ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

The Quantization Algorithm Based on DRL

Can Pan

School of Agricultural University of Hunan, Changsha, 410000, China
1942219467@qq.com

Abstract: *Maximizing returns is always people's investment goal. Gold and Bitcoin have some hedging ability, and their prices fluctuate greatly, making them popular varieties for investors. However, the market is risky, and different economic, political and environmental factors will impact the market. As a result, Bitcoin and gold prices fluctuate sharply, leading to uncertain investment and uncertain returns. To maximize returns, we model time and price, analyze and model the price trends of gold and bitcoin in problems, give the best trading model, and analyze the transaction cost sensitivity of transaction risk and trading strategies. Predictive techniques are used to move the future and to build some heuristic-based robots to make decisions. Use (DRL) deep reinforcement learning to simulate stock trading based on the Markov decision-making process of avoiding risk aversion, reducing trading costs, maintaining liquidity, and assuming that the stock market is not affected by enhanced trading agents. The second step is to use the risk analysis of the Sharp ratio to compare with the portfolio-managed Markowitz effective boundary model to select a portfolio with medium risk and high yield in the third step, to effectively calculate the sensitivity of the transaction strategy to the transaction cost, we used the time series to calculate the change difference between the final result and the standard value after the transaction cost change. A system of equations based on a number of yield values can be established to meet the optimization requirements, introduce computational historical data, and replace the relevant transaction costs with the above model to clarify the change curve of the final return results. Later, ac- cording to the needs of the trading strategy, complete the sensitivity analysis of the transaction strategy on the trading cost (calculate the impact of variable factors on profit), determine the sensitivity anal- ysis mode, and evaluate the advantages and disadvantages of the model combined with the above results. Ultimately, our strategy and models are passed on to traders as memos.*

Keywords: *Maximizing returns, deep reinforcement learning, Markowitz' efficient frontier, Markov decision-making process, sensitivity analysis mode*

1. Introduction

1.1. Background

Market traders often buy and sell volatile assets with the goal of maximising their total returns. There is usually a commission on every sale. T wo of those assets are gold and bitcoin. There is volatile property in the market, and its value constantly fluctuates over time. Traders can maximize total returns by constantly buying and selling volatile property to continuously increase the total value of the holding property. But different investment projects require a different commission for each sale. Investing is not blind, the key issue of trading decision is to execute the right decision [1] at the right time. When holding a variety of property such as gold, Bitcoin needs more strategy. Investors decide whether to buy the property or sell or continue to hold the property in the portfolio. We can provide us with the current transaction optimal strategy in the daily price changes of investment property known to date, finally achieving the purpose of maximizing the corresponding returns.

1.2. Problem Restatement

To explore the best investment trading strategy, our team has studied it with gold and Bitcoin.Bitcoin can be traded every day, gold can only be traded on trading days, and they have different trading commissions.Our team studied investment trading strategies by performing different portfolios for gold and bitcoin during the five-year trading period from November 9,2016 to October 9,2021.And completed the following questions:

Task1 Develop a trading model based only on the daily price data as of the day, giving the best trading

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

strategy of the day, including how to buy, sell and continue to hold property in the portfolio

Task2 Ensure that the trading strategy providedby the model is optimal in the current environment

Task3 Test the sensitivity analysis of the transaction cost of the transaction strategy

Task4 Make a memo of no more than two pages to show traders our model, strategy, and results.

It needs to start at $1,000 by November 9, 2016. Used since November 2016 A five-year trading period from 9 October to 9 October 2021. On each trading day, traders will have a portfolio of cash, gold and bitcoin [C, G, B]. Dollar, Troy ounce and Bitcoin. The initial state is [1000, 0, 0]. The commission cost for each transaction is α% of the transaction amount. Assume alpha gold =1% and alpha Bitcoin = 2%. There is no cost to holding assets.

1.3. Our Work

For Task1, we propose a model framework as shown in Figure 1;

For Task2, we compare our model with Markowitz model;

For Task3, one approach is to use forecasting techniques to predict the movement of the stock and build some heuristic based bot that uses the prediction to make decisions. Another approach is to build a bot that can look at the stock movement and directly recommend the actions — buy/sell/hold.

