

Portfolio Construction in China's Stock Market: Take Automobile Industry as Example

Qinyan Xie*

School of Finance, Jiangxi University of Finance and Economics, Nanchang, China

*Corresponding author: 2201904751@stu.jxufe.edu.com

Abstract. With the development of new energy vehicles, investors focus on the automobile subsidy in rural areas, and automobile section of China's A share has been active recently. Based on ARMA model and Mean-Variance model, this paper hopes to give advice on portfolio construction. All the closing price of the stocks are download from AkShare database, from the 9th June, 2020 to the 9th June, 2022. After the stationary test and correlation test for time series, make predictions about return rates for the following 20 days based on ARMA model, construct a profitable portfolio based on Mean-Variance model, compare the accumulative returns and Sharp ratios of constructed portfolio and the one with equal weights. The result reveals that the constructed portfolio outperforms the latter portfolio obviously, so it can be considered that the construction is reasonable. This paper follows the current affairs about the policies on car subsidy in rural areas, uses statistical methods to provide suggestions of wealth allocation for investors, expands the application of the two models.

Keywords: portfolio construction; automobile industry; ARMA model; mean-variance model.

1. Introduction

In 2009, the policies on car subsidy implemented for the first time, which not only improved the consumption structure in rural areas, but stimulated the development of the automobile industry. Because of the implementation, the integral economic development in China has been promoted a lot.

Recent years, people's awareness of environmental protection have been raised, the government kept emphasizing the importance of the notion of green environment. Accordingly, new energy related things are greatly encouraged. So, in automobile industry, the market for new energy vehicles has expanded a lot. However, the outbreak of COVID-19 had hit the whole economic in all countries heavily, people's production capacity and consumption willingness had reduced, and they are more inclined to save money. The macroeconomic growth had seen a downward. In March 2022, the epidemic spread again in many cities in China, which lay a big challenge for China's economic growth. It is urgent to promote the economy and stimulate consumption. However, in automobile industry, the urban automobile consumption market has reached a high level and has changed from an incremental market to a stock market, so it is important to stimulate the rural market to promote consumption and revitalize the market.

Policy implementation is closely related to stock market changes. In April and May this year, many investors have speculated about automobile policies to the countryside. At the beginning of June, it was obvious that the automobile sector in stock market was highly active, and there were many capital inflows. Under this context, this paper hopes to provide a reasonable portfolio construction suggestion by ARMA model and Mean-Variance model.

According to the literature before, Auto-Regressive Moving Average model has been applied to predictions by many scholars. Niu et al. used ARMA model to predict the price of Ya pear in Hebei province, the results show that the model has good prediction results [1]. Sun used ARMA-LSTM combined model to predict the railway passenger flow, and the results reveal the improvements on the accuracy of passenger flow prediction and the accuracy of the model [2]. But ARMA model has some defects in long-term prediction, and the error is relatively large than short-term prediction. The short-term prediction is better than that of long-term prediction [3].

In addition, Risk and expected return are the two core issues that investors are most concerned about, it is important to achieve optimal asset allocation on the basis of balancing the two issues. Mean-Variance model is proposed by Markowitz, who published a book called *Portfolio Selection*

[4], which realize the minimum risk when given the expected return, or the maximum return when given the risk. It is widely applied to construct portfolios. Wang used Mean-Variance model to construct portfolio of occupational pension, and the portfolio is proved to be profitable [5]. Chen used this model to construct portfolio of foreign exchange [6], the result reveals that it is helpful for investors to allocate wealth reasonably among various currency pairs, so as to achieve the purpose of obtaining the highest expected rate of return under certain risk conditions and minimizing investment risks and losses under certain expected rate of return conditions.

The following of this paper use reference to the researches before, selects 6 stocks in the automobile sector in China's A-share market, uses the ARMA model to predict the following 20-day stock yields, uses Mean-Variance model to construct a reasonable portfolio based on the predictions, the results show prediction accuracy and portfolio profitability, which provide suggestion of wealth allocation for investors.

