Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
C		**2205329**

Quantitative Investment Strategy for Gold and Bitcoin

Summary

In the context of global inflation, it becomes more and more important to choose and invest in safe haven assets. With the continuous circulation of gold and bitcoin, market traders are constantly walking in the market, buying and selling assets with large fluctuations (gold and bitcoin). They use their unique vision and strategies maximize the value of their total assets. Inspired by the past stream of gold and bitcoin daily prices, we develop a trading model and use the $1000 capital on September 11, 2016 to achieve higher returns within five years using the best trading strategies.

Firstly, we establish a combined model training based on **Random Forest** and **Long Short-Term Memory (LSTM)**[1]. It can learn the closing price and change trend of gold and bitcoin on specified dates in the first four months, and then forecast the buying situation and closing price trend of gold and bitcoin after December 11, 2016. The **New Quantitative investment Strategies based on portfolio theory** are used to determine the specific proportion of gold and bitcoin buying, selling and holding on that day. Combining the above forecasts and allocation models, **final value** we arrive at on September 10, 2021 is **$155521.73**.

Secondly, we not only use the **evaluation index P** to select the best decision model - random forest from many machine learning models, but also add **random factors** to simulate the random decision to find **optimal planning**. The final value is reduced to some extent by random sowing of the volatility (threshold) in the setting judgment conditions, random perturbation of the predicted growth rate in a certain range and random selection of allocation strategy combination by adding random factors. Thus, the optimality of our combined model and strategy is proved. Meantime, the most appropriate volatility (threshold) is 14.08%.

Thirdly, we change the transaction cost by [-80%, +80%] to have the **sensitivity analysis**. Through program simulation, we discuss the impact of the transaction costs of gold and Bitcoin on final returns respectively and in combination. By calculating the sensitivity and final income, we conclude that the change of transaction costs between -20% and 20% can still obtain a higher final return. That is, our model has high model **robustness**. For the change of transaction costs within this range, we don't need to change our strategies and still have a better target value. Beyond this range, we need to adjust the volatility to update our strategies to achieve better returns.

After the modeling is completed, we analyze the advantages and disadvantages of the model, and talk about how to improve our model and strategies in the end.

Keywords: Random forest, Term mermory, quantitative investment, Random factors, Sensitivity analysis

Contents

1 Introduction

1.1 Background

Gold is one of the rarest and most valued metals. It is not only a special currency for reserve and investment, but also an important material for jewelry industry, electronics industry, modern communications, aerospace and aviation industries. The concept of Bitcoin was first proposed by Satoshi Nakamoto on November 1, 2008 and was officially born on January 3, 2009. It's a peer-to-peer, decentralized digital asset. In 2009, the Bitcoin network was born by its founders who created the first Bitcoin block with a strong scarcity. Market traders often buy and sell volatile assets with the goal of maximizing total returns. There is usually a commission on each purchase, and two of the assets[2] we talk about are gold and bitcoin.

1.2 Problem Restatement

Let us now give the Problem Restatement.

- Develop a model to give a best daily strategy based only on price data up to that day and get the final worth on 9/10/2021 of the initial investment $1000 by using our model and strategy.
- Provide evidence of the model provides the best strategy.
- Determine the sensitivity of the strategy on transaction costs and how transaction costs affect the strategy results.
- Communicate the strategy, model, and results to the trader in a memorandum of at most two pages.

1.3 Our Work

We establish a combined prediction model, which is based on Random Forest and Long Short-Term Memory (LSTM)[3] to train the closing price and trend of gold and bitcoin on the specified date in the first four months, so as to predict whether gold and bitcoin will be bought on the specified date and the changing trend of the closing price after December 11, 2016. The New Quantitative investment Strategies based on portfolio theory is used to determine the specific proportions of gold and bitcoin buying, selling and holding that day. Then, we get the final value of our cost on September 10, 2021. After that, we add random factors to simulate random decisions. The evidence that the final value is reduced to a certain extent proves that the model and asset allocation strategy we choose are correct and better. Moreover, we analyze the sensitivity of strategy to transaction costs by controlling the fluctuation of transaction cost, and explain the influence of transaction costs on strategy and final result. After analyzing the strengths and weaknesses of our model and strategy and coming to a final conclusion, we write the model, strategy and result in a memo.

