Proceedings of CECNet 2022
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220575

Prediction and Analysis of Stock Logarithmic Returns Based on ARMA-GARCH Model

Hao Qi [1] and Yuanshen Wang
School of Mathematics and Big Data, Chongqing University of Arts and Sciences, China

Abstract. With the rapid development of economy, investment has become a hot word. Many people hope to find an investment method to make profits. Many investment methods such as stocks, wealth management and funds have emerged. In the process of investment, forecasting the trend of investment products is one of the most important links. This paper analyzes the time series of APPLE, AMERICAN AIRLINES and AMD based on ARMA-GARCH model, and evaluates the model according to AIC, BIC, HQIC and other indicators to select the optimal model. The research results show that ARMA (3,2) - GARCH (1,1) model is applicable to APPLE stock logarithmic profit Prediction, ARMA (2,2) model is applicable to AMERICAN AIRLINES stock logarithmic profit prediction, and ARMA (2,2) - GARCH (1,1) is applicable to AMD stock logarithmic profit prediction.

Keywords. Stock, Return, Prediction, ARMA, GARCH

1. Introduction

With the development of economy and the improvement of residents' income, residents hope to find an investment method that can be used as a sideline, such as stocks, funds and real estate. At the same time, investors are also very careful to choose the investment products, hoping to maximize profits (see [1]). Therefore, before investing, investors need to predict the price of the investment product in the future according to the price of the investment product in the past period to predict whether the purchase of the product will generate profits and evaluate the investment risk.

This paper mainly uses Autoregressive Moving Average Model and Generalized Autoregressive Conditional Heteroskedasticity Model(ARMA-GARCH) to predict the APPLE stock returns, American Airlines and Advanced Micro Devices(AMD). The ARMA model can predict stationary time series data, and it reduces the steps of difference in the ARIMA model proposed by Box, G. and Jenkins, G (see [2]). In 1982, the Model was proposed by Engle was ARCH, and this model was suitable for heteroscedastic financial time series. Bollerslev proposed a model called GARCH in 1986, and established a conditional heteroscedastic time series model of investment price and return in 1987 (see [3-5]).

[1] Corresponding Author, Hao Qi, *Chongqing University of Arts and Sciences, China*; Email: qih33333@163.com

2. Related Works

Stock is a very common option in investment. Many people hope to gain income by investing in stocks. The information measurement of expected investment is very important in stock trading. Shareholders need to know the daily closing price, increase and other important information of stocks in real time (see [6]).

As stock forecasting is a hot research field, Jörg Gottschlich and Oliver Hinz [7] use BFO and ABFO technologies to develop a prediction model with better accuracy and faster convergence than those based on genetic algorithm (GA) and particle swarm optimization (PSO) for stock prediction (see [7]). Using historical information of the stock closing price to predict the future stock price trend by using random walk model is very difficult. Therefore, Ritika Singh and Shashi Srivastava respectively built three kinds of neural networks, including deep neural network (DNN), to predict stock returns, and concluded that DNN has a better prediction effect than RBFNN and RNN (see [8]). Wenjie Lu et al. combined the Convolutional neural networks (CNN), bi directional long short term memory (BiLSTM) and Attention Mechanism (AM) to form the CNN-BiLSTM-AM stock price prediction. The experimental results show that CNN-BiLSTM-AM has the highest accuracy in predicting stock returns compared with the prediction results of eight methods, including MLP, CNN and RNN (see [9]).When selecting investment products such as stocks, traders need to use a variety of forecasting techniques to obtain information about the stock market, rather than a single method. Phichhang Ou discussed ten different data mining technologies, such as linear discriminant analysis (LDA), support vector machine (SVM) and least squares support vector machine, and applied them to the price trend prediction of Hang Seng Index in Hong Kong stock market. The experimental results show that support vector machine and least squares support vector machine have superior performance in predicting the price trend of Hang Seng Index in Hong Kong stock market (see [10]).