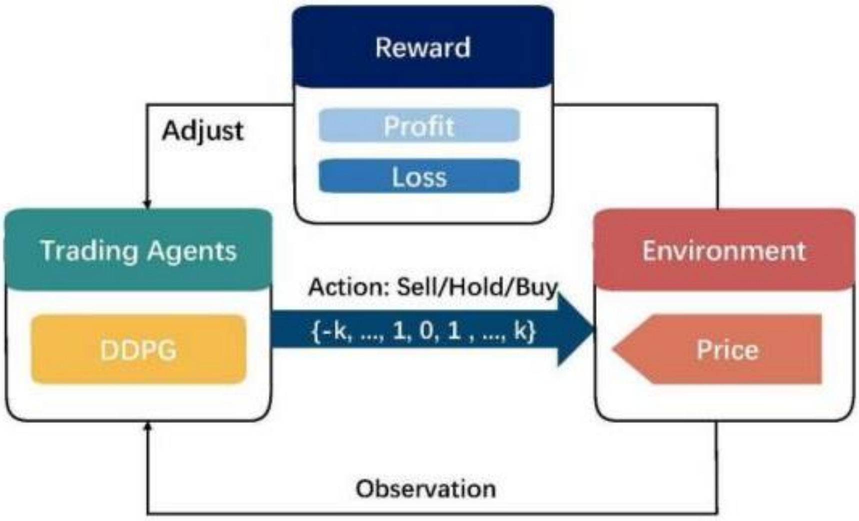

Figure 1: Framework

2. Assumptions

Assumptions 1: Our trading actions are regulated and we can buy and sell freely

Justification: We assume that our trading orders are legitimate and we can buy and sell or short-sell with no restrictions.

Assumptions 2: The prices are influenced by both systematic risks, and non-systematic risks

Justification: We assume that the prices of gold and bitcoin are affected by systematic risk inherent to the entire market or market-segment.The non-systematic risk, the industry or company specific risk which is inherent in every investment also have an impact on prices of gold.

Assumptions 3: Assume the research data is true and reliable.

Justification: We assume that the data provided is real and reliable. Based on this, we can simulate more real scenarios and establish reasonable models.

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

3. Notation

Symbol Description is shown in table 1.

Table 1: Symbol Description

Symbol	Description
α	the percentage of transaction costs of every single transaction
C_{equity}	arbitrary constant
$P_{date,equity}$	prices of each equity on each trading date
n	the total number of trading days of each equity
w_t	the portfolio weights of on trading date t
r_t	return on trading date t
r_p	the expected return of portfolio
r_f	the risk free rate
σ_p	the portfolio standard deviation
X^i	investment proportion coefficient of risky assets i
$w_{i,j}$	the co-variance matrix of the yield rate of n types of portfolio investment of i^{th} risky asset and j^{th}

4. Model Preparation

The data provided consist of the closing price of a troy ounce gold and the price of a bitcoin of each day during the five-year period commencing 9/11/2016. The data is pre-processed in the following fashion:

- The gold is not traded when the market is closed and as a result many dates are absent from the data. Additionally, some dates included in the data set do not have the corresponding closing price. All these dates are labelled as "not trading gold".

- It is assumed that the trader is able to trade both gold and bitcoin in fractional shares. To simplify our model, we divide all prices of each equity by an arbitrary constant C_{equity},where $C_{equity} \geq P_{date,equity}$ for all dates, and the resultant values $C_{equity} \geq P_{date,equity}$ are used as the minimum units of trade for the equity.

- For the model to mimic the decisions of human traders and fully exploit the patterns of the data, several common technical indicators are added to each equity of each date, which are:

– the n-day simple moving average, i.e. the average closing price for previous n trading days, where we take n as 5, 10, 30 to capture the trend of the prices of different length of periods;

– the difference of the price of the current day and the previous day;

– the prices of previous n days, where we take n as 1, 2, 3, 4, as the most recent prices have the largest influence on future prices based on the common random walk hypothesis[];

– the moving average convergence divergence (MACD), calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA;

– the signal line, which is a nine-day EMA of the MACD;

– the volatility, calculated by the standard deviation of previous 20 trading days, which measures the intensity of recent fluctuation;

– the bias, which is the percentage of the deviation of current day price from the 5-day simple moving average.

because the calculation of some indicators requires data of previous dates, the data set is dropped until 21/10/2016 so that all dates have all indicators stated.

Similar to the tutorial FinRL: Multiple Stock Trading, we model the portfolio management process as a Markov Decision Process (MDP). We then formulate our trading goal as a maximization problem. The algorithm is trained using Deep Reinforcement Learning (DRL) algorithms and the components of the reinforcement learning environment are:

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

• **Action:** The action space describes the allowed actions an agent can take at a state. In our task, the action $\omega(t) \in RN$corresponds to the portfolio weight vector decided at the beginning of time slot, and should satisfy the constraints: firstly, the each element is between 0 and 1, secondly the summation of all elements is 1.

• **Reward function:** The reward function $(s(t)$,$\omega(t)$,$s(t+1)$ is the incentive for an agent to learn a profitable policy. We use the logarithmic rate of portfolio return: In($\omega(t)^T y(t)$) as the reward, where $y(t) \in RN$ is the price relative vector.

