Academic Journal of Business & Management
ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

Optimization of Portfolio Investment Strategy Based on Markov Decision Process Model

Xiangzhao Hao, Runxing Jia

Tianjin university, Tianjin, 300192, China

Abstract: *With the continuous development of the world financial market, investors are often faced with a wide variety of assets but do not know how to choose. In order to obtain the maximum benefits, how to choose the optimal investment strategy has become a top priority. In this paper, based on Markov decision process model, we study the optimal strategy for investment with two important assets in the market, gold and Bitcoin. we solve the problem with a Markov decision process model and measure the risk of the trade by financial indicators such as deviation rates and bullish indicators. Then a difference-in-difference formula for the asset appreciation rate is introduced to optimize the model by incorporating the long-term impact of trading on assets. Also based on the iterative algorithm of the sample path, an iterative algorithm of the strategy that can be applied online is designed, and it was finally calculated that after five years of investment using this strategy under the initial condition of $1000, $348287.46 could be obtained.*

Keywords: *Markov decision process model; Portfolio investment strategy; Proportional transaction costs*

1. Introduction

With the development of world's financial investment market business and the continuous reform of the capital market in recent years, the needs of investors are increasing, and asset trading strategies are becoming even more abundant. From the trading market data, high returns and high risks are often proportional. When faced with a variety of risky assets, every investor will face such a question, how to choose the best return? This is a portfolio strategy problem.

Portfolio investment strategies vary somewhat for different types of assets, where assets with proportional fees are common. Gold and Bitcoin are among these assets. Bitcoin, as an encrypted payment method using blockchain technology, suffers from drastic price fluctuations, high volatility and high downside risks. Gold reflects the stability of the financial market and is the wind vane of market changes.

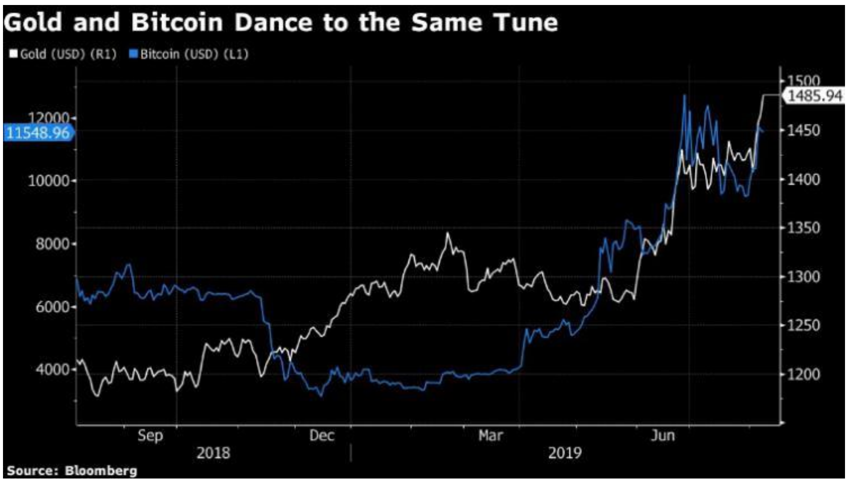

Figure 1: Gold and Bitcoin dance

We will find the best trading strategy so that traders can get the maximum return in these five years and prove that the strategy is optimal.

ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

2. Data Preprocessing

For preprocessing the initial data, first plot the value/days curve for gold and Bitcoin respectively:

Figure 1: Bircoin trend

Figure 2: Gold trend

Then, consider the transaction risk defined as follows:

$$R = Y + B$$

Among them, Y is the deviation rate of the asset on the day, and B is the bull market evaluation index. The average value of the bull market evaluation index is greater than the bull market, and the average value is less than the bear market. The deviation rate is given by formula $\mathrm{Y} = \frac{\mathrm{X} - \overline{\mathrm{X}}_{\mathrm{n}}}{\overline{\mathrm{X}}_{\mathrm{n}}}$, X represents the value of the day, $\overline{\mathrm{X}}_{\mathrm{n}}$ is the average price of the previous n days, and B satisfies the following formula:

$$B = \alpha \frac{\sum_{j=i-N}^{i} R_j}{N} + \beta \frac{\sum_{j=i-N}^{i} Y_j}{N}$$

α and β are weights, and α+β=1, the setting is more subjective, and R is the increase of the asset on the day, which is obtained by comparing it with the previous day. For investment in gold, the mean value of the bull market evaluation index is: 0.57, N is selected for 90 days, and n is selected for 15 days. For investment in Bitcoin, the mean value of its bull market evaluation index is: 0.52, N is selected for 30 days, and n is selected as 5 days, for both assets, the weights α and β are both chosen to be 0.5.