The ARMA-GARCH model has a wide range of applications, in which ARMA can predict stationary time series data. Before analyzing and predicting the time series data, it is necessary to perform unit root test (ADF test) on the data. When the data is verified to be a stationary time series, ARMA model can be used for prediction. Jingli Yang and Xiaofeng Zhou will apply ARMA model to PM 2.5. The results showed that the ARMA model can achieve the expected prediction effect (see [11]). Mohammad Valipour et al. used ARMA model and ARIMA model to predict the inflow of Dez Dam Reservoir every month from 1960 to 2007 (see [12]). J. L. Torres et al. verified that the wind speed prediction performance of ARMA is significantly better than that of the persistence model, especially in the long-term prediction (see [13]).

When ARMA model is used alone, the variance equation will be ignored and many parameters need to be estimated. The combination of ARMA model and GARCH model can overcome this shortcoming, and GARCH model can fully describe the volatility process of asset returns (see [14]). During the investigation of ARMA-GARCH model, we found that Heping Liu and Jing Shi used ARMA-GARCH model to forecast electricity prices (see [15]). Alicja Lojowska et al. used ARMA-GARCH model to predict wind speed (see [16]). ARMA-GARCH can capture the most important features of the data in a satisfactory way, such as distribution, time related structure and periodicity. The experimental results show that by simulating the wind speed with extremely high energy content characteristics, this model can help to consider the extreme situation of wind power generation (see [12]).

3. Methods

3.1 ARMA Model

ARMA (p, q) model is mainly aimed at stationary time series data. It is composed of AR (p) model and MA (q) model. p is the order of automatic regression process, and q is the order of the sliding average process. This model takes into account the values, prediction errors and random items of data in the past period. The basic model is shown in (1), α_m and β_n are autocorrelation coefficients and moving average coefficients, respectively.

$$R_t = \sum_{m=1}^{p} \alpha_m R_{t-m} + \varepsilon_t + \sum_{n=1}^{q} \beta_n \varepsilon_{t-n} \quad (1)$$

$\{R_t\}$ represents stock logarithmic returns time series data of logarithmic stock returns, $\{\varepsilon_t\}$ represents the error terms. The model can also be expressed in the form shown in (2).

$$R_t = \alpha_1 R_{t-1} + \alpha_2 R_{t-2} \cdots + \alpha_p R_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots \beta_q \varepsilon_{t-q} \quad (2)$$

The stationarity is a necessary condition for forecasting stock data with ARMA model. Therefore, before building the ARMA model, it is necessary to perform ADF test on the time series data of stocks (see [17]). After the test data is stationary time series data, ARMA can be used to analyze and predict the data. Since the average value of the logarithmic profit data of stocks is close to zero and the variance is constant, the logarithmic profit data of stocks are considered as stationary time series data (see [18]).

For the established model, we can evaluate the excellence of the ARMA model according to the value of AIC criteria (see [17]). The value of AIC calculation formula is shown in (3), L represents the maximum likelihood of the model, and k represents the number of variables.

$$AIC = -2\ln(L) + 2k \quad (3)$$

In the process of establishing the model, multiple groups of p and q values can be selected for modeling, AIC of each model can be calculated, and the model with the lowest AIC value can be selected as the prediction model for the logarithmic return of stocks.

The flow chart of ARMA model for time series prediction is shown in Figure 1.

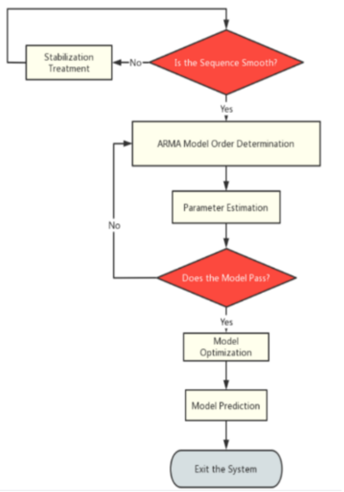

Figure 1. ARMA Flow Chart

3.2 GARCH Model

GARCH(i, j) is expressed by (4) and (5), a_p and b_q are coefficients, σ^2 represents the conditional standard deviation at time t, z_t represents the white noise with $\mu = 0$ and unit variance.