• **State:** describes an agent's perception of a market. The state at the beginning of time slot is $s(t) =$

• **SRL Algorithms:** We use two popular deep reinforcement learning algorithms: Advantage Actor Critic (A2C) and Proximal Policy Optimization (PPO).

• **Environment:** Dow Jones 30 constituent stocks during 01/01/2009 to 09/01/2021

Action: portfolio weight of each stock is within [0,1]. We use softmax function to normalize the actions to sum to 1.

State: Covariance Matrix, MACD, RSI, CCI, ADX, state space shape is (34, 30). 34 is the number of rows, 30 is the number of columns.

Reward function: r(s, a, s′) = p_t, p_t is the cumulative portfolio value.

Environment: portfolio allocation for Dow 30 constituents.

5. Model 1

Table 2: Portfolio allocation using deep reinforcement learning (DRL)

Protocol 1 Portfolio allocation using deep reinforcement learning (DRL)

1): **Input:** s, state space includes covariance matrix for stocks and technical indicators

2): **Output:** Final portfolio value
3): Initialize P_0 =$1,000,000, ω_0=($\frac{1}{m}$,...,$\frac{1}{m}$), P_0 is the initial portfolio value,ω_0 is the initial portfolio weights, m is the number of stocks in the portfolio;
4): **for** t = 1,..., n **do**
5): Portfolio manager DRL observes a state *s* and outputs a portfolio weights vector ω_t;
6): Normalize the weights ω_t to sum to 1;
7): Calculate stock returns vector γ_t = ($\frac{v_1,t-v_1,t-1}{v_1,t-1}$,... ,$\frac{v_m,t-v_m,t-1}{v_m,t-1}$), *v* is the closing price;
8): Portfolio incurs period return $\omega_t^T\gamma_t$;
9): Update portfolio value $P_t = P_{t-1}$×(1+$\omega_t^T\gamma_t$);
10): **end**

The model (table 2) is based on Deep Reinforcement Learning (DRL), which combines the framework of reinforcement learning and deep learning techniques.

The Markov Decision Process (MDP) is used to describe the trading process. We have the following definitions:

• **A portfolio** *P* = (C,n_G,n_B) is the allocation of equities on a given date, where $C \in [0, +\infty)$ is the amount of cash in the portfolio, and n_G,$n_B \in$ N respectively represent the number of shares of gold and bitcoin in the portfolio;

• **An observation** *O* is the share price p_G, p_B, whether gold is traded, and other technical indicators of the equities on a given date;

• A given date is thus represented as **a state** *S* = (*P*, *O*);

• Given a state *S*, **the value** of an equity *E* is $v_E = n_E p_E$, and **the value** *V* of a state *S* is the sum of cash and the market values of all equities in the portfolio under current market prices,$V = C + v_B + v_G$;

• **An action** $A = (a_C, a_G, a_B)$ where $0 \leq a_C, a_G, a_B \leq 1$, $a_C + a_G + a_B = 1$, its elements respectively representing the intended ratio of value allocated to cash, gold and bitcoin, is taken by **an agent** given a state *S* under a certain policy π;

• **An equity allocation function** *f* takes an action *A* and produces a new portfolio *P* under certain

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

restrictions;

• **A policy** π is the agent's behavior function describing the trading strategy at state *S*, standing for the probability distribution of actions at state *S*;

• **The reward** *R* of an action *A* under a state *S*, which results in the new state *S′*, is the difference of the values of the two states $R = V(S') - V(S)$;

• **Q-value** $Q_\pi(S, A)$ is the expected value of reward of taking action *A* at state *S* under policy π.

(More restraints of the definitions)

The environment is modelled to simulate that of a trader in the gold and bitcoin market. To train the model each environment randomly picks a *N*-day period and slice the data only from these dates.The portfolio of the trader is initialized as (1000, 0, 0). The environment then informs the agent of the first date's state. Then on each day, it sequentially takes the following steps:

• The environment receives the action *A* given by the agent and updates the traders portfolio as $P' = f$ (A) according to the equity allocation function.

• The environment shifts to the next date. It calculates the reward of given the new date's market and informs the agent of the reward as well as the new date's state.

A naive way of defining *f* is to find the closest number of shares of such a value ratio. We define such an *f* as f_0.

The agent needs to develop a set of rules that decide the best award based on the given state and reward. This is taken care of with the use of a DdpgAgent by TensorFlow. The agent is trained with γ = 0.05, batch size 60 and iteration number 400. (Details about random policy, replay buffer, etc.)