Calculate the initial data to obtain the purchase risk map of gold and Bitcoin:

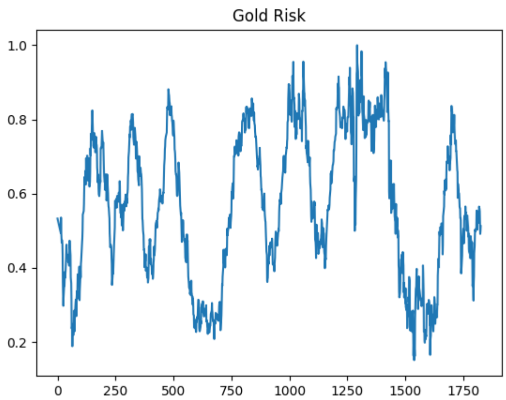

Figure 3: Gold risk curve

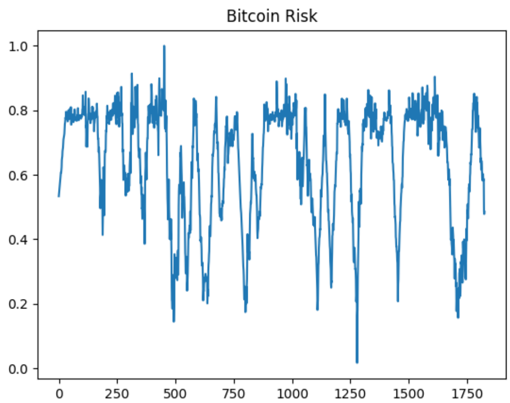

Figure 4: Bitcoin risk curve

3. Model Establishment

For risky assets like Bitcoin, considering the basic geometric Brownian motion model, the price change of the i-th risky asset obeys:

$$p_i(t+1)/p_i(t) = e^{r_i + \sigma_i \omega_{i,t}} \quad (1)$$

ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

Where r_i is the appreciation rate, σ_i is the volatility, $\omega_{i,t}$ is a random number obeying the standard normal distribution.

At each moment, traders will decide how to allocate the ratio of risk-free assets to risky assets in the portfolio according to the established investment strategy. Let $U_i(t)$ denote the sale of risk-free assets at time *t* to buy the i-th risky asset, and $V_i(t)$ denote that at time *t* the amount of funds flowing from the i-th risky asset to the risk-free asset, The fee ratios of the risk assets are denoted as $\lambda_i \in [0,1]$ and $\mu_i \in [0,1]$ respectively. According to the above description, the dynamic changes of the asset portfolio satisfy the following equation:

$$S_0(t+1) = e^{r_0}[S_0(t) - \sum_{i=1}^{N}(1+\lambda_i)U_i(t) + \sum_{i=1}^{N}(1-\mu_i)V_i(t)] \tag{2}$$

$$S_i(t+1) = e^{r_i+\sigma_i\omega_{i,t}}[S_i(t) + U_i(t) - V_i(t)] \tag{3}$$

In this problem, we regard cash as a risk-free asset S_0, so we know that under the constraint of ontology, the appreciation rate of cash is 1 (that is, it does not increase in value), so $r_0 = 0$,The above formula can also be rewritten as:

$$S_0(t+1) = S_0(t) - \sum_{i=1}^{N}(1+\lambda_i)U_i(t) + \sum_{i=1}^{N}(1-\mu_i)V_i(t) \tag{4}$$

Considering Gold and Bitcoin as two risky assets, the core of this problem is the combination optimization problem of one risk-free asset and N risky assets. Generally speaking, the handling fee for purchasing or throwing each risky asset the proportions vary. Depending on the meaning of the question, for the same risky asset in this question, the proportion of the handling fee for throwing or buying is the same. It can be observed that gold is 1% and Bitcoin is 2%.