$$\sigma_t^2 = a_0 + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \cdots + a_i + b_1\sigma_{t-1}^2 + b_2\sigma_{t-2}^2 + \cdots + b_q\sigma_{t-p}^2 \quad (4)$$

$$\varepsilon_t = z_t\sigma_t \quad (5)$$

$$0 < \sum_{m=1}^{i} a_m + \sum_{n=1}^{j} b_n < 1 \quad (6)$$

In order to select the optimal GARCH model, we can choose three criterion including Akaike Information Criterion (AIC) (see [17]), Bayesian Information Criterion (BIC), and Hannan Quinn Information Criterion (HQIC) to evaluate the GARCH model in this paper (see [19]). The calculation formulas of AIC, BIC, and HQIC are shown in (3), (7), and (8), k represents the number of variables, N represents the number of observations, L represents the maximum likelihood of the model.

$$BIC = ln(N)\,k - 2ln\,(L) \quad (7)$$

$$HQIC = 2\,ln(ln(N))\,k - 2ln\,(L) \quad (8)$$

3.3 ARMA-GARCH Model

According to the model ARMA(p, q) and the model GARCH(i, j), we can build up the model ARMA(p, q)-GARCH(i, j). Before constructing the GARCH model, it is necessary to test the ARCH effect of the sequence. The GARCH model can be used when there is no ARCH effect. The standard test is the p value of chi square test. When the P value is less than 0.05, it indicates that the time series does not have ARCH effect, and GARCH modeling can be performed. ARMA (p, q) - GARCH (i, j) model models the mean and variance. ARMA (p, q) model models the mean, and GARCH (i, j) model models the variance.

4. Experimental Analysis and Results

4.1 Data Set

The data set used in the experiments involved in this paper is the stock closing prices of APPLE Inc(APPLE), AMERICAN AIRLINES and Advanced Micro Devices(AMD) from January 4, 2021 to September 30, 2022. The original data is shown in Table 1, and the data distribution and data trend are shown in Figure 2.

Table 1. Original Data

APPLE		AMERICAN AIRLINES		AMD	
Date	**Closing Price**	**Date**	**Closing Price**	**Date**	**Closing Price**
2021-1-4	129.41	2021-1-4	15.13	2021-1-4	92.30
2021-1-5	131.01	2021-1-5	15.43	2021-1-5	92.77
2021-1-6	126.60	2021-1-6	15.52	2021-1-6	90.33
2021-1-7	130.92	2021-1-7	15.38	2021-1-7	95.16
2021-1-8	132.05	2021-1-8	15.13	2021-1-8	94.58
……		……		……	
2021-9-13	149.55	2021-9-13	19.31	2021-9-13	104.80
2021-9-14	148.12	2021-9-14	19.21	2021-9-14	105.73
2021-9-15	149.03	2021-9-15	19.38	2021-9-15	105.60
2021-9-16	148.79	2021-9-16	19.89	2021-9-16	106.22
2021-9-17	146.06	2021-9-17	19.73	2021-9-17	103.88
……		……		……	
2022-9-26	150.77	2022-9-26	11.86	2022-9-26	66.30
2022-9-27	151.76	2022-9-27	12.27	2022-9-27	67.17
2022-9-28	149.84	2022-9-28	12.75	2022-9-28	68.36
2022-9-29	142.48	2022-9-29	12.25	2022-9-29	64.14
2022-9-30	138.20	2022-9-30	12.04	2022-9-30	63.36