(some graphs here)

After training the model is run on the entire data set to obtain the overall profit rate of the strategy.(We need some indicators of the model's performance.)

When the transaction fee is zero the trained model with f_0 performs quite well with profit rate ...

(This is actually question 3) However when the transaction fee gets higher the model tends to degenerate... (some terrible graphs here)

This is due to the high frequency of transaction taken by the model. We mitigate this by redefining *f*. In order to reduce unnecessary transaction fee, no transaction will be conducted if the change rates of n_G,n_B are both less than β%.

When β takes different values we have ... (graphs here)

Thus a reasonable value for β by comparison is ...

In this section we will use the Markov Decision Process (MDP) to model the stock trading process and the objective function will be returned when it achieves its maximum.

We aim to describe the stochastic process of dynamic stock market, a MDP was deployed as following:

• **State:** $s = [p, h, b]$: a vector that includes stock prices $p \in RD+$, the stock shares $h \in ZD+$,and the remaining balance $b \in R+$, where *D* denotes the number of stocks and *Z+* denotes non-negative integers.

• **Action:** a vector of actions over *D* stocks. The actions could be taken upon each of share includes selling, buying, or holding, representing for decreasing, increasing, and no change of the stock shares *h*, respectively.

• **Reward:** $\gamma = (s, a, s')$: the incentive mechanism for an agent to learn a better action. At state *a*,the firsthand reward of taking action *a*, and then arriving at the new state *s′*.

• **Q-value:** $Q_\pi(s, a)$: under policy π, the expected value of reward of taking action *a* at state *s*.From figure 2, the state transition of stock trading is shown. In the portfolio, there are three possible actions could be taken on stock $d (d = 1, 2, 3, \ldots\ldots, D)$.

• Selling $k[d] \in [1, h[d]]$ shares results in $h_{t+1}[d] = h_t[d] - k[d]$, where $k[d] \in Z+$ and $d = 1, \ldots, D$.

• Similarly, holding results in $h_{t+1}[d] = h_t[d]$.

• Finally, buying $k[d]$ shares results in $h_{t+1}[d] = h_t[d] + k[d]$. Action is taken at time t and then the stock price changes at time $t + 1$. Consequently, the portfolio may be updated, for example, from portfolio value 0 to portfolio value 1. Where the portfolio value is $pTh + b$.

Now, consider the following constraints and assumptions, which are mainly used for practical reasons: risk aversion, transaction costs, liquidity.

• Liquidity: we assume that stock market will not be affected by our reinforcement trading agent. These orders can be rapidly executed at the closing price.

• Balance b being non-negative: the balance should not be less than 0, i.e negative. At time t, the stocks are divided into sets for selling S, buying B, and holding H, where $S \cup B \cup H = 1, , D$. Meanwhile, there is no overlappings. Let $pBt = [pit : i \in B]$ and $ktB = [kti : i \in B]$ be the vectors of price and number of buying shares for the stocks in the buying set. We can similarly define pSt and ktS for the selling stocks, and pHt and ktH for the holding stocks. Therefore,there is a relationship stating the constraints:

$$b_{t+1} = b_t + \left(pSt\right)TktS - \left(pBt\right)TktB \geq 0 \tag{1}$$

• Transaction cost: transaction costs are included within each of the trade. We assume our commission to be 0.1% for gold and 0.2% for bitcoin of each trade as:

$$c_t = pTkt * 0.1\%, c_t = pTkt * 0.2\% \tag{2}$$

• Risk aversion for market crash: there is always a possibility that some events might cause the market crash. To control this, financial turbulence index turbulence is used to measure the extreme value movements of the asset:

$$turbulencet = \left(y_t - \mu\right)\Sigma -1\left(y_t - \mu\right)' \in R \tag{3}$$

6. Model 2

6.1. Risk Analysis by Sharpe Ratio

The sharpe ratio is calculated as:

$$sharpe\ ratio = \frac{\bar{r}_p - r_f}{\sigma_p} \tag{4}$$

where $\bar{r}_p$ is the expected portfolio, r_f is the risk free rate, and σ_p is the portfolio standard deviation.

6.2. Comparison with Markowitz' Efficient Frontier

6.2.1. Markowitz' Efficient Frontier

Markowitz's Efficient Frontier which is used to manage portfolio can be expressed as the following mathematical model:

$$\min \sigma_p^2 = \mathbf{X}'\Omega\mathrm{X} = \sum_{i=1}^{n} X_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} X_i X_j \sigma_{ij} \tag{5}$$

$$s,t \quad r_p = \sum x_i r_i$$

$$\sum x_i = 1$$

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

Where r_p is the expected yield rate of return on portfolio investment, σ_p^2 is the variance of the yield rate of portfolio investment. X = (X1, · · ; Xn)\ represents n types of investment proportion coefficient vector of risky assets. $\Omega = (\sigma_{ij})_{n \times n}$ represents the co-variance matrix of the yield rate of n types of portfolio investment, $\sigma_{ij} = cov(ri, rj)$, i, j = 1, · · ·, n, standing for two risky asset co- variance, and ri, i = 1, ··; n represents the yield rate of ith asset.