2 Assumptions and Notations

2.1 Assumptions

To simplify the problem, we make the following basic assumptions, each of which is properly justified.

- When trading, we follow the utility maximization principle.
- The initial funds are all invested.
- Market traders are risk avoiders, choosing the portfolio with the lowest risk under the condition of equal returns.

- Gold and bitcoin can be wirelessly subdivided.

2.2 Notations

Throughout the paper, we define all the variables used in Table 1.

Table 1: Definition of the Symbol

Symbol	*Definition*	*Unit*
t	Time	\
M_A	Total value of gold we have	USD
M_B	Total value of Bitcoin we have	USD
$Cash$	Total value of cash we have	USD
d_A	The amount of money to be redistributed to buy gold	USD
d_B	The amount of money to be redistributed to buy Bitcoin	USD
C_i	One day's closing price	\
$C_{A,t}$	One day's closing price of gold	USD per troy ounce
$C_{B,t}$	One day's closing price of Bitcoin	USD per bitcoin
A	One day's amount of gold we have	ounce
B	One day's amount of Bitcoin we have	bitcoin
S	Sharpe Ratio	\
R	Expected rate of return	\
r	Risk-free rate	\
σ	Standard deviation of rate of return	\
R_p	Earnings expectations of the portfolio assets	\
R_i	Earnings expectations for an individual asset	\
W_i	Investment proportion of the ith asset	\
N	Number of asset types	\
ρ_{ij}	Correlation coefficient between i assets and j assets	\
p	Proportion of funds allocated by gold	\
q	Proportion of funds allocated by Bitcoin	\
k	Volatility (threshold) we set	\
α_{gold}	Commission rate of gold	\
$\alpha_{bitcoin}$	Commission rate of Bitcoin	\
$Commission_t$	Total commission charges to be paid after this distribution	USD
P_t	Evaluation index	\
bin_t	Classification results of evaluation	\

3 Data Analysis

The COVID-19 pandemic has had a serious impact on the financial market. Investors are turning their eyes from risky assets to hedge assets. At the same time, cryptocurrencies represented by Bitcoin and Tether are introduced to the market as new risk averse assets due to their scarcity. We can see from the Figure 1 that since the outbreak of COVID-19 in 2019, the traditional hedge asset gold and the new hedge asset bitcoin have risen to varying degrees. Bitcoin is also called "new gold" by the media. Although it has not formed a consensus in the market due to its short birth time and has great risks, it has been in an upward trend in the long run. Therefore, in the context of global inflation, it is very important to balance the investment

