

Asset Forecasting Analysis Based on ARIMA Model and BP Neural Network

Hanyin Yang

Department of business administration, Zhongnan University of Economics and Law, Wuhan, Hubei, 430073

vj162774@163.com

Abstract. This paper forecasts the trend of the asset based on historical prices. First, through the establishment of exponential smoothing method, *ARIMA*, BP neural network and other models, the trader invest in three assets: gold, bitcoin and cash in USB. The thesis is based on historical price to predict the trend of assets, determine whether traders should purchase, hold or sell and what percentage of the asset, and evaluate its future value. This paper first predicts the future returns and volatility of the two assets. In the initial forecast, the exponential smoothing method and *ARIMA* model are used to predict premiums and future returns. BP neural network is used to predict the future earnings in the middle and late forecast. First, for the first 60 days, we sit tight and wait for the data to accumulate. After 60 days, by looking back at historical data and setting appropriate technical indicators, the secondary trend curve and risk exposure curve of gold and Bitcoin can be obtained respectively. Once we have the curves, the commission, expected rate of return, and volatility of the two markets are combined, we will set up a scoring system to score the daily trading feasibility. Finally, we simulate the transaction, allocate the investment share, get the asset accumulation curve, and complete the decision.

Keywords: Gold; Bitcoin; *ARIMA*; Neural Network.

1. Introduction

To maximize the total return, market traders usually consider commissions and many other factors to decide the best portfolio strategy. In this problem, the initial capital is \$1000, and the tradable assets are gold and bitcoin. Based on the unknown of the future price trends, we are required to construct models and develop trading strategies to buy or sell both assets to maximize the total return.

By observing the trend charts of gold and bitcoin, we found that the daily capital flows of both are in fluctuations. It can be seen from the figure that for gold, the early, middle and late stages are in a slow-changing trend. In contrast, bitcoin is relatively flat in the early and middle stages, while the price changes sharply in the late stage. Because BP neural network needs more historical data, for the early stage, we use the *ARIMA* model to make the initial prediction. Under the condition of sufficient historical data, the neural network is used for prediction in the middle and late stage, and the trading strategy is determined based on the factors that are closely related to the prediction judgment and price fluctuation.

2. Model assumptions and notation

Combined with the actual problem, in order to ensure the accuracy and rationality of the model solution, this paper eliminates the interference of some factors and makes the following assumptions for the model.

- Bitcoin and gold split indefinitely.
- There is no arbitrage in the market.
- Investors pursue the highest return with the lowest risk.
- Money has no time value.
- Non-systemic risks are not considered.

The symbols and their meanings are as following Table 1.

Table 1. The symbols and their meanings

R_t	Expected return
σ_t	Risk on the current day
n_t	Residuals

3. Model construction and solving

3.1 Date Processing

We first process the given data, since gold is not traded on non-working days, we adopt the price of the previous trading day as the gold price of the day, then establish a new indicator to determine whether gold trading can be carried out on that day. In addition, we calculate the 5-day and 30-day averages of bitcoin prices, the 10-day and 60-day averages of gold prices, and the residuals of the two assets.

3.2 Prediction Model

3.2.1 Prediction Model of early stage

(1) Exponential Smoothing Method (ES)

Exponential smoothing method [1] is characterized by giving different weights to the past observations. The weights of recent observations are larger than those of long-term observations. The predicted values are the weighted sum of previous observations, which is consistent with the prediction based on historical data, mentioned above. In general, the first-order exponential smoothing method can be expressed as.

$$S_t = a \cdot y_t + (1 - a) \cdot S_{t-1} \tag{1}$$

(2) Moving Average Model (MA)

Moving average model [2] sets up a discrete linear system. $u(n)$, the input variate, is a zero-mean white noise sequence variances variance, and the output variate is $x(n)$. The relationship between the output variate and the input variate of this discrete linear system can be expressed by the following difference equation.

$$x(n) = \sum_{r=0}^M b_r u(n-r) \tag{2}$$

Its system function is.

$$H(Z) = \frac{X(Z)}{U(Z)} = \sum_{r=0}^M b_r Z^{-r} \tag{3}$$

In the above formula, $X(Z)$ is the Z-transformation of the output signal $x(n)$, $U(Z)$ is the Z-transformation of the input signal $u(n)$, and $b_r (r = 0, \dots, M)$ is the coefficient.