2. Data and methodology

2.1 Data Selection

Select 6 distinctive stocks from the Listed Companies in the automobile sector of China's A-share market. Among them, in Shanghai Stock Exchange, there are 4 chosen stocks. The ticker symbol of the first stock is 600066, which has a high share dividends rate, and the historical stock dividend rate is 62.69%. The ticker symbol of the second stock is 600686, its PE ratio is relatively high in automobile industry, also the procyclical effect is obviously shown on its stock price. The ticker symbol of the third stock is 601127, according to the first-quarter report of the company, its rate of gross profit margin had increased significantly. The ticker symbol of the fourth stock is 601633, the advantage of the company is its abundant free cash flow, and the 2021 annual report shows that the free cash flow is RMB 4.41 billion and the operating cash flow is RMB 35.316 billion. Besides, in Shenzhen Stock Exchange, there are 2 chosen stocks. The ticker symbol of the first stock is 000572, the company has poor cash flow, which is -299 million yuan from operating activities in the first quarter report. The ticker symbol of the second stock is 002594, the number of shareholding institutions decreased significantly, 1244 in 2021 annual report and 405 in 2021 third quarter report. The data of the six stocks are all from AkShare database in python, with the time span from the ninth of June, 2020 to the ninth of June, 2022. Table 1 is the descriptive statistics results of the variables.

Table 1. Descriptive statistics of stock yields

Ticker symbol	600066	600686	601633	601127	000572	002594
Count	482.0000	482.0000	482.0000	482.0000	482.0000	482.0000
Mean	-0.000613	0.000130	0.004098	0.003057	0.002330	0.003375
Std.dev	0.023571	0.029111	0.048247	0.038110	0.036535	0.035442
Min	0.091840	-0.105385	-0.105570	-0.105378	-0.105504	-0.105427
25% quantile	-0.013720	-0.014718	-0.025680	-0.022554	-0.020674	-0.017886
50% quantile	0.000000	0.000000	0.001999	0.001087	0.000000	0.002642
75% quantile	0.010712	0.012249	0.027876	0.024841	0.024893	0.022971
Max	0.095656	0.095962	0.095767	0.095589	0.096442	0.095326

2.2 Methodology

2.2.1 ARMA model

2.2.1.1 Introduction of ARMA model

ARMA is an abbreviation of Auto Regressive Moving Average, proposed by Box and Jenkins in Time Series Analysis Forecasting and Control published in 1978 [7]. ARMA is a random process

composed of autoregressive and moving average. It is usually recorded as ARMA (p, q). The mathematical expression is as follows.

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (1)$$

{ e_t } is a white noise sequence, The Auto-Regressive Moving Average model parts' lag order is represented by p and q, respectively. When p equals 0, the ARMA (p, q) model degenerates into an AR (p) model, and when q equals 0, it degenerates into an MA (q) model. It should be noted that ARMA model can only classify stationary time series. If the time series is non-stationary, it needs to be preprocessed to generate a new stationary time series before it is applied [8]. In practice, logarithmic difference or difference is usually used for preprocessing, and the stable sequence after d times of difference can be written in the form of ARIMA (p, d, q), and the value of d usually does not exceed 2.

2.2.1.2 Determination of lag orders

Autocorrelation graphs and partial autocorrelation graphs can be used to calculate the lag order p and q, if the partial autocorrelation function graph is truncated and the autocorrelation function graph of a stationary series is trailing, the AR (p) model can be established for the series; if the autocorrelation function graph of a stationary series is truncated and the partial autocorrelation function graph is tailed, the MA (q) model can be established for the series; if the autocorrelation function graph and partial autocorrelation function graph of a stationary series are both trailing, thus the ARMA (p, q) model can be established. In addition, the value of AIC and SC should be used to judge the lag order p and q. If the selection of p and q make the value of AIC and SC minimized, then they are the best orders [9].

2.2.1.3 Modeling steps

The ARMA model is used to fit the time series of stock prices. First, match the appropriate values of p and q for each stock. Second, predict the values and the trend of the yield of each stock in the next 20 days.

2.2.2 Mean-Variance Model

2.2.2.1 Introduction of Mean-Variance model

Mean-variance model, also known as Markowitz model, is a portfolio optimization model proposed by Markowitz in 1952. It combined the risk and return of investment, the model put forward a quantitative method to measure financial risk for the first time, which defined risk as the volatility of yield [10]. Also, the model helps people to choose the optimal weights for each asset in the portfolio from all possible weights.