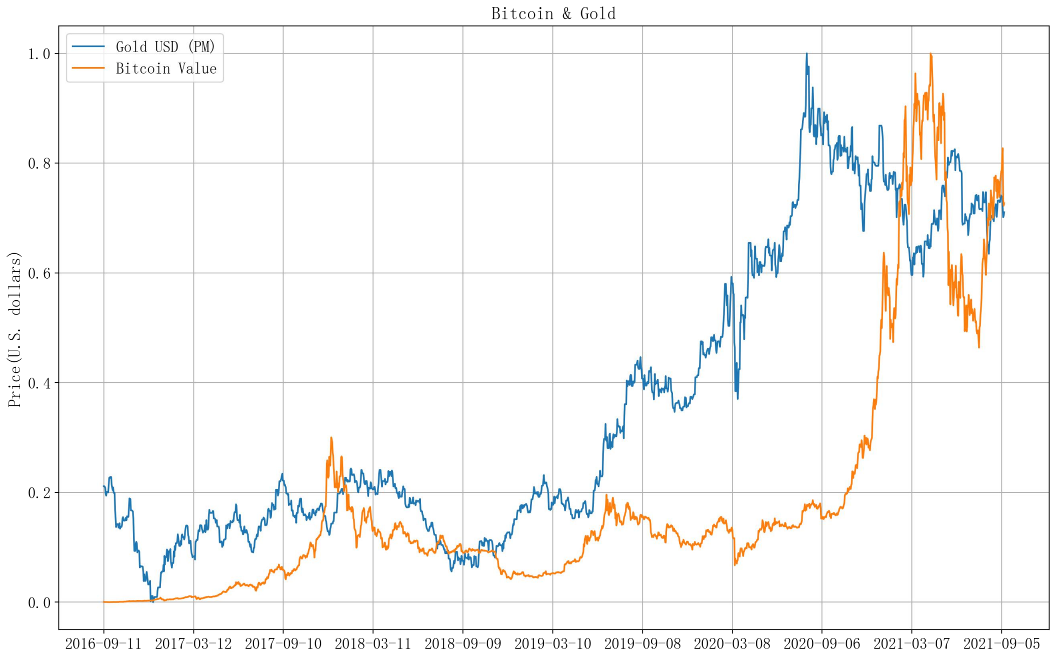

Figure 1: Normalized gold and bitcoin value series curve

of traditional hedge assets and new hedge assets and achieve risk hedging. Through correlation analysis, we calculated that the correlation coefficient between the two, which has a certain correlation. Therefore, we believe that the prediction of the trend of the two can be solved with the same model to a certain extent. And we can assume that with the continuous maturity of the cryptocurrency market, the price trend of bitcoin begins to be similar to that of traditional assets.

4 The Combined Prediction Model based on RF and LSTM

If we want to determine on which day the initial \$1000 will be allocated and how it will be allocated, we need to use the combined prediction model of Random Forest and Long Short-Term Memory (LSTM) to predict the buying situation and closing price change of the next day based on known historical data.

Besides, based on the reason that the use of predictive algorithms to determine the future trend of stock market prices contradicts the basic rules of the efficient market hypothesis of finance (Fama and Malkiel(1970))[4], We can't really predict future advantages based on the data we have. That's because if future advantage can be predicted, subsequent prices will be corrected to prevent this from happening. Therefore, the combined prediction model established here is only limited to the next day's prediction.

4.1 The Random Forest Model

4.1.1 The Introduction of the RF Model

To achieve higher returns in the future, we urgently need to know what to do with our assets that day, whether to buy, sell or hold. So, we choose the RF model and make decisions about the allocation of assets for that day based on what we expect to happen tomorrow.

A random forest, which is a bunch of classification trees classifying the input vectors. Each tree is a classification, and the input vector is "voted" on. The forest is the tree with the most votes. Now we need to find N data in the forest to train. Each training session is replaced from the original data. This sample becomes the growth of the forest tree. Each tree is grown as much as possible without pruning. After each tree is built, all data is run under the tree and proximity is calculated for each pair of cases. If two cases occupy the same end node, their proximity increases by one. At the end of the run, the proximity is normalized by dividing by the number of trees. Proximities is used to replace missing data, locate outliers, and generate a low-dimensional view of the data.

4.1.2 The Constitution of the RF Model

First of all, we define returns as the criterion of growth, as shown in the following formulas:

$$returns_t = \ln\left(\frac{C_{t+1}}{C_t}\right), \tag{4.1}$$

$$directions = \begin{cases} 1 & , returns > 0, \\ 0 & , returns = 0, \\ -1 & , returns < 0, \end{cases} \tag{4.2}$$

Where, the value of direction is 1, 0, and -1 respectively indicating that it can be bought, neither rising nor falling (negligible), and can not be bought.