(3) Autoregressive Integrated Moving Average model

ARIMA (p, d, q) model [3] is the extension of model *ARMA*(p, q). *ARIMA*(p, d, q) model can be expressed as.

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \tag{4}$$

Through the thesis, SPSS software and $ARIMA(p, d, q)$ model are used to conduct a stepwise analysis, using the known historical data of gold and bitcoin. The forecast value is then calculated, and the subsequent forecast price is more precisely adjusted based on the forecast premium. The model can be expressed as.

$$PRemium_t = (\exp P_t - P_t) / P_t \tag{5}$$

(4) Initial Stage Empirical Prediction

The *Ljung – Box test* statistical test [4] is used to perform on whether there is a lag in the time series. This test is mainly used to examine whether a series of observations in a certain period is random independent, and exclude the autocorrelation between observations. The result can avoid autocorrelations that reduce the accuracy of time-based forecasting models (such as transition diagrams) and lead to the misinterpretation of the data.

The result of *Ljung – Box test* is $p - \text{value} < 2.2e - 16$, which is a significant non-white noise sequence at the 99% level, so the time sequence is meaningful to establish [5]. The analysis recursion is as follows.

$$\text{Premium } P_t = \alpha + \sum_{m=1}^t \beta_m \text{ Premium } P_{t-m} + \sum_{n=1}^t \beta_n \text{ Premium}_{t-n}^2 \tag{6}$$

After adjusting the parameters, the value of *AIC* is the smallest under this parameter configuration, so this parameter model is selected.

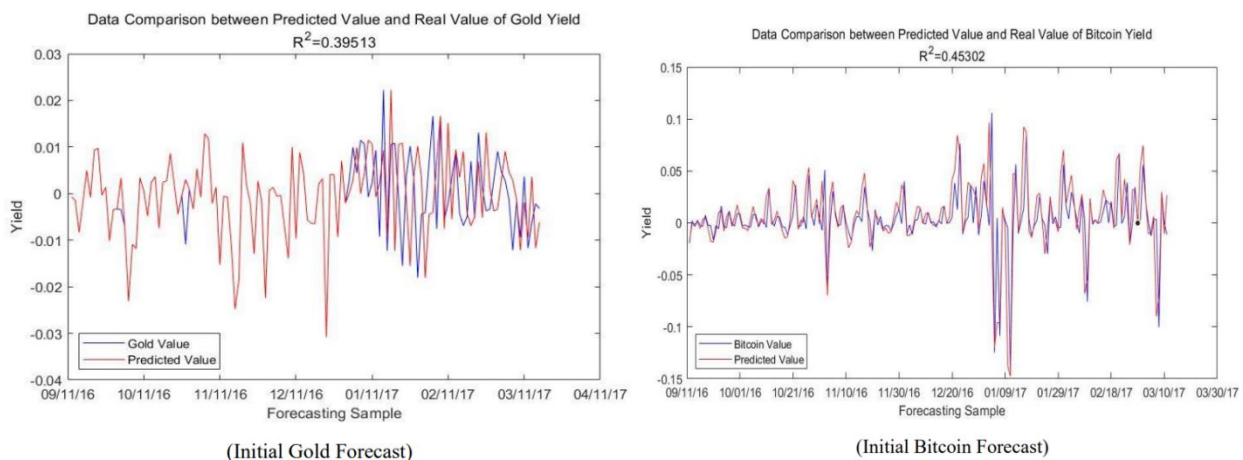


Fig. 1 Gold and bitcoin initial price predictions compared to the real value

Figure 1 shows the *ARIMA* model-based on gold and bitcoin based on historical data. The forecast series from September 11, 2016 to April 11, 2017 is selected as the initial forecast. However, due to the uncertainty of the future and the lack of the previous data, the prediction of the *ARIMA* model is not very accurate, so we consider replacing the model during the mid-to-late forecast.

However, BP neural network needs more historical data, so when BP neural network data is insufficient in the early stage, *ARIMA* model is still used.

3.2.2 Prediction Model of Mid-Late stage

(1) Back Propagation Neural Network

The calculation process of the backpropagation neural network [6] is made up of three segments, a positive-going calculation process and a reverse calculation process. In forward propagation, the operation mechanism is as follows: from the input layer through the hidden layer, then processing to the output layer. The state of each layer of neurons only affects the state of the next layer of neurons.