This approach proposes a framework to evaluate the risk and returns of a portfolio.

The line of efficient frontier shows the portfolios with the highest returns for a given risk profile.

For our evaluation, we designed an agent to pick a moderate risk high reward portfolio from an efficient frontier graph calculated at every time step based on the previous 30-time steps' performance. Average returns: -1%

6.2.2. Comparison

Below is a side-by-side comparison of the 2 policies in the same environment:

RL grows portfolio to 31%, Markowitz' shrinks to 96% Our model has several advantages over Markowitz:

(a) We can see that the efficient boundary is not effective for highly volatile assets such as Bitcoin. However, during times of increased volatility or when all assets are going down, the RL decides to hedge the losses by selling the assets and increasing the cash in hand — very smart strategy when we haven't enabled a short-sell option.

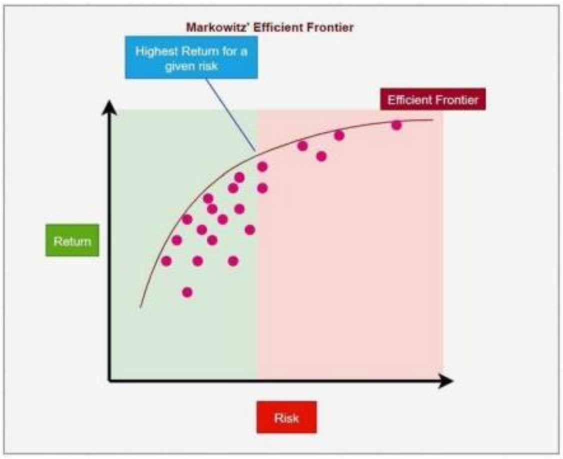

Figure 2: Markowitz' Efficient Frontier

(b) The traditional Markowitz model uses the average return and risk of assets in a period of time to analyze the risk and return, but the actual situation is that both investment return and investment risk change in a period of time.

(c) The theory assumes a single-term investment, in which capital is allocated once at the beginning of the period and nothing is changed thereafter. This is not an actual investment behavior, nor does it apply to multi-stage problems. But the RL strategy seems to be to identify bursts of small surges in price and capitalize on that immediately.

(d) The classical mean-variance model is often threatened by real data. In the real financial market, the distribution of a series of returns often deviates from the normal distribution, presenting kurtosis and skewness, which makes the variance of returns unsuitable for risk measurement.

Therefore, although Markowitz's portfolio theory has laid a foundation for the development of modern portfolio research, it still has some limitations when facing practical problems. We see that RL consistently outperforms Markowitz' approach in our experiments.

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

7. Model3

7.1. Problem solution

To effectively calculate the sensitivity of the transaction strategy to the transaction cost, we used the time series to calculate the change difference between the final result and the standard value after the transaction cost change.

We aimed to find transaction costs with higher yields, with the change difference measured as:

$$\begin{cases} \sqrt{\left(x_{o1}-x_1\right)^2+\left(y_{o1}-y_1\right)^2}=r_1 \\ \sqrt{\left(x_{o1}-x_3\right)^2+\left(y_{o1}-y_3\right)^2}=r_1 \\ \left(x_1-x_3\right)^2+\left(y_1-y_3\right)^2=2r_1^2-2r_1^2\cos a \end{cases} \tag{6}$$

$$\begin{aligned} &\left(z-z_1\right)^2+\left(y-y_1\right)^2=r_1^2 \\ &\left(z-z_2\right)^2+\left(y-y_2\right)^2=r_2^2 \\ &\left(z-x_n\right)^2+\left(z-y_n\right)^2=r_n^2 \end{aligned} \tag{7}$$

We build a system of equations based on multiple yield values, so that they can meet the optimization requirements.

$$\begin{aligned} &2z\left(z_n-z_1\right)+2y\left(y_n-y_1\right)=z_n^2-z_1^2+y_n^2+r_1^2-r_n^2 \\ &2z\left(z_n-z_2\right)+2y\left(y_n-y_2\right)=z_n^2-z_2^2+y_n^2+r_2^2-r_n^2 \\ &\cdots \\ &2z\left(z_n-z_{n-1}\right)+2y\left(y_n-y_{n-1}\right)=z_n^2-z_{n-1}^2+y_n^2+r_{n-1}^2-r_n^2 \end{aligned} \tag{8}$$