We select the closing price data of gold and Bitcoin in the first 120 days respectively, and divide them into each column containing the direction of five consecutive days as training data and the direction of the next day as label data for training. After the training of Random Forest classification, we select the direction data of the 120th day and the first four days to predict the direction data of the 121st day, and the subsequent prediction method is the same.

4.1.3 The Prediction Results of the RF Model

Finally, we can use the direction value predicted by random forest and the probability distribution of different values to allocate assets in combination with the expected return rate predicted later. The result of random forest prediction of Bitcoin is shown in Figure 2.

Date	input1_bin	input2_bin	input3_bin	input4_bin	direction
2016-09-18	1	-1	-1	1	-1
2016-09-19	-1	1	-1	-1	-1
2016-09-20	-1	-1	1	-1	-1
2016-09-21	-1	-1	-1	1	-1
2016-09-22	-1	-1	-1	-1	-1
...	...	...	...	...	...
2021-08-29	-1	1	-1	1	-1
2021-08-30	-1	-1	1	-1	-1
2021-08-31	-1	-1	-1	1	1
2021-09-01	1	-1	-1	-1	1
2021-09-02	1	1	-1	-1	1

(a) Predicted direction

Date	input1	input2	input3	input4	returns
2016-09-18	0.007451	-0.003404	-0.002083	0.002559	-0.002275
2016-09-19	-0.002275	0.007451	-0.003404	-0.002083	-0.002511
2016-09-20	-0.002511	-0.002275	0.007451	-0.003404	-0.016199
2016-09-21	-0.016199	-0.002511	-0.002275	0.007451	-0.002441
2016-09-22	-0.002441	-0.016199	-0.002511	-0.002275	-0.005606
...	...	...	...	...	...
2021-08-29	-0.003251	0.043624	-0.041895	0.025385	-0.001860
2021-08-30	-0.001860	-0.003251	0.043624	-0.041895	-0.036132
2021-08-31	-0.036132	-0.001860	-0.003251	0.043624	0.001721
2021-09-01	0.001721	-0.036132	-0.001860	-0.003251	0.035557
2021-09-02	0.035557	0.001721	-0.036132	-0.001860	0.009497

(b) Predicted direction's probability distribution

Figure 2: Prediction results

4.2 Long Short-Term Memory Model

Above, we obtain whether to buy gold or Bitcoin on the specified date through Random Forest prediction. Below, we use the Long Short-Term Memory model[5] to predict the closing price and change trend of gold and Bitcoin on the specified date respectively, which help us to carry out the asset allocation strategies based on portfolio theory.

4.2.1 The Introduction of the LSTM Model

We use LSTM to the prediction of the closing price and trend of gold and bitcoin on specified dates, which is a special Recurrent Neural Network (RNN) model[6] to better solve the problem of gradient extinction and gradient explosion during long sequence training. It is also better suited to the long series of data for analyzing graphic changes in the actual daily closing prices of gold and Bitcoin.

LSTM has a repeating module chain form common to recurrent neural networks, in which four layers interact in a particular way.

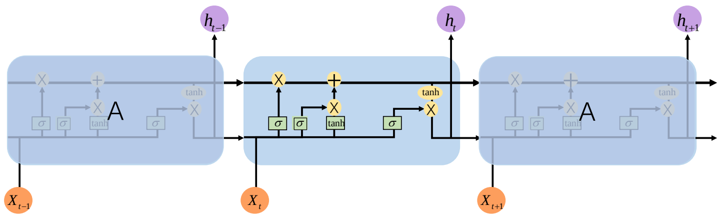

Figure 3: Loop module structure of LSTM

In the figure 3, each row carries a vector that outputs from one node to the input of the other nodes.The pink circles represent point-to-point operations, such as vector addition and dot product, while the green boxes are learning neural network layers. A merge of the lines indicates a connection while a crossing of the lines indicates that their contents are being copied. And the copy will go to a different location.