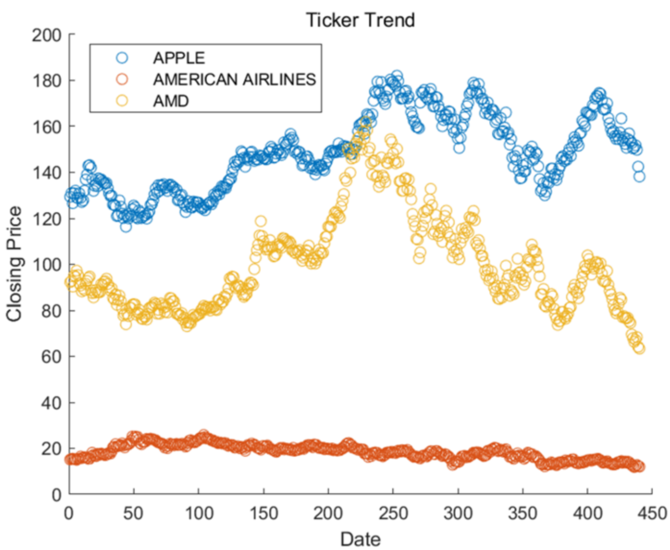

Figure 2. Data Distribution and Trend

4.2 Data Preprocessing

In order to make the data meet the assumptions of the linear model, we convert the stock closing price data into logarithmic income receipts, and apply ARMA-GARCH model to predict the logarithmic income of Apple, American Airlines and AMD. The logarithmic return formula is shown in (9), X_t represents the closing price on the t_{th} day.

$$R_t = ln(X_t) - ln(X_{t-1}) = ln\,(\frac{X_t}{x_{t-1}}) \qquad (9)$$

The trends of logarithmic return time series data of APPLE, AMERICAN AIRLINES and AMD are shown in Figure 3 to Figure 5, respectively.

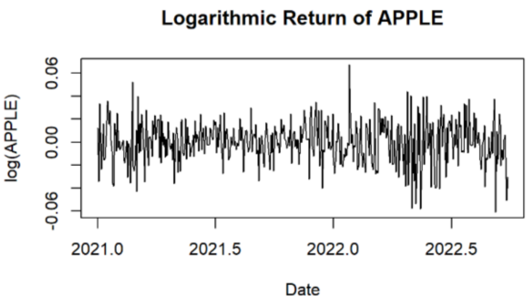

Figure 3. Logarithmic return of APPLE

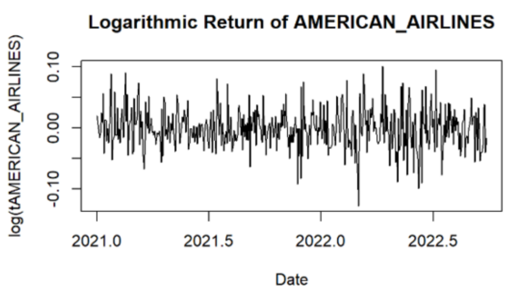

Figure 4. Logarithmic return of AMERICAN AIRLINES

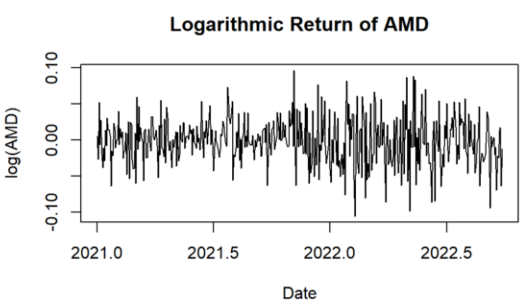

Figure 5. Logarithmic return of AMD

4.3 ARMA

Since ARMA model can predict and analyze stationary time series data, ADF test is required for data before ARMA model is applied to predict and analyze time series data. When the p-value obtained through the test is less than 0.01, the original assumption that the data is an unstable time series is rejected, indicating that the data is a stable time series data, which can be predicted and analyzed using ARMA model. The ADF Test Results table is shown in Table 2.

Table 2. ADF Test Results

	APPLE	AMERICAN AIRLINES	AMD
		ADF Test	
p-value	<0.01	<0.01	<0.01

According to the ADF results shown in Table 2, the p-value of APPLE and AMERICAN AIRLINES is less than 0.01. The original assumption is rejected, which indicates that the logarithmic return data of the three stocks are stable time series data, and ARMA model can be used for prediction and analysis.