The derivation of the simplification can be obtained:

$$M=\begin{bmatrix} 2\left(z_n-z_1\right) & 2\left(y_n-y_1\right) \\ 2\left(z_n-z_1\right) & 2\left(y_n-y_2\right) \\ M & M \\ 2\left(z_n-z_{n-1}\right) & 2\left(y_n-y_{n-1}\right) \end{bmatrix} \tag{9}$$

$$N=\begin{bmatrix} z_n^2-z_1^2+y_n^2-y_1^2+r_1^2-r_n^2 \\ z_n^2-z_2^2+y_n^2-y_2^2+r_2^2-r_n^2 \\ \cdots \\ z_n^2-z_{n-1}^2+y_n^2-y_{n-1}^2+r_{n-1}^2-r_n^2 \end{bmatrix} \tag{10}$$

$$Z=\begin{bmatrix} x \\ y \end{bmatrix} \tag{11}$$

Get its optimized location:

$$Z_{\text{est}}=\left(M^TM\right)^{-1}M^TN \tag{12}$$

If the time series $\{Y_t\}$ satisfied:

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

$$Y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

The partial autocorrelation function of the sample is:

$$\varphi_{kk} = \begin{cases} \hat{\rho}_1 \\ \dfrac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j} \cdot \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j} \cdot \hat{\rho}_j} \end{cases} \tag{13}$$

We conducted a comprehensive analysis to obtain the optimization mode under different circumstances. We calculated the maximum return value when the transaction cost is the gold transaction cost and the bitcoin transaction cost combination as 0.005,0.01,0.02,0.02,0.04,0.05 and 0.1, as shown in figure 3~5.

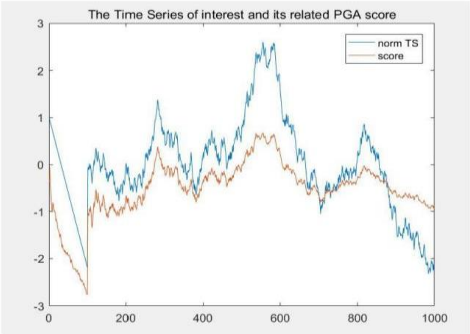

Figure 3: Gold transaction cost 0.005, Bitcoin transaction cost 0.01

In fact, in order to maximize the investment utility, we need to calculate the historical data and substitute the relevant transaction costs with the model described above, so as to clarify the change curve of the final benefit results.

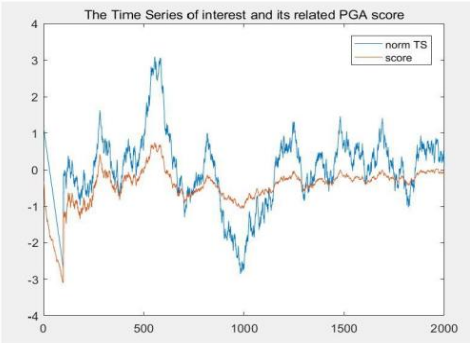

Figure 4: Gold transaction cost 0.02, Bitcoin transaction cost 0.04

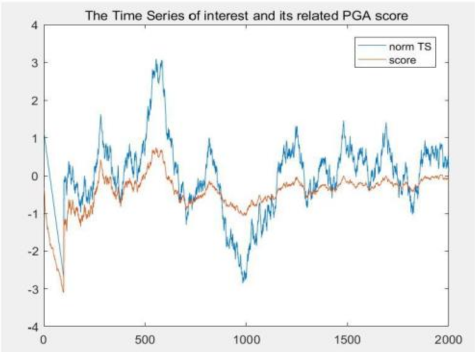

Figure 5: Gold transaction cost 0.05, Bitcoin transaction cost 0.01

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

We can calculate the maximum value the regression fitting algorithm:

$$\sum_{i=1}^{n} y_i \alpha_i = 0 \tag{14}$$

$$\alpha_i \geq 0, i = 1, 2, \ldots, n$$

The maximum value of the following function is solved below for α_i:

$$\mathrm{Q}(\alpha) = \sum_{\mathrm{i}=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \left(x_i x_j \right) \tag{15}$$