At time *t*, the total value of the portfolio is recorded as:

$$M(t) = S_0(t) + \sum_{i=1}^{N} S_i(t) \tag{5}$$

The trader's goal is to achieve a long-term average increase in the total value of the portfolio. The value rate *η* is the largest, that is:

$$\eta = \lim_{T\to\infty}\frac{1}{T}E[ln\, M\,(T)] \tag{6}$$

The aim of this paper is to maximize the long-term average value-added rate η by calculation. Therefore, $g^d(x)$ (potential energy) is introduced in this paper to represent the contribution of the previous strategy to the long-term value-added rate, and f^d is introduced to represent the single-stage value-added rate of assets. This translates η into a combination of $g^d(x)$ and f^d. Considering that the combined problem is solved in a finite state space, **markov decision model** is adopted in this paper, and P^d is regarded as the state transition function, so the combined problem can be transformed into $f^d + P^d g^d$.

After introducing the difference model, through the performance sensitivity analysis (**See the second task for the proof process**), it is proved that the optimal strategy can be obtained by using the following strategy iteration formula

$$f^{\hat{d}} + P^{\hat{d}} g^{\hat{d}} = \underset{d}{argmax}\{f^d + P^d g^{\hat{d}}\} \tag{7}$$

It should be noted that the variables in $g^d(x)$, $f^d(x)$, $P^d(x)$have different meanings. The meaning of the former is the proportion of each asset in the initial investment, and the latter represents the sample path.

In formula (7), f^d is the difference of one-stage asset value-added rate between the two strategies in the current state, $P^d g^{\hat{d}}$ is the long-term impact of the current transaction on the long-term average value-added rate of total assets in the future. If we know what these two parts are, we can get the best solution from formula (8).

When we calculate the f^d, we do not directly use parameters such as r_i and σ_i, but estimate the single-stage average value-added rate through the geometric Brownian motion model on the strategy path, which can reduce the number of parameters due to errors caused by the choice.

$$f^d(x) = \frac{\sum_{t=0}^{T-1}[I(x|X(t))f^d(X(t),p(t),p(t+1))]}{\sum_{t=0}^{T-1} I(x|X(t))} \tag{8}$$

Where: $I(x|X(t))$ is $f^d(X(t),p(t),p(t+1))$ defined as:

$$f^d(X(t), p(t), p(t+1)) = ln\left[\frac{p_0(t+1)}{p_0(t)}(s_0(t) - \sum_{i=1}^{N}(1+\lambda_i)u_i(t) + \sum_{i=1}^{N}(1-\mu_i)v_i(t)) + \sum_{i=1}^{N}\frac{p_0(t+1)}{p_0(t)}(s_i(t) + u_i(t) - v_i(t))\right] \quad (9)$$

In this formula, the indicative function can be defined as

$$I(x|X(t)) = \begin{cases} 1, & X(t) \in x \\ 0, & X(t) \notin x \end{cases} \quad (10)$$

Through the above calculation process, we have realized the mapping from the strategy path to the single-stage average value-added rate. Next, we only need to measure the contribution of the initial investment asset ratio to the long-term average value-added rate.

Because the asset portfolio problem is a continuous state space problem, only First estimate the performance potential of a finite number of states, we first construct a state space $S_M = \{X_1, X_2, \ldots, X_M\}$ of a finite number of states, which represents the state space of M states, we first need to measure the contribution of a single state to the long-term average contribution rate, where *X(t)* represents the The proportion of the total asset value of each asset value in the state, as follows:

$$g^d(X_m) \approx \frac{\sum_{t=0}^{T-L+1}[\rho(X_m, X_m(t))\sum_{t=0}^{L-1} f^d(X(t+l)]}{\sum_{t=0}^{T-L+1}\rho(X_m, X_m(t))} \quad (11)$$

Among them, $\rho(X_m, X(t))$ is the weight coefficient. It can be viewed on the formula that we need to select L discrete states, and then calculate the contribution of the subset X (m) of S to the average value-added rate. We choose a window of size L, calculate the average value-added rate of a subset of the state space in this window, and use this value as the contribution value of the subset X(m) to the long-term average value-added rate.