In order to ensure the accuracy of prediction and analysis, eight ARMA models are selected for data modeling and analysis, and the AIC of the model is calculated ($N = 440, k = p + q + 2$), p represents the autoregressive term, and q represents the number of moving average terms. The model with the minimum AIC is selected as the optimal model. All the AIC values of the eight models are shown in Table 3.

Table 3. AIC Values of ARMA(p, q)

	APPLE	AMERICAN AIRLINES	AMD
Model		**AIC**	
$ARMA(0,1)$	-2253.98	-1735.26	-1756.76
$ARMA(1,1)$	-2251.94	-1733.27	-1755.03
$ARMA(1,2)$	-2249.98	-1731.68	-1753.04
$ARMA(2,1)$	-2251.18	-1731.63	-1753.04
$ARMA(2,2)$	-2251.92	-1737.32	-1757.97
$ARMA(3,1)$	-2250.37	-1732.37	-1751.22
$ARMA(3,2)$	-2255.88	-1728.59	-1749.05
$ARMA(3,3)$	-2250.02	-1735.37	-1754.76

According to the AIC of the three stocks obtained by applying the ARMA model, select the ARMA (3, 2) model for APPLE stocks, ARMA (2, 2) model for AMERICAN AIRLINES, and ARMA (2, 2) model for AMD stocks. The prediction results are shown in Figure 6, Figure 7 and Figure 8.

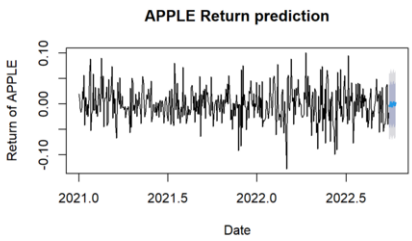

Figure 6. Prediction of APPLE Stock Return

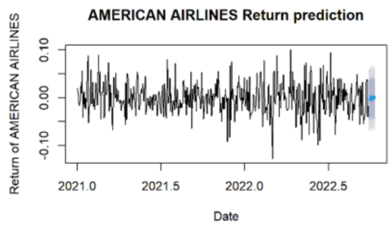

Figure 7. Prediction of AMERICAN AIRLINES Stock

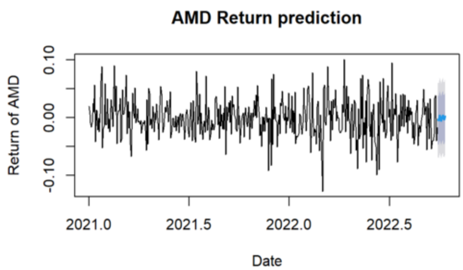

Figure 8. Prediction of AMD Stock Return

4.4 GARCH

Before establishing the GARCH model, it is necessary to test the data for the existence of the ARCH effect. If the p-value obtained through the test is less than 0.01, the original assumption that the data has the ARCH effect is rejected, indicating that the data does not have the ARCH effect, and the GARCH model can be established. The test results of ARCH effect are shown in Table 4.

Table 4. ARCH Test Results

	APPLE	AMERICAN AIRLINES	AMD
		ARCH Test	
p-value	0.0006408	0.07996	3.436e-06

According to the test results of the ARCH effect, the p-value of the logarithmic return data of APPLE and AMD stocks is less than 0.01. Rejecting the original hypothesis, it can be explained that there is no ARCH effect in the logarithmic return data of APPLE and AMD stocks. GARCH model can be used to analyze the logarithmic return data of APPLE and AMD stocks. However, the p-value of the AMERICAN AIRLINES stock logarithmic return data is 0.07996, greater than 0.01, and the original assumption cannot be rejected. Therefore, it cannot be explained that there is no ARCH effect in the AMERICAN AIRLINES stock logarithmic return data, so the GARCH model cannot be used to analyze the AMERICAN AIRLINES stock logarithmic return data.