The model is based on the following simplified assumptions. First, investors estimate the risks of portfolios according to the variance, or mostly standard deviation of the predicted return. Second, investors make decisions solely based on the risks and returns of the stocks. Third, investors expect to get the maximum return at a certain risk level, correspondingly, at a certain level of return, investors hope to minimize the risk [11].

Mathematical expression of daily average rate of return and mathematical expression of standard deviation of daily rate of return are as follows.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2)$$

$$E(R) = \frac{\sum_{t=1}^n R_t}{n} \quad (3)$$

$$\sigma(R) = \sqrt{\frac{\sum_{t=1}^n (R_t - \bar{R})^2}{n - 1}} \tag{4}$$

R_t represents the yield of the stock on the day, P_t represents the closing price of the stock on the day, and the opening price of the stock on the day is represented by P_{t-1} . $E(R)$ represents the expected return or daily average return of the stock, $\sigma(R)$ represents the standard deviation of stock return, and n represents the number of dates observed [12].

Calculation of the annualized rate of return and the standard deviation of annualized rate of return are as follows.

$$\text{Annualized return} = E(R) \times 252 \tag{5}$$

$$\text{Annualized standard deviation} = \sigma(R) \times \sqrt{252} \tag{6}$$

2.2.2.2 Modeling steps

First, use the forecast returns of each stock by ARMA model, randomly generate 1000 groups of simulated weights for 1000 portfolios, draw the effective frontier and calculate the Sharpe ratio of each portfolio to find the one with the maximum Sharpe ratio, then construct the portfolio with the optimal weights for each stock.

Second, create a portfolio with equal weights, calculate the accumulated yield of the optimal portfolio with maximum Sharp ratio and the portfolio with equal weights, contrast the performances of two portfolios.

3. Results

3.1 Forecast Results

3.3.1 Data validation

Before prediction of time series, the first step is the stationarity test. Here use ADF test and Ljung-Box test.

ADF test is also known as Unit Root Test, if the unit root exists, the relationship between the independent variable and the dependent variable is deceptive, the process is actually a random walk. This kind of regression is also called spurious regression. There is a unit root is the H_0 hypothesis of ADF test [13]. There are three confidence levels (10%, 5%, 1%), if the significance test statistics are less than the three above, then there should be (90%, 95%, 99%) confidence to reject the original hypothesis.

Ljung-Box test is a statistical test for whether there is lag correlation in time series, which is based on a series of lag orders to judge whether the correlation or randomness of the sequence population exists. The significance test is ditto.

Table 2 shows p values of ADF test and Ljung-Box test for each stock respectively, it can be seen clearly the time sequences of stock price are all stable with the confidence level of 95% in two tests.

Table 2. p values of ADF test and Ljung-Box test

Ticker symbol	ADF test	Ljung-Box test
600066	$2.55e^{-07}$	0.0019
600686	$1.07e^{-25}$	0.0245
601127	0.00	0.0430
601633	$1.09e^{-11}$	0.0298
000572	$4.01e^{-07}$	0.0171

002594	$2.71e^{-25}$	0.0245
--------	---------------	--------

3.1.2 Prediction outcomes

Fig.1 shows the 20 days of predicted returns of each stock by ARMA model. The ticker symbols of the stocks are 600066, 600686, 601127, 601633, 000572, 002594 respectively in order.