For the strategy path, we need to look at the T-1 strategy from a global perspective, and the impact of each strategy on the average value-added rate, we need a value to measure the average impact of the strategy path on the growth rate, so we only need to the contribution rate of the single-step strategy is weighted to measure the impact of the strategy path on the long-term average value-added rate. The formula is as following:

$$g^d(x) \approx \sum_{m=1}^{M}\kappa(X_m, x)g^d(X_m), x \in S - S_M \quad (12)$$

Among them, the $\kappa(X_m, x)$ function is the weight function, which represents the average contribution of the limited strategies to the growth rate. To satisfy $\sum_{m=1}^{M}\kappa(X_m, x) = 1$

In the decision-making process, we need to determine the probability of making a decision in the strategy path, that is, the transition function, the state transition The function means the probability of making the strategy in the strategy path, so for all points on the strategy path (a finite number of states (it can be assumed that the decision state is discrete and finite)), finally, we have to measure the previous The contribution of the one-step strategy to the average growth rate of the portfolio transaction. This calculation method and the direct use of the growth rate (changed to the symbol) and volatility (unchanged) avoid calculation errors, because this strategy measures the impact of each step of the strategy on the future asset growth rate from the entire state space, rather than directly using the derived parameters to estimate.

Finally, estimates $P^d g^{\hat{d}}$ are needed. Because of the non-linear nature of the portfolio problem, it is not easy to estimate the probability of state transition and calculate $P^d g^{\hat{d}}$, but it is possible to estimate them directly as a whole. Using $g^d(X(t+1))$ as a replacement, get $P^d g^{\hat{d}}$'s estimate

$$P^d g^d(x) \approx \frac{\sum_{t=0}^{T-1}[I(x|X(t))g^d(t+1)]}{\sum_{t=0}^{T-1}I(x|X(t))} \quad (13)$$

At this point, the calculation methods have been given for the quantity we need to calculate, modeling complete.

4. Transaction strategy

Based on the model established above, we derive the following iteration methods to generate optimal strategies:

Step 1: Initialize, select the parameters in the algorithm, including T, M, etc., take the proportion of the initial assets to the total assets as the investment strategy, denoted as d_0, and set k=0, and choose a

ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

next investment strategy d1;

Step 2: Evaluate the strategy, conduct transactions by dk, obtain the sample path, and know the historical price change space. We use the previous formula (write the number) to estimate the various parameters in the key expression;

Step 3: Strategy improvement, construct the next strategy d(k+1) based on the previous strategy, until d(k+1) achieves the maximum value of the key expression and the algorithm stops when d(k+1) <= d(k), the optimal investment strategy is d(k), otherwise let k=k+1, and return to step2.

Based on the above strategy iteration method, the program flow diagram is obtained as follows

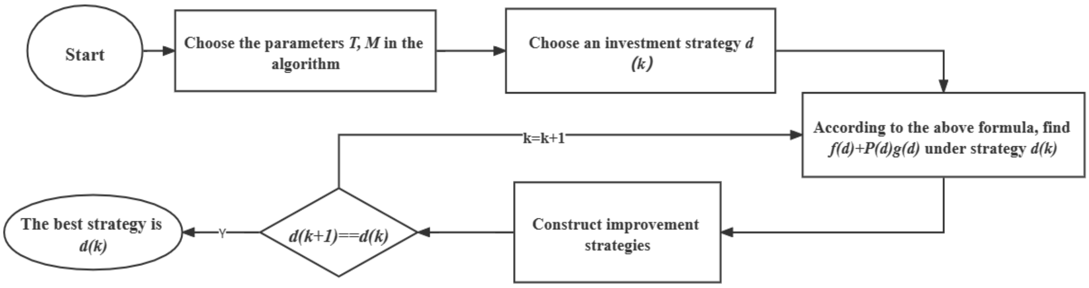

Figure 5: Program flow chart

5. Analysis and Evaluation of results

The final calculation result is $348,287.46, the initial asset is $1,000, and the final holdings of cash, Bitcoin and gold are [348.70,7.50,0.10]. According to the corresponding price of each asset on the day, the asset value at this time can be obtained as 348.70 + 7.50 * 46368.69 + 0.10 * 1794.6 = 348293.34 US dollars.