If the desired output data cannot be reached at the output layer, reverse propagation will be carried out to return the error signal along the connection path, and the error signal will be minimized by modifying the parameters of each neuron.

BP neural network has strong nonlinear mapping ability and flexible network structure, which is suitable for financial data forecast in the case of a large number of historical data [7].

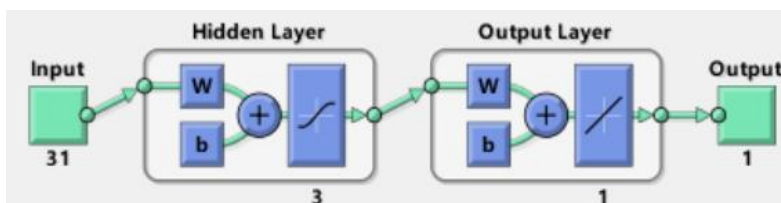


Fig. 2 Structure diagram of 3-layer BP neural network

(2) Mid-late Empirical Prediction

This paper uses Matlab software and selects the BP neural network model to analyze the historical data of gold and bitcoin. Backpropagation neural network is a multi-layer feedforward network trained by an error backpropagation algorithm. Its topology includes the input layer, hidden layer and output layer [8]. Weights and thresholds of the network are constantly adjusted through Back Propagation to minimize the sum of error squares of the network.

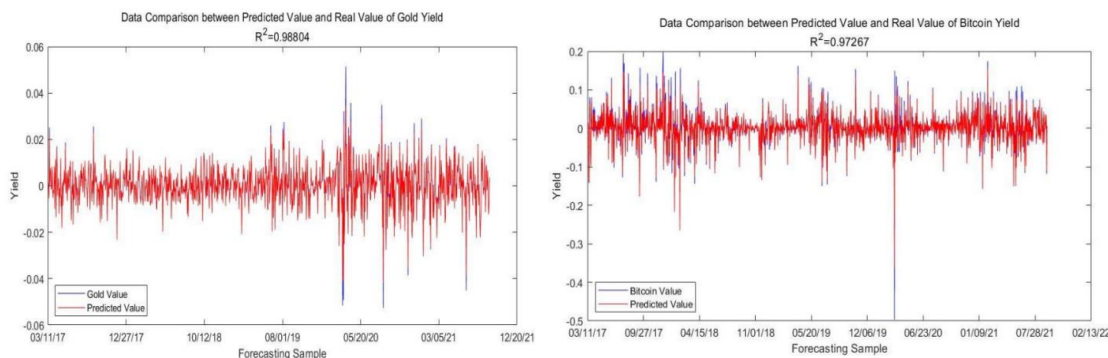


Fig. 3 Gold and bitcoin mid-late stage price forecast compared to true value

Figure 3 shows a comparison of the BP neural network's predictions with real data, based on historical data for gold and bitcoin. The data from March 11, 2017 were selected for mid to late prediction. According to the prediction graph, it can be seen that the predicted value of the model is close to the actual value, which preliminarily shows that the prediction of the neural network is good.

Through calculation, the R^2 value of gold is 0.98804 and the R^2 value of bitcoin is 0.97267, which indicates that the model has a high degree of fit. This shows that the BP neural network model can predict the future trend of gold and bitcoin well based on the obtained historical data.

3.3 Transaction Model

3.3.1 Trend Grading System

After obtaining the forecast data, we formulate a trading model according to the return and fluctuation characteristics of the data. Due to insufficient understanding of the market in the first 60 days, the 60-day moving average is usually associated with a secondary trend, so the volatility is large. Therefore, for the first 60 days, we chose not to trade, but to make predictions. After 60 days, we made an analysis based on the past daily increasing trend, and set 10 days as a threshold value [9].