In order to test the accuracy of GARCH, the GARCH (1,1) model was established. GARCH (1,2) model and GARCH (1,3) model were used to model and analyze APPLE and AMD stocks. The AIC, BIC and HQIC obtained from the three types of models are gradually increasing, so GARCH (1,1) is selected to model and analyze two stocks.

The evaluation table for modeling and analyzing APPLE and AMD stocks using GARCH model is shown in Figure 5 and Figure 6. It can be seen intuitively that GARCH (1,1), GARCH (1,2) and GARCH (1,3) have the same evaluation results on the modeling and analysis of APPLE stock. In the evaluation results of AMD stock modeling and analysis, GARCH (1,1) is slightly better than GARCH (1,2) and GARCH (1,3). So we choose GARCH (1,1) model to model and analyze APPLE and AMD stocks.

Table 5. Model of APPLE

APPLE			
Model	**AIC**	**BIC**	**HQIC**
$\boldsymbol{GARCH(1,1)}$	-5.215375	-5.178158	-5.200691
$\boldsymbol{GARCH(1,2)}$	-5.215375	-5.178158	-5.200691
$\boldsymbol{GARCH(1,3)}$	-5.215375	-5.178158	-5.200691

Table 6. Model of ADM

AMD			
Model	**AIC**	**BIC**	**HQIC**
$\boldsymbol{GARCH(1,1)}$	-4.066688	-4.029472	-4.052005
$\boldsymbol{GARCH(1,2)}$	-4.061805	-4.015284	-4.043451
$\boldsymbol{GARCH(1,3)}$	-4.057147	-4.001316	-4.035116

4.5 ARMA-GARCH

To sum up, the model selected to predict the logarithmic returns of APPLE, AMERICAN AIRLINES and AMD stocks is shown in Table 7.

Table 7. Model Selection Results

Stock	Model
APPLE	$\boldsymbol{ARMA(3,2) - GARCH(1,1)}$
AMERICAN AIRLINE	$\boldsymbol{ARMA(2,2)}$
AMD	$\boldsymbol{ARMA(2,2) - GARCH(1,1)}$

4.6 Model Accuracy

We can verify the accuracy of the model, by detecting the error of the prediction results. The error of the results selected by APPL, AMERICAN AIRLINES and AMD are shown in Figure 9 to Figure 11, respectively. Take standard deviation as the main error measurement standard, we can find that the prediction errors of APPLE and AMD is less than 0.03, and the prediction error of AMERICAN AIRLINES stocks is less than 0.1, which indicates that the model is accurate and effective in predicting the logarithmic returns of stocks.