The chart below shows the history of Bitcoin transactions:

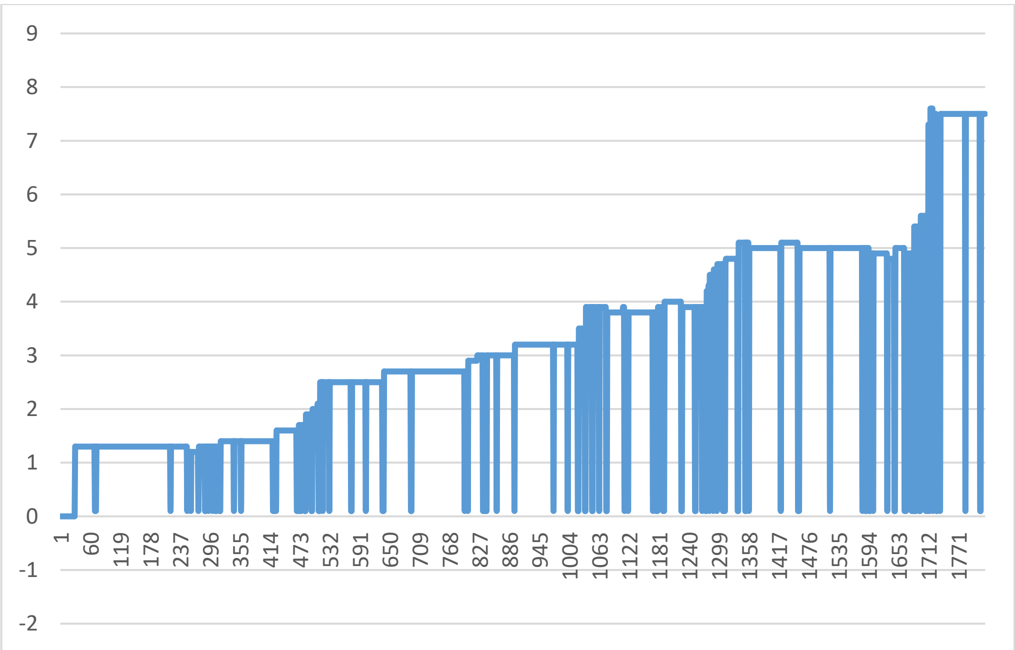

Figure 6: History of Bitcoin transactions

ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

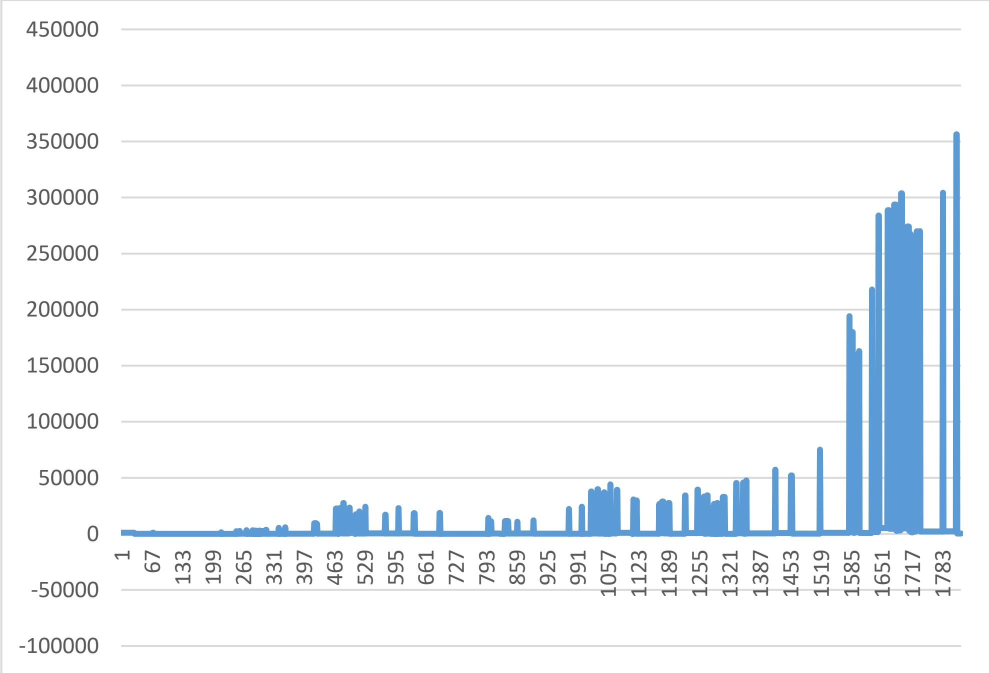

Figure 7: History of cash transactions

From the historical data, it is known that the change of gold trading is basically unchanged. After buying 0.1 unit of gold, the trading volume of gold has almost no change. According to the analysis of the price change curve of Bitcoin and gold,

According to the figure2\figure3 in data pre-processing, the appreciation rate of Bitcoin is significantly higher than that of gold. In the same period of time, for example, by the 1633rd day, Bitcoin has increased by more than 100 times compared with the first day, while the trend of gold is relatively stable. , there has not been a large increase, so the trading curve of gold is very stable, because its total value remains relatively stable, stability is the best trading strategy, and for Bitcoin, because of its large value volatility, so the trading curve Greater volatility is in line with customer needs and is more likely to maximize returns.