Then, LSTM goes through three stages. Selectively forgetting the input from the previous node, forgetting the unimportant and remembering the important is "forgetting stage". The "selective memory stage", which inputs from this stage is selectively remembered. And the "output stage" that determines which will be considered outputs of the current state.

The mdel of LSTM remembers the need for long-term memory and forgets the unimportant information, so as to achieve memory processing of long sequence data through the gating state to control the transmission state.

4.2.2 The Constitution of the LSTM Model

We first establish the structure of LSTM, a special circulating neural network. It is followed by the linear activation function Relu, which formula is as follows:

$$f(x) = \max(0, x), \tag{4.3}$$

and a recurrent neural network[7], which together form the model that predicts the closing prices and trends of gold or Bitcoin. As shown in Figure 4.

We divide the closing price data of gold and Bitcoin in the first 120 days after max-minimum standardization into 90 training data, and each training data is composed of the training data of consecutive 30 days and the label data of the next day. Refer to figure 5.

Then, we train 90 pieces of data that we already have. After that, we start with the 120th day and use the 121st day forecast to make asset purchases with the principal. When it reaches the 121st day, the actual closing price of the 121st day is obtained. The new data of the 121st day

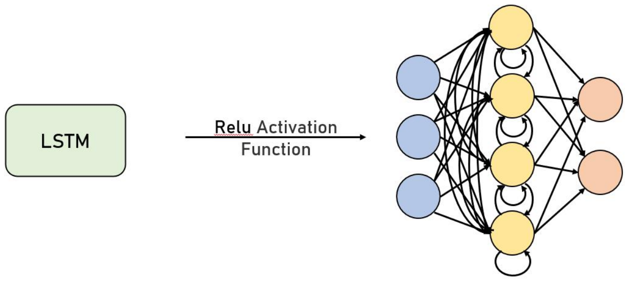

Figure 4: The structure of our model

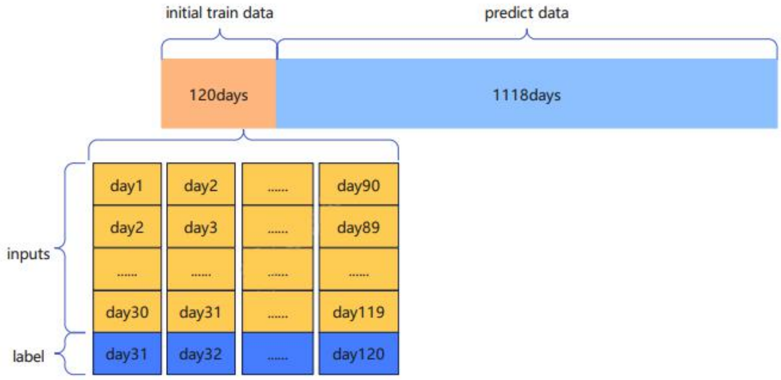

Figure 5: LSTM model's training data description diagram

is used as label data, and the consecutive 30 days before it are used as training data to update the network and make the prediction of the next day. The expected rate of return is predicted by the predicted value and will be used in the asset redistribution strategies.

4.2.3 The Predicted Closing Price and Trend Result of the LSTM Model

The closing price and its change trend predicted on the 121st day and after using LSTM are shown in Figure 6.

We have carried out smoothing processing, data standardization and data denoising processing for many times, and changed different forecast days and data combinations. The results all show different degrees of lag, as shown in the figure above. Considering the reasons that the use of prediction algorithms to determine the future trend of stock market prices is inconsistent with the basic rules of efficient market hypothesis in finance (Fama and Malkiel(1970)), we finally decide to use the data of the previous few days to forcast and calculate the expected rate of return.