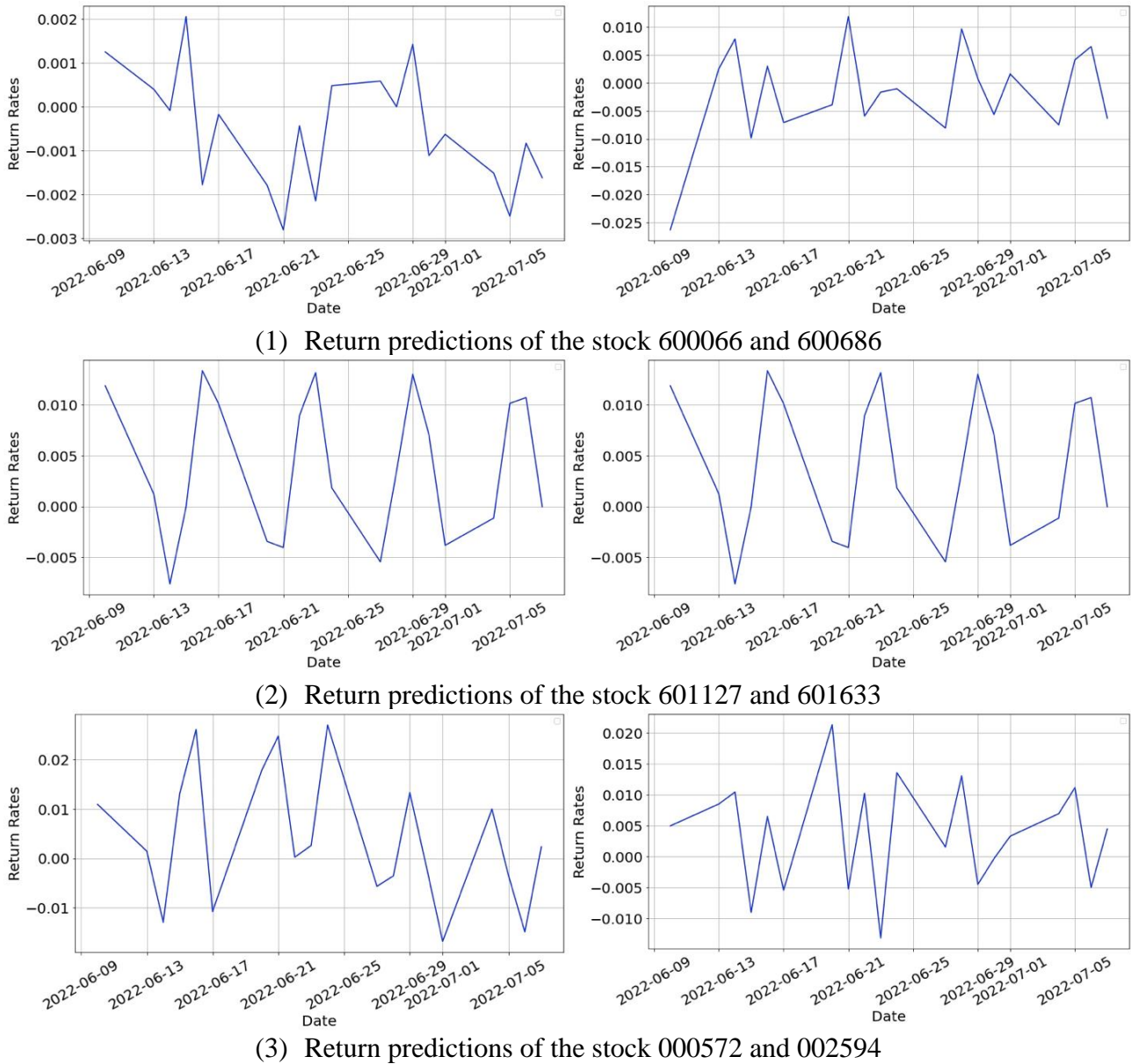


Fig. 1 Return predictions of six stocks

Test the residuals of the prediction results with certain confidence level, the calculate results shows that all the residuals are white noise. It can be considered that the predictions are accurate with the confidence level. Table 3 shows the p values of residual series of each stock.

Table 3. P values of residuals

Ticker symbol	600066	600686	601127	601633	000572	002594
P values	0.2053	0.9221	0.9457	0.8680	0.4661	0.5886

3.2 Portfolio Construction

3.2.1 Data validation

Before portfolio construction, it is important to test correlation among variables. Covariance matrix is a matrix of covariance differences between two or more variables arranged in sequence. It is used to measure whether the variation trend of two variables' deviations is consistent.

Table 4 shows the covariance matrix of the stocks in the portfolio. It can be clearly seen that the covariances between variables are all low, thus the predicted returns can be seen as uncorrelated.

Table 4. Covariance matrix

	600066	600686	601127	601633	000572	002594
600066	0.000460	-0.000019	-0.000019	-0.000019	-0.000286	-0.000280
600686	-0.001194	-0.002779	-0.002779	-0.002998	-0.002998	0.001360
601127	-0.000019	0.012526	0.012526	0.001539	0.001539	-0.004815
601633	-0.000019	0.012526	0.012526	0.001539	0.001539	-0.004815
000572	-0.000286	0.001539	0.001539	0.045808	0.045808	0.003730
002594	-0.000280	-0.004815	-0.004815	0.003730	0.003730	0.019279

3.2.2 Portfolio Construction

Randomly generate 1000 groups of simulated weights for 1000 portfolios, calculate and contrast Sharpe ratio of each portfolio to find the one with the maximum Sharpe ratio.