If $\alpha*$ is the optimal solution, then $w^* = \sum_{i=1}^{n} \alpha^* y \alpha_i$

$$L(w,b,\alpha) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{l} \alpha_i \left(y_i \cdot \left(\left(x_i \cdot w \right) + b \right) - 1 \right)$$

$$\frac{\partial}{\partial b} L(w,b,\alpha) = 0 \frac{\partial}{\partial w} L(w,b,\alpha) = 0$$

$$\sum_{i=1}^{l} a_i y_i = 0 \quad w = \sum_{i=1}^{l} \alpha_i y_i x_i$$

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j \right) \tag{16}$$

$$\alpha_i \geq 0, i = 1, \ldots, l, \quad \text{and} \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i \cdot \left(x \cdot x_i \right) + b \right)$$

$$f(x) = \operatorname{sgn}\left\{ \left(w^* \cdot x \right) + b^* \right\} = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i \left(x_i \cdot x \right) + b^* \right\}$$

What we should calculate is the plane:

$$y_i \left[\left(w \cdot x_i \right) + b \right] - 1 + \varepsilon_i \geq 0, i = 1, 2, \ldots, n$$

$$F_\sigma(\varepsilon) = \sum_{i=1}^{n} \varepsilon_i^\sigma \tag{17}$$

$$\| w \|^2 \leq c_k \tag{18}$$

Computing the maximum value is considered:

$$\phi(w,\varepsilon) = \frac{1}{2}(w,w) + C\left(\sum_{i=1}^{n} \varepsilon_i \right) \tag{19}$$

Using such a model, we will get the maximum investment utility and calculate the change situation and the return maximum. As can be seen from the figure, the different transaction costs will change the location of their maximum value, based on the fact that the different transaction costs can be changed in the time series by affecting the investment expectations.Different transaction costs do not have much impact on the maximum, presumably because the maximum has been greatly affected by the daily price in the past, and long transactions have stabilized it.

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

8. Sensitivity Analysis

In fact, we can complete the analysis of the sensitivity of the transaction strategy to the transaction cost according to its requirements, and we determine its sensitivity analysis mode.

Determination scheme for the optimal values:

$$R_i = \sqrt{(Z_i - z)^2 + (Y_i - y)^2}$$

$$R_i^2 = (Z_i - z)^2 + (Y_i - y)^2 = K_i - 2Z_i z - 2Y_i y + z^2 + y^2$$

$$K_i = Z_i^2 + Y_i^2$$

$$R_i^2 = (R_{i,1} + R_1)^2 = R_{i,1}^2 + 2R_{i,1}R_1 + R_1^2$$

It can be obtained after the merger:

$$R_{i,1}^2 + 2R_{i,1}R_1 + R_1^2 = K_i - 2Z_i z - 2Y_i y + z^2 + y^2$$

$$R_1^2 = K_1 - 2Z_1 z - 2Y_1 y + z^2 + y^2$$

Do bad can get:

$$R_{i,1}^2 + 2R_{i,1}R_1 = K_i - 2Z_{i,1}Z - 2Y_{i,1}y - K_1$$

$$-2Z_{i,1}z - 2Y_{i,1}y = R_{i,1}^2 + 2R_{i,1}R_1 - K_i + K_1$$

$$\begin{cases} Z_{i,1} = Z_i - Z_1 \\ Y_{i,1} = Y_i - Y_1 \quad i = 2\text{L L} \\ K_i = Z_i^2 + Y_i^2 \end{cases}$$

Its expression for sensitivity analysis for three different transaction cost ratios:

$$z = \frac{\frac{z_1}{l_1 + l_2} + \frac{z_2}{l_2 + l_3} + \frac{z_3}{l_1 + l_3}}{\frac{1}{l_1 + l_2} + \frac{1}{l_2 + l_3} + \frac{1}{l_1 + l_3}}$$

$$y = \frac{\frac{y_1}{l_1 + l_2} + \frac{y_2}{l_2 + l_3} + \frac{y_3}{l_1 + l_3}}{\frac{1}{l_1 + l_2} + \frac{1}{l_2 + l_3} + \frac{1}{l_1 + l_3}}$$

$$x = \frac{\frac{x_1}{l_1 + l_2} + \frac{x_2}{l_2 + l_3} + \frac{x_3}{l_1 + l_3}}{\frac{1}{l_1 + l_2} + \frac{1}{l_2 + l_3} + \frac{1}{l_1 + l_3}}$$

Figure 6 shows the relationship between gold transaction cost and total assets when keeping the bitcoin transaction cost at 0.01; Figure 7 shows the relationship between bitcoin transaction cost and total assets when keeping the gold transaction cost at 0.01. The generation of results is more sensitive to changes in gold transaction costs. As can be seen in figures 6 and 7, as the transaction cost increases, the total assets ultimately obtained also decline significantly. The point itself that increasing costs make profits decrease is obvious. At the same time, our model avoids unnecessary costly waste and chooses to remain in a wait-and-see mode, so the cost also influences the trading strategy to some extent, making the model degenerate. The rate of decline of total assets in both charts is fast and slow, but the effect shown in the graphs shows that the degree of decline of the curve is still in a relatively stable state, indicating the health of our model.