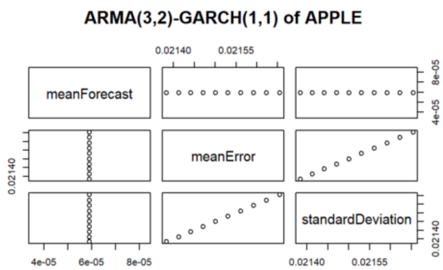

Figure 9. The Error of APPLE

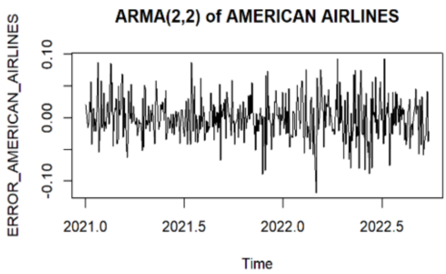

Figure 10. The Error of AMERICAN AIRLINES

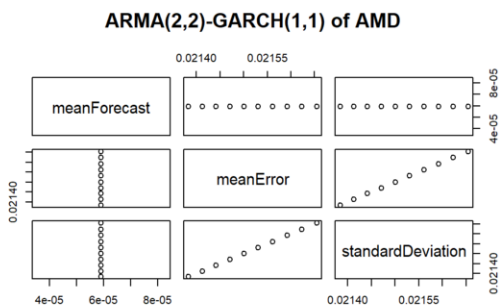

Figure 11. The Error of AMD

5. Conclusions

This study set out to apply time series analysis to predict and analyze APPLE, AMERICAN AIRLINES and AMD. This paper combines ARMA (p, q) model and GARCH (i, j) model for modeling and analysis. The results of time series analysis show that ARMA (3,2)-GARCH (1,1) model is applicable to APPLE stock, ARMA (2,2) model is applicable to AMERICAN AIRLINES stock, and ARMA (2,2)-GARCH (1,1) is applicable to AMD stock. In future work, we will continue to collect the closing price data of the three stocks to continuously improve the accuracy of the prediction.

Reference

[1] Mak, M. K., & Ip, W. (2017). An exploratory study of investment behaviour of investors. International Journal of Engineering Business Management, 9, 184797901771152. doi: 10.1177/18 47979017711520
[2] Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
[3] Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation [J]. Econometrica. 1982, 50: 987-1008.
[4] Bollerslev, Tim.Generalized Autoregressive Conditional Heteroscedasticity [J]. Journal of Econometrics, 1986, 31: 307-327.

[5] Bollerslev. A Conditional Heteroskedasticity Time Series Model for Speculative Prices and Rates of Return [J]. Review of Economics and Statis-tics. 1987, 69: 542-547.
[6] Lamont, O. A. (2000). Investment Plans and Stock Returns. The Journal of Finance, 55(6), 2719–2745. doi:10.1111/0022-1082.00304
[7] Gottschlich, J., & Hinz, O. (2014). A decision support system for stock investment recommendations using collective wisdom. Decision Support Systems, 59, 52–62. doi:10.1016/j.dss.2013.10.00
[8] Singh, R., & Srivastava, S. (2016). Stock prediction using deep learning. Multimedia Tools and Applications, 76(18), 18569–18584. doi:10.1007/s11042-016-4159-7
[9] Lu, W., Li, J., Wang, J., & Qin, L. (2020). A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications. doi:10.1007/s00521-020-05532-z
[10] Ou, P., & Wang, H. (2009). Prediction of Stock Market Index Movement by Ten Data Mining Techniques. Modern Applied Science, 3(12). doi:10.5539/mas.v3n12p28
[11] J. Yang and X. Zhou, "Prediction of PM2.5 Concentration Based on ARMA Model Based on Wavelet Transform," *2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2020, pp. 63-66, doi: 10.1109/IHMSC49165.2020.00022.
[12] Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, 433–441. doi:10.1016/j.jhydrol.2012.11.017
[13] Torres, J. L., García, A., De Blas, M., & De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy, 79(1), 65-77. doi: 10.1016/j.solener. 2004.09.013
[14] Zhang Dongxu, Empirical analysis of the fluctuation of the Shanghai Stock Index return rate based on the ARMA-GARCH models [D]. Tsinghua University, 2016
[15] Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. Energy Economics, 37, 152–166. doi:10.1016/j.eneco.2013.02.006
[16] Lojowska, A., Kurowicka, D., Papaefthymiou, G., & van der Sluis, L. (2010). Advantages of ARMA-GARCH wind speed time series modeling. 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems. doi:10.1109/pmaps.2010.5528979
[17] J. Qin, Z. Tao, S. Huang and G. Gupta, "Stock Price Forecast Based on ARIMA Model and BP Neural Network Model," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 426-430, doi: 10.1109/ICBAIE52039.2021.9389917.
[18] Wei-Min Li, Jian-Wei Liu, Jia-Jin Le, & Xiang-Rong Wang. (2005). The financial time series forecasting based on proposed ARMA-GRNN model. 2005 International Conference on Machine Learning and Cybernetics. doi:10.1109/icmlc.2005.1527274
[19] Jia Chunmiao, Fu Zhongning, Ma Yaling, Li Jianguo, Chen Hulin. Freight Volume Forecast of Gan-Ning District Railway Based on ARMA Model and Multiple Regression [J]. Comprehensive Transportation, 2022,44 (09): 147-154