4.3 The Summary of the Combined Prediction Model

4.3.1 Clustering Analysis

We predict whether to buy gold or Bitcoin on a specified date and the probability distribution of different directions through Random Forest, which is based on machine learning. We also predict the closing price and its change trend of gold and bitcoin on a specified date through LSTM, which is based on recurrent neural network. So we can calculate the required expected return rate and asset standard deviation data. Based on the prediction results of Random Forest

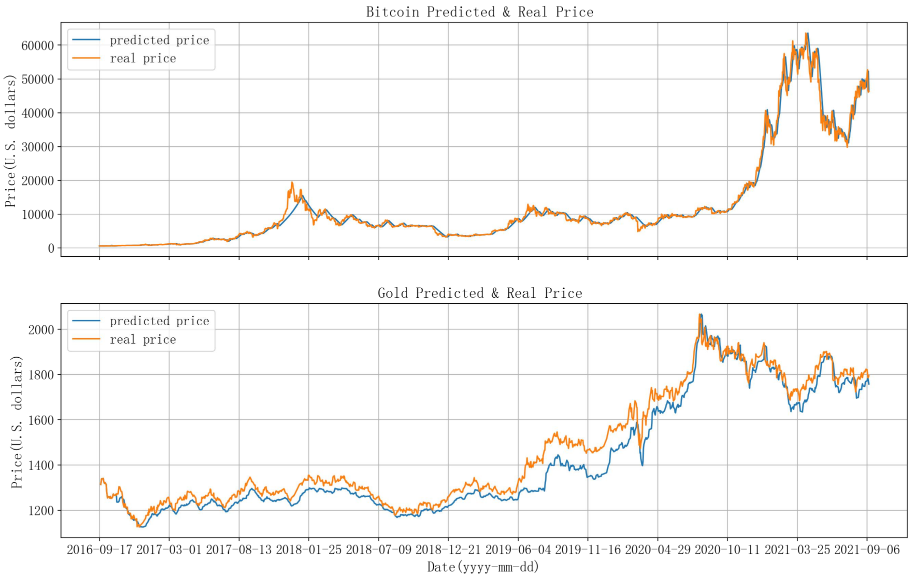

Figure 6: LSTM prediction curve of Bitcoin and gold

and LSTM combined prediction model, a new quantitative investment strategy based on asset portfolio theory will be carried out to find a portfolio allocation method with maximum profit and minimum risk.

5 New Quantitative investment Strategies based On Portfolio Theory

We successfully use the four-month wait and see period to predict whether to buy gold or bitcoin after January 11, 2020, as well as their daily closing prices and trends[8]. We hope to find the characteristic points in line with Sharpe Ratio on January 12, 2020 according to the above predicted closing prices of the two assets of the next day, so as to determine the asset portfolio allocation method that meets the maximum profit and minimum risk. Of course, the same thing happens every day after that.

Therefore, we discuss different cases to determine the allocation strategy of gold and Bitcoin under different processing conditions. In this paper, Monte Carlo method is used to form effective boundary and Sharpe Ratio theory in Modern Portfolio[10] Theory (MPT). The daily optimal investment ratio scheme is determined and the sharpe ratio peak in effective frontier is found based on our new quantitative investment strategies.

5.1 Sharpe Ratio

We want to help investors choose portfolios[9] that maximize expected returns at a given level of risk, or those that minimize risk at a given level of expected return. This is also the core idea of Sharpe's formula, which takes benefits and risks into comprehensive consideration, as shown in the following formula:

$$S = \frac{R - r}{\sigma}, \tag{5.1}$$

5.2 Markowitz Model and Efficient Boundary

Based on the above theoretical basis, we need to introduce Markowitz model in order to achieve the purpose of analyzing the portfolio composed of multiple risky assets. According to the expected rate of return predicted by the LSTM model and under the conditions of our assumptions, the expected return rate of the portfolio can be calculated by the following formula:

$$E\left(R_p\right)=\sum_{i=1}^{N} W_i E\left(R_i\right), \tag{5.2}$$

Among them, due to the non opening time of gold, we will skip the non opening time and connect the opening time data calculation when we calculate.