Fig.2 shows the mean-variance of 1000 random portfolios and emphasize the portfolio the maximum Sharpe ratio. It can be seen that its volatility is 0.0699, return rate is 0.9383.

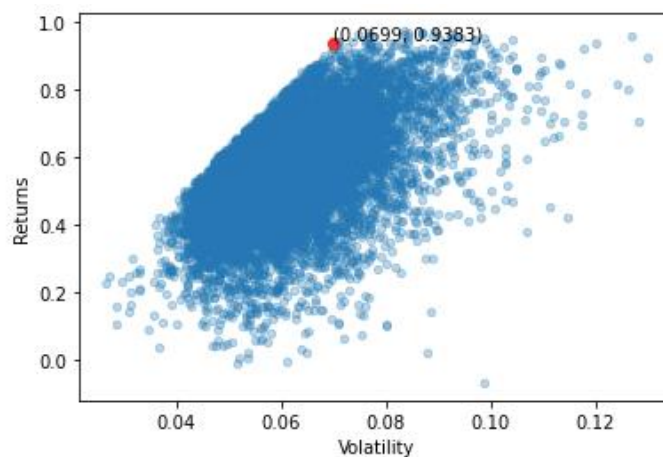


Fig. 2 Mean-variance graph with maximum Sharpe ratio

Calculate the weights for each stock in the portfolio with the maximum Sharpe ratio, the outcomes are 0.0183, 0.0085, 0.2416, 0.3414, 0.0386, 0.3517 respectively for each stock.

3.2.3 Portfolio performance

Create a portfolio with equal weights, calculate the returns of the optimal portfolio and the portfolio with equal weights, which are shown in Table 5. Draw the accumulative returns graphs of the two portfolios, which is shown in Fig.4, contrast their performances. The red line in the plot represents the accumulative returns of the portfolio with maximum Sharp ratio, the blue one represents that of the portfolio with equal weights. It can be seen clearly that the performance of portfolio with maximum Sharpe ratio is better than equal weights one, with higher rate of return.

Table 5. Contrast of two portfolios

	Portfolio with Sharp ratio	Portfolio with equal weights
Annualized rate of return	0.9382	0.5565
Sharp ratio	0.5616	0.4457
Variance of daily returns	0.0226	0.0153