As long as the asset's rising and falling direction changes or continues to depreciate within 10 days, it is considered that these 10 days follow the downward trend. On the contrary, if the asset continues to rise within 10 days, it is considered that these 10 days obey the upward trend. After that, the 10-day bias of the first 60 days can be obtained by calculation. So we designed a scoring system to

evaluate whether the market is in an ascent stage or a declining stage. The initial values of the system are formulated as follows.

$$\text{score} = \frac{2}{3} \times (\text{Average yield of earlier 60 days}) + \frac{1}{3} \times (10 - \text{day bias of earlier 60 days}) \quad (7)$$

The above formula is designed because we usually believe that the ascent and decline direction of assets tomorrow is usually the same as today, that is, if it rises today, it is very likely that it will continue to rise the next day. Therefore, we set a larger weight on the average return rate of the past assets to obtain the above formula and the initial value. The initial value is shown in Figure 4.

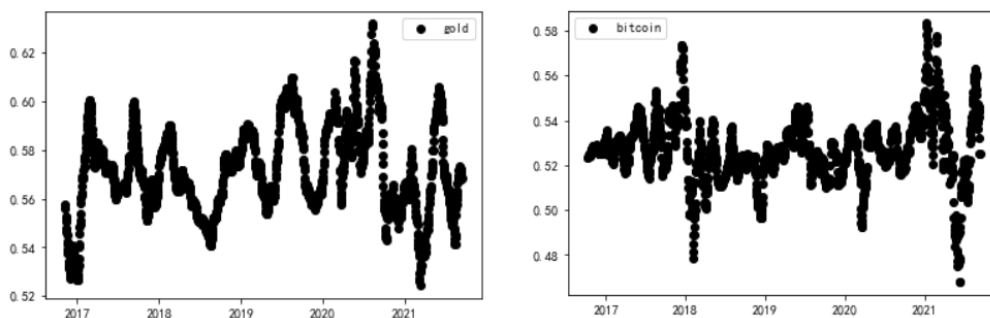
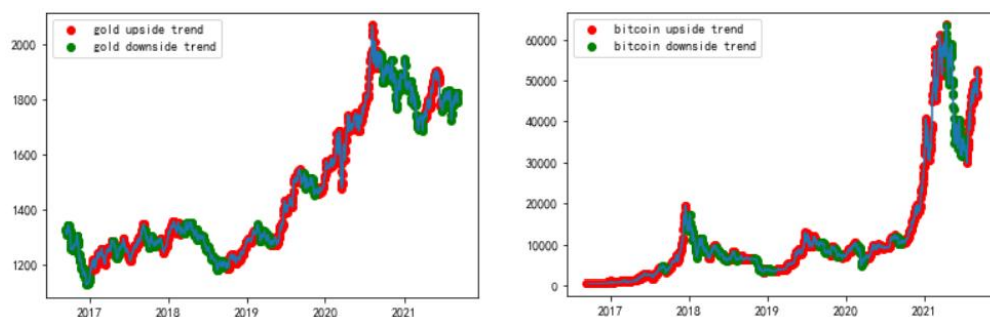


Fig. 4 Initial value

After obtaining the initial value, the distribution curve of its rise and fall can be exported. However, there is a problem that cannot be ignored in the previous mechanism, that is, if it is judged that there is a rising trend on the day, it requires that there is an upward trend all the 10 days before.

According to the mechanism, if there is a downward trend only on the previous day of the first 10 days, the current day will be judged to have a downward trend, which will cause errors in the calculation results of the day. To eliminate the error, we define the initial value as 0. If the calculation result of the day shows that there is an upward trend, the value of the previous 60 days will be increased by 1, and if there is a downward trend, it will be reduced by 1.

Through the above operation, it can be seen whether it is going up or down. The result is shown in Figure 5.



(Fluctuation Distribution Chart of Gold)

(Fluctuation Distribution Chart of Bitcoin)

Fig. 5 Gold and Bitcoin volatility distribution chart

3.3.2 Quantification of Risk Exposure

During the trading, in addition to the yield rate of return will affect the final decision, the risk is also a factor that cannot be ignored. Based on this, in order to quantify the risk exposure during the trading [10], we define it exists a positive correlation between purchase risk, upward trend and 10-day convergence divergence, then give the weights of these two variables, 2 3 and 1 3 respectively. We obtained the purchase risk curves of gold and bitcoin, and quantified the risk exposure. The buying risk chart is shown in Figure 6.