6. Model Prove

To prove that the strategy is optimal, only need to prove the strategy iteration formula mentioned in Task 1 can get the optimal solution. First we assume there is a better strategy h than origin strategy d . Mark $\{X^d(t), t \geq 0\}$ and $\{X^h(t), t \geq 0\}$ as strategy d and h 's Markov process. Similarly, other strategy-related quantities are differentiated by different superscripts.

According to Poisson equation[2], the long-term average value-added rate of total assets under the two investment strategies meets the following lemma.

$$\eta^h - \eta^d = \pi^h[(f^h - P^h g^d) - (f^d + P^d g^d)]$$

The direct explanation of the performance difference formula is that under two different investment strategies, the long-term average value-added rate of total assets is determined by two parts, one part is the difference between the current two strategies in one-stage value-added rate. The other part is the long-term impact of the current transaction on the long-term average appreciation rate of total assets in the future.

According to the difference formula, an iterative algorithm to solve the optimal investment strategy can be designed. First, we need to redefine the relationship between the two functions.

In the portfolio optimization problem, because the volatility of risky assets is described by Gaussian distribution, which is supported on any subset of the state space, for all stationary strategies d, their corresponding stationary probability measures are satisfied $\pi^d(B) > 0, \forall B \in \boldsymbol{B}$**,** So you can omit the subscript from the relational symbol π^d ,then the following optimal conditions and equations are obtained.

ISSN 2616-5902 Vol. 4, Issue 4: 89-95, DOI: 10.25236/AJBM.2022.040418

$$f^{\hat{d}} + P^{\hat{d}} g^{\hat{d}} = \max_{d}\{f^{d} + P^{d} g^{\hat{d}}\}$$

This means that when the above formula is true, an optimal solution can be obtained. An iterative algorithm can be designed to solve the optimal investment strategy. The idea is to construct an improvement strategy based on any strategy d_0 and iterating at the d_k

$$d_{k+1} = arg\max_{d}\{f^{d} + P^{d} g^{\hat{d}_k}\}$$

Iteration stops if there is only a difference between d_{k+1} and d_k with the set whose measure is zero. In each iteration, the asset value-added rate of the improved strategy increases. The optimal conditions and equations guarantees that the strategy obtained when the algorithm stops is optimal.

7. Conclusion

The analysis is concise and intuitive, the results are more appropriate to the real situation, and the comparison of the return of the investment strategy, the derivation of the optimality equation, and the design of the strategy iteration algorithm are all based on the difference formula of the appreciation rate of the investment strategy. However, the relationship between risk assets is not taken into account, and it is considered that risk assets are independent of each other, but in fact some risk assets are related to each other, and this correlation may affect the strategy later, and if this relationship is preconceived, it may be difficult to solve the optimal strategy.

References

[1] Taksar M, Klass M J, Assaf D. A diffusion model for optimal portfolio selection in the presence of brokerage fees [J]. Mathematics of Operations Research, 1988, 13(2): 277-294.

[2] Huang Yonghao, Chen Xi. Investment Portfolio Optimization with Proportional Fees Based on Sensitivity Analysis [J]. Control and Decision, 2014, 29(07): 11811186.

[3] Meng Xianghuan. Research on the application of risk portfolio investment model based on multiobjective planning [D]. Wuhan University of Science and Technology, 2013.

[4] Cao X R. Stochastic learning and optimization: A sensitivity-based approach [M]. New York: Springer, 2007: 341-380.

[5] PaperWeekly, When Deep Learning Meets Quantitative Trading - Models, https://blog.csdn.net/c9Yv2cf9I06K2A9E/article/details/106561139, [2021-2-19]

[6] Meng Xianghuan. Research on the application of risk portfolio investment model based on multi-objective planning [D]. Wuhan University of Science and Technology, 2013.

[7] Huang Wenlan. Research on Portfolio Optimization Based on Multi-Objective Evolutionary Algorithm [D]. Harbin University of Commerce, 2020. DOI: 10.27787/d.cnki.ghrbs.2020.000514.