 Model Prediction and Analysis

Due to the limitations of the model, only the comparison between the real values of the four

periods and the predicted values is given here, and the results are shown in Table 2.

Table 2 Model Prediction Results

Date	Real value	Predicted value	Error	Relative error (%)
20190704	89.9	90.98	-1.08	1.2
20190705	90.85	90.16	0.69	0.7
20190708	88.59	89.69	-1.10	1.2
20190709	87.53	89.25	-1.72	1.97

Average relative error

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{4}\frac{|y_i - \bar{y}|}{y_i} = 1.29\% \tag{8}$$

3 Conclusions

It is not difficult to draw the following conclusions from the analysis of prediction results:

(1) The ARIMA-GARCH model used in this paper solves the modeling problem of non-stationary time series, and achieves good results in short-term prediction of data, which shows that the model is feasible. Although the static model is used in this paper, it is theoretically possible to make dynamic prediction, which makes the model more flexible.

(2) Only random factors can affect the predicted value of stock price. If something happens to the company, or if the international or domestic market environment fluctuates, the sample data will change, and the model parameters will also change, which shows that the model is still very sensitive. Therefore, when selecting data, it is necessary to prevent market fluctuations and various uncertain factors.

(3) To further promote the development of the insurance industry, first, we must ensure the stability of the development of the industry. Enterprises themselves should pay attention to management, the market, adapt to changes, constantly innovate product types, and have a sense of urgency in times of peace. Then, government departments should strengthen supervision and appropriate guidance, safeguard the safety of people's property, and stabilize the development of the insurance industry [7].

Second, improve the efficiency and quality of insurance claims. This requires optimizing the claims process and improving the claims quality. At the same time, we should pay attention to the training of employees in the claims department, and further improve their service quality and skill level.

Third, improve policies and regulations. Because policies and regulations lag behind the development of the industry, resulting in more regulatory loopholes and inadequate supervision. Therefore, we can look at the world and learn from the appropriate policies of other countries to supervise China's insurance industry to ensure its normal and efficient operation [8].

References

[1]   SY. Sun, WX. Sun, *Study on the fluctuation law of RMB exchange rate after joining SDR —— Empirical analysis based on ARIMA-GARCH model* [J]. Economic Problems, 02, 42-47, (2019).

[2]   Wang Xiaoquan, Shao Chunfu, Yin Chaoying, Ji Xun, Guan Ling. Short-term traffic flow prediction method based on ARIMA-GARCH-M model [J]. Journal of Beijing Jiaotong University, **42**, 04, 79-84, (2018).

[3]   Fang Yan, Geng Xueyang, Qin Shanshan. Study on the price forecast of media sector index in Shanghai and Shenzhen stock markets —— Analysis based on ARIMA—GARCH model [J]. Price Theory and Practice, 01, 102-105, (2018).

[4]   Feng Tieying, Luo Tianheng. Stochastic prediction of fertility rate based on ARIMA-GARCH model [J]. statistics and decision, 24, 21-24, (2015).

[5]   T. Ding, Feng Donghan, Lin Xiaofan, Chen Jingwen, Chen Lixia. Ultra-short-term wind speed prediction based on modified ARIMA-GARCH model [J]. Power Grid Technology, **41**, 06, 1808-1814, (2017).

[6]   Dong Xiaohong. Research on Financial Mismatch and Loan Default Rate Fluctuation of Commercial Banks —— Empirical Test Based on ARIMA-GARCH Model [J]. Jianghan Forum, 05, 23-27, (2016).

[7]   Sun Qixiang, Fan Juanjuan. Experience of Insurance Development in New China [J]. China Finance, 19, 115-118, (2019).

[8]   Hao Chen, Qian Jing, Fu Jinwei, Cui Guangyao. Research on the Development and Optimization of Insurance Governance in China [J]. Southwest Finance, 01, 41-50, (2018).