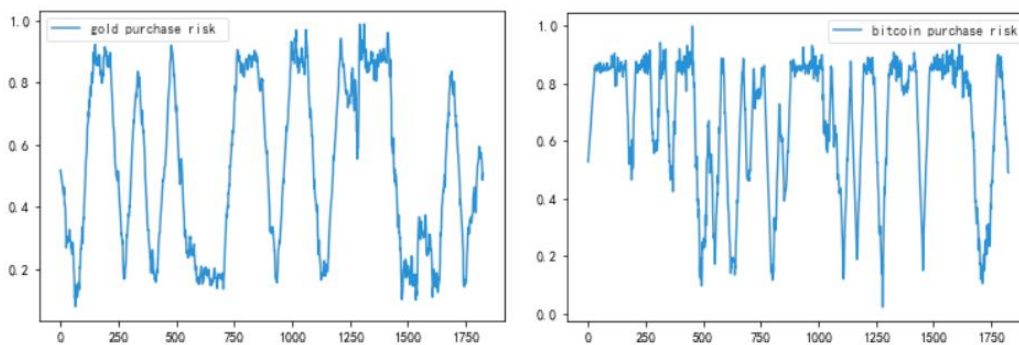


Fig. 6 Buying risk chart

The value of the ordinate represents the size of the risk exposure, the larger the value is, the higher the purchase risk is. Therefore, it is not a wise investment strategy to buy or add positions at this time.

The risk exposure curve of bitcoin changes more frequently, and there are more values close to 1, which indicates that the investment risk of bitcoin is higher than gold. However, the risk premium of bitcoin also brings more expected returns for it, making up for the excess risks that investors need to bear.

3.4 Portfolio Strategy

After excluding 5% of abnormal data, we normalize our score to facilitate our subsequent scoring [11]. In particular, abnormal data was recorded as ± 1 . We then establish a basic threshold based on the price forecast in the previous period, bitcoin's performance in the first 30 days, gold's performance in the first 60 days, and use this threshold to guide our investment.

Our purchase amount is set as follows.

$$\begin{aligned} \text{sell shares} &= \text{holdings shares} * (1 - \text{score} + \text{threshold of sell}) \\ \text{purchase shares} &= [\text{cash} * \text{score} (1 - \text{premium})] / \text{price} \end{aligned} \tag{8}$$

To sum up, the overall flow chart of the strategy is shown in Figure 7.

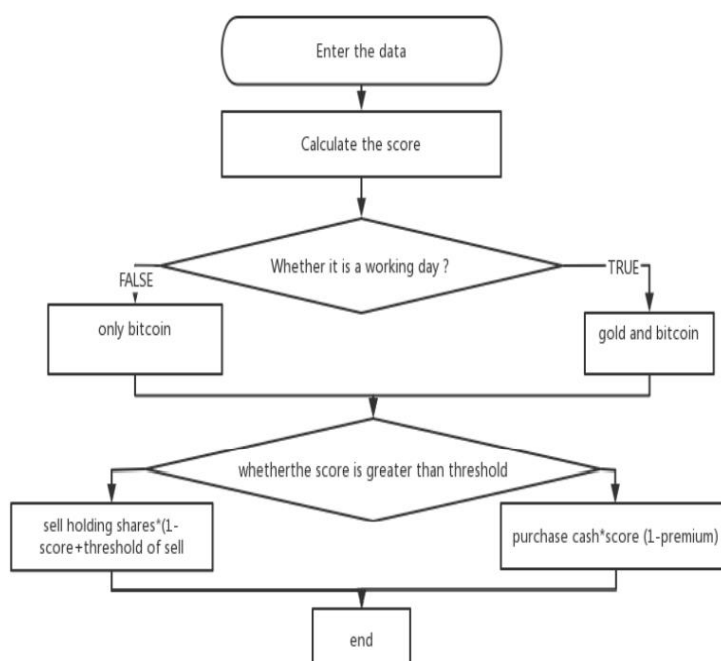


Fig. 7 The overall flow chart of the strategy

According to our strategy, as of September 10, 2021, the total accumulated asset value is \$318,499.0744. The asset chart is shown in Figure 8.