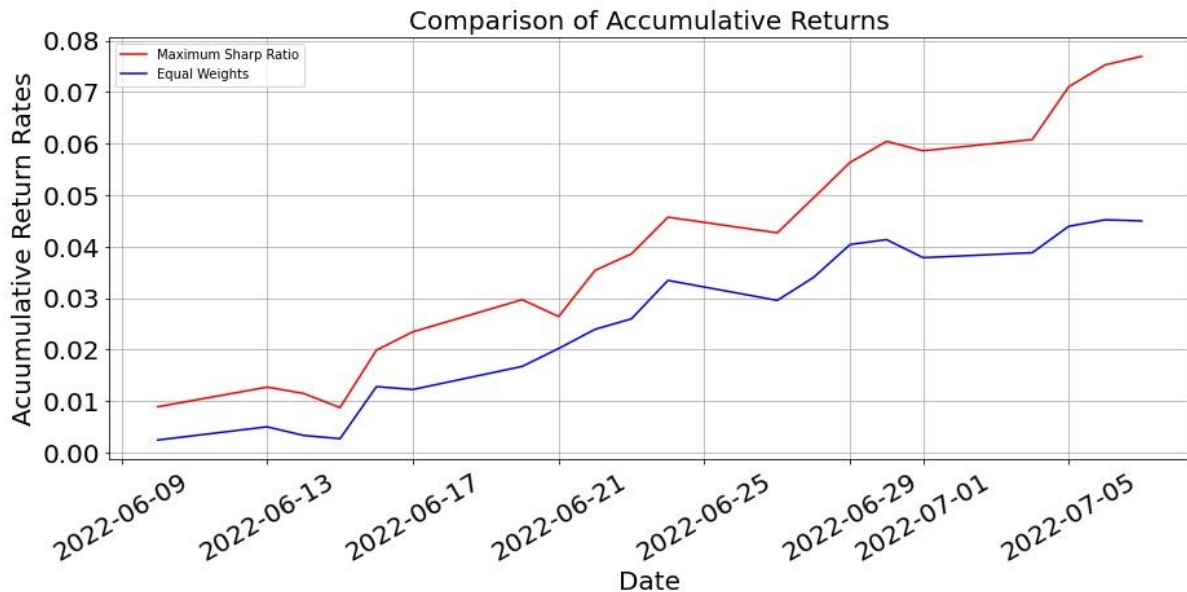


Fig. 4 Accumulative returns

4. Conclusion

In recent years, with the increasing expectation of the automobile subsidy in rural areas, automobile section of China’s A share has been active for a period. Under the context, this paper hopes to give advice on portfolio construction. The data is the closing price of 6 stocks from automobile sector of China's A-share market with the period of 2 years, use ARMA model to make predictions about the following 20-day returns, the result shows the residuals are all white noise, reveals that the predictions are accurate with the confidence level of 95%. Use Mean-Variance model to construct a reasonable portfolio with maximum Sharp ratio, compare the annualized yields and the variance of daily returns of constructed portfolio and the one with equal weights to test the performance of the constructed portfolio. The result reveals that the former one outperforms the latter obviously, so it can be considered that the constructed portfolio is successful.

This paper follows current affairs, uses strict statistical methods, provides suggestion for portfolio construction in the automotive field under the current context, expands the practical application of Auto-Regressive Moving Average model and Mean-Variance model, also provides practical guidance for investors. However, there are still some improvements. This paper only relies on ARMA model for return predictions, which may not be accurate enough. In the future, it is planned to use another model or more models for prediction to improve the accuracy of prediction and the effectiveness of portfolio construction.

References

- [1] NIU Guicao, ZHOU Jihong, MA Hongyan, DU Xiaodong, WANG Junqin. Price prediction and early warning of Hebei Ya pear based on ARMA model. Shandong Agricultural Science, 2021,53 (11): 151-156. (In Chinese)
- [2] SUN Yue, SONG Xiaoyu, JIN Liting, LIU Tong. Railway passenger flow forecast based on ARMA-LSTM combined model. Computer Applications and Software, 2021,38 (12): 262-267+273. (In Chinese)

- [3] LIU Jie. Application of ARMA model in stock price prediction -- Taking Gree Electric Appliance as Example. *China Management Informationization*, 2021, 24(11):153-155. (In Chinese)
- [4] Markowitz H.M. Portfolio selection. *The Journal of Finance*, 1952, 7(1):77-91.
- [5] WANG Zhongdong. Research on portfolio selection of occupational pension of the People's Bank of China--Based on Markowitz mean-variance model. *Financial Economy*, 2020 (06): 80-86. (In Chinese)
- [6] CHEN Min. Analysis of foreign exchange portfolio optimization strategy based on mean-variance model. Chongqing: Chongqing University, 2019. (In Chinese)
- [7] Box G.P.E, Jenkins G.M. *Time series analysis: Forecasting and control*. San Francisco: San Francisco Press, 1978.
- [8] HU Kang, GE Ruiyang. Price prediction of CSI 300 index based on ARMA model. *Economic Management Abstracts*, 2021 (18): 169-171. (In Chinese)
- [9] WANG Xiaopeng, CAO Guangchao, DING Shengxi. Modeling and prediction of precipitation time series analysis in the southern Qinghai plateau based on Box-Jenkins method. *Journal of Applied Statistics and Management*, 2008 (04): 565-570. (In Chinese)
- [10] HE Chaolin, XU Qian. Portfolio choice under the mean-variance model with parameter uncertainty. *Journal of Donghua University (English Edition)*, 2015, 32(03): 498-503.
- [11] ZHU Shihao, SHI Jingtao. Optimal reinsurance and investment strategies under mean-variance criteria: Partial and full information. *Journal of Systems Science & Complexity*, 2022:1-22 (In Chinese)
- [12] ZUO La. Research on portfolio strategy based on mean variance model. *Financial Theory Research*, 2021 (05): 9-16. (In Chinese)
- [13] MA Zixuan. Application of ARMA model in stock short-term prediction. *Guangxi Quality Supervision Guide*, 2020 (05): 152-153. (In Chinese)