In addition to returns, we also need to pay attention to risks. The risk measurement formula of the portfolio is as follows:

$$\operatorname{Var}\left(R_p\right)=\sum_{i=1}^{N} W_i^2 \sigma_i^2+\sum_{i=1}^{N} \sum_{j \neq 1}^{N} W_i W_j \rho_{ij} \sigma_i \sigma_j, \tag{5.3}$$

The final risk (standard deviation σ) is:

$$\sigma_p=\sqrt{\sum_{i=1}^{N} W_i^2 \sigma_i^2+\sum_{i=1}^{N} \sum_{j \neq 1}^{N} W_i W_j COV_{ij}}, \tag{5.4}$$

By observing the above formula, we found that the direct of function fitting has the certain difficulty. So we adopt the Monte Carlo method for effective boundary fitting and generate the return rate and standard deviation of 10000 investment proportion portfolio randomly by using this method. The portfolio on the left is connected by a curve to form the effective boundary of the portfolio. The vertex of the curve is the investment plan with the least risk.

5.3 Optimal Portfolio Point M and Commissions

On the basis of Figure 7, we find the most convex Y-axis point on the effective boundary. It is our optimal portfolio point based on the Sharpe Ratio mentioned above, which is point M, as shown in figure 7. We can get the horizontal and vertical coordinates of point M, corresponding to the standard deviation and the expected return rate.

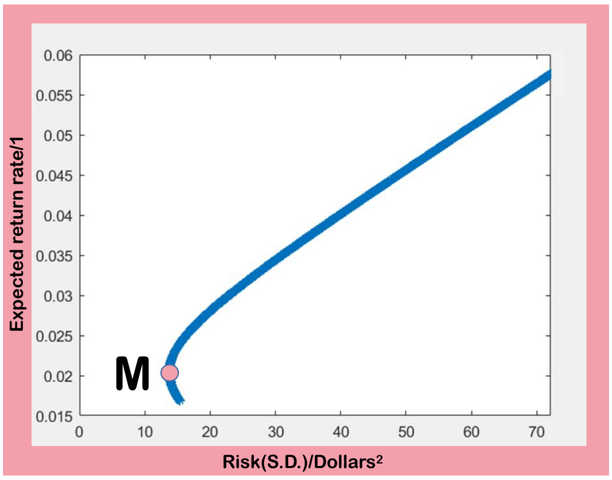

Figure 7: Effective boundary and point M

According to Formula 5.5 , the asset allocation plan of the day can be worked out.

$$\left\{\begin{array}{c} pE\left(R_A\right)+qE\left(R_B\right)=M_y \\ p\sigma_A+q\sigma_B=M_x \end{array}\right. , \tag{5.5}$$

Before truly determining the allocation of assets for the day, a final comparison is made to judge that whether the expected return of the allocation is greater than the commissions spent on purchasing and selling the allocation. We'll discuss this in the following steps.

5.4 The Step of Our New Quantitative Investment Strategies

After accumulating the previous points, we will explain the asset allocation strategies used in detail.

We already have the probability distribution of whether gold and Bitcoin will be bought on a given date, the next day's closing price and its trend and the expected return rate of each asset based on the portfolio prediction model. Using these projections, we redistribute our assets on a daily basis.

The initial values of M_A, M_B, Cash are 0, 0, \$1000 (updated over the course of the investment). We set volatility (threshold) k as 14.08%, which can get the highest final value through experiments. The forecast data are processed to get the corresponding closing prices of gold and Bitcoin at the corresponding times (we set the closing prices of gold to 0 on the days when gold is closed). Here's the actual asset allocation process for a day.

1) We use Random Forest to predict the current direction's value and probability distribution of gold and Bitcoin. Respectively according to the process of Figure 9 to make preliminary judgment and determine the current processing situation of gold and bitcoin.