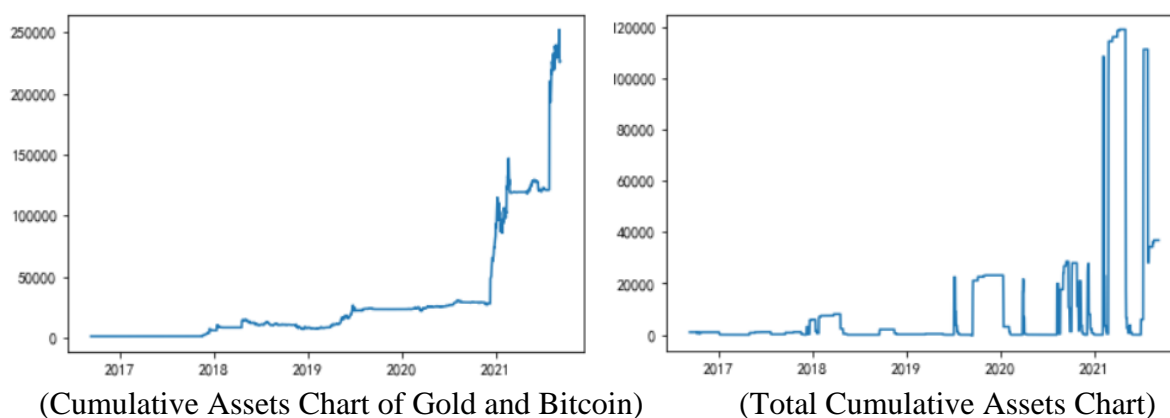


Fig. 8 The asset chart

4. Conclusion

By observing the trend charts of gold and bitcoin, we found that the daily capital flows of both are in fluctuations. It can be seen from the figure that for gold, the early, middle and late stages are in a slow-changing trend. In contrast, bitcoin is relatively flat in the early and middle stages, while the price changes sharply in the late stage. Because BP neural network needs more historical data, for the early stage, we use the ARIMA model to make the initial prediction. Under the condition of sufficient historical data, the neural network is used for prediction in the middle and late stage, and the trading strategy is determined based on the factors that are closely related to the prediction judgment and price fluctuation.

References

- [1] Ansari Saleh Ahmar, F. Fitmayanti & R. Ruliana. (2021). Modeling of inflation cases in South Sulawesi Province using single exponential smoothing and double exponential smoothing methods. *Quality & Quantity* (prepublish), doi:10.1007/S11135-021-01132-8.
- [2] Toledo Nhoriel I.. (2021). The Autoregressive Integrated Moving Average Model in Forecasting Philippine Peso - United States Dollar Exchange Rates. *Journal of Physics: Conference Series* (1).
- [3] Alabdulrazzaq Haneen, Alenezi Mohammed N., Rawajfih Yasmeen, AlGhannam Bareeq A., Al Hassan Aber A. & Al Anzi Fawaz S.. (2021). On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics* (prepublish).
- [4] Chen Y. Ting, (2002). On the Robustness of Ljung– Box and McLeod– Li Q Tests: A Simulation Study." [J]. *Economics Bulletin*, 2002, 3(17): 1-10.
- [5] Chen Tengfei, Chen Guoping, Chen Weiting, Hou Shuo, Zheng Yuxuan & He Huan. (2021). Application of decoupled ARMA model to modal identification of linear time-varying system based on the ICA and assumption of "short-time linearly varying". *Journal of Sound and Vibration* (prepublish).
- [6] Xinhong Ni & Zeng Yang. (2021). Inverter fault diagnosis based on BP neural network. (eds.) *Proceedings of 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE)*(pp.203-208)
- [7] Taehyun No & Taewook Lee. (2020). Wild bootstrap Ljung-Box test for residual autocorrelation in vector autoregressive models. *Journal of the Korean Data and Information Science Society* (3), doi:10.7465/jkdi.2020.31.3.477.
- [8] Zhang, L., Rong, X., & Du, Z. (2014). Precommitted investment strategy versus timeconsistent investment strategy for a dual risk model. *Discrete Dynamics in Nature and Society*, 2014

- [9] Regression_based estimation of dynamic asset pricing models [J]. Tobias Adrian, Richard K. Crump, Emanuel Moench. *Journal of Financial Economics*. 2015 (2)
- [10] Harvey C R. The risk exposure of emerging equity markets [J]. *The world bank economic review*, 1995, 9(1): 19-50.
- [11] Ren F, Lu Y N, Li S P, et al. Dynamic portfolio strategy using clustering approach [J]. *PloS one*, 2017, 12(1): e0169299.