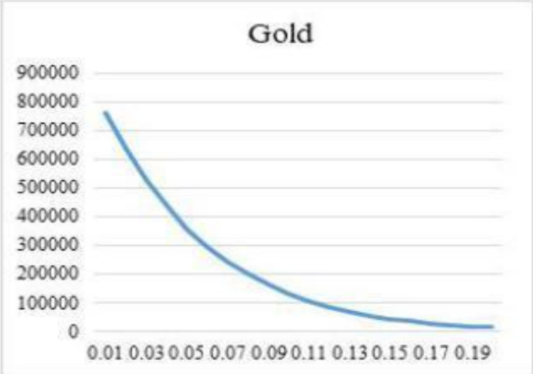

Figure 6: Relationship between gold transaction cost and total assets

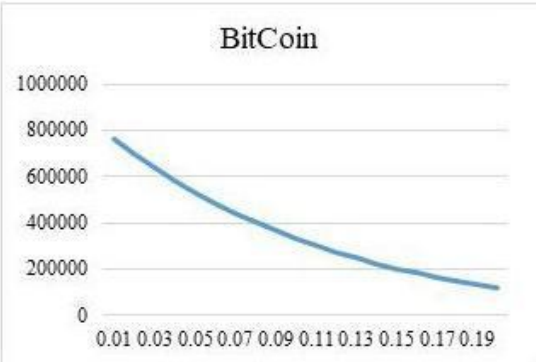

Figure 7: Elationship between bitcoin transaction cost and total assets

9. Strengths and Weaknesses

The model can give an optimization and improvement scheme for the relevant trading mode in a relatively short time, so that its efficiency, cost, and quality are all at a good level, and its effect is good, which can be dealt with for such problems.The model has a good application effect for such optimization problems, and can easily find its optimal trading mode, which is of great significance to its future research and development, and needs to be further improved through multiple modes. At the same time, it can effectively regulate the possible risk value. After optimizing according to the goal of risk reduction and utility improvement, we can get its optimal value. When optimization, the model needs comprehensive analysis of multiple data to meet the specific requirements. Various conditions and values should be kept within a reasonable range. This is more difficult, and it needs to further optimize the model and control it.At the same time, for reinforcement learning and other methods, the relevant parameters need to consider, the adjustment of hyperparameters has high requirements, and the optimization algorithm should be used to obtain a more reasonable learning rate, iteration number and other parameters.

9.1. Strengths

(1) The gold and bitcoin price prediction models based on DRL are scientific and reasonable, and can not only grasp the general characteristics of price data, but also grasp the time series characteristics of price data.

(2) Sensitivity analysis of the model is carried out, which proves the validity of the model under various conditions and has strong robustness.

ISSN 2616-5775 Vol. 5, Issue 9: 98-110, DOI: 10.25236/AJCIS.2022.050915

9.2. Weaknesses

(1) In the early stage of prediction, the integrated system will suffer from shock due to lack of experience, resulting in poor decision-making results. If there are other investment goals to pre-train this system, it may achieve better results.

(2) The integrated system does not consider the decay of past experience over time, but only updates with the number of successful decisions.

10. Memorandum

Washington State Department of Agriculture Team Analysis and improvement in the reported sighting data February 21, 2022 [Memorandum]

Intention

We are here to provide our interpretation of the data provided by the public reports and to provide a method to prioritize public reports for further investigation.

Models for improving the report system

Possible Improvements

There are several improvements that can useful to help get information from the public and reduce the possibility of mistaken identification.

References

[1] Shahriar Shafiee, Erkan Topal, An overview of global gold market and gold price forecasting, Resources Policy, Volume 35, Issue 3,2010,Pages 178-189, ISSN 0301- 4207, https://doi.org/10.1016/ j.resourpol.2010.05.004.
[2] Letteri, I., Della Penna, G., De Gasperis, G., and Dyoub, A.. "A Stock Trading System for a Medium Volatile Asset using Multi Layer Perceptron", 2022.
[3] M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi. "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in IEEE Access, vol. 8, pp. 150199-150212, 2020, doi: 10.1109/ACCESS.2020.3015966.
[4] C. Zheng and J. Zhu. "Research on stock price forecast based on gray relational analysis and ARMAX model," 2017 International Conference on Grey Systems and Intelligent Services (GSIS), 2017, pp. 145-148, doi: 10.1109/GSIS.2017.8077689.
[5] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman. "Stock price prediction using LSTM, RNN and CNN-sliding window model," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1643-1647, doi: 10.1109/ICACCI.2017.8126078.