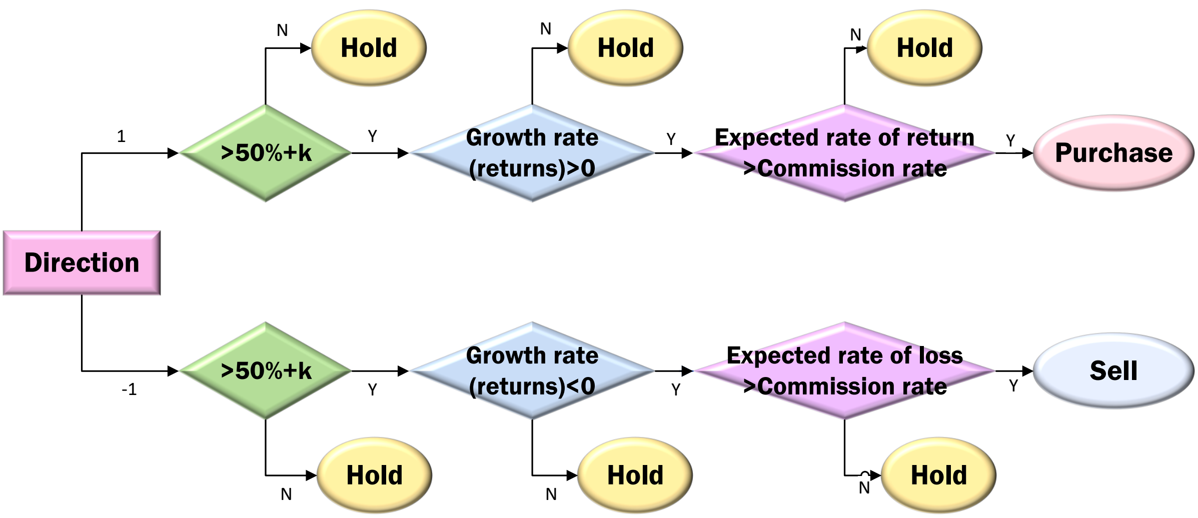

Figure 8: Process for preliminary judgment of gold and bitcoin's management

2) According to the judged buying, holding and selling of gold and Bitcoin on the day, the asset allocation strategy is selected in combination with Table 2 (gold is always held on the date when it does not open).

   Among them, "M" represents using point M to determine the daily allocation method of gold and Bitcoin. "Cash→M_A" means if we have money, we'll use it to buy gold.

Table 2: Select asset allocation strategy

Gold \ Bitcoin	Purchase	Hold	Sell
Purchase	M	$Cash$→M$_A$	M_B& $Cash$→M$_A$
Hold	$Cash$→M$_B$	\	M_B→ $Cash$
Sell	M_A&$Cash$→M$_B$	M_A→ $Cash$	M_A&M_B→ $Cash$

"M_B&Cash→M$_A$" represents continuing to buy gold using the amount of Bitcoin sold and the funds held. "\" indicates that the way of asset allocation is not changed.And the same thing.

3) Allocation is performed using the selected allocation strategy. M_A, M_B, Cash are updated after allocation. Supposed that we choose M's strategy, which determines that the best asset allocation ratio for gold and Bitcoin is p : q. M_A, M_B, Cash update formulas are as follow.

   We divide the updating process into two parts. The first part is to use cash to buy a certain amount of gold and Bitcoin to make their value ratio close to p : q, if there is cash available.The formulas are below.

$$f=\begin{cases}\frac{\frac{q}{p}\times M_A-M_B}{1-\alpha_A}, & \frac{M_B}{M_A}\le\frac{q}{p},\\ \\ \frac{\frac{p}{q}\times M_B-M_A}{1-\alpha_B}, & \frac{M_B}{M_A}>\frac{q}{p},\end{cases}\tag{5.6}$$

$$M_A=\begin{cases}M_A+f\times(1-\alpha_A), & f\le Cash and\frac{M_B}{M_A}>\frac{q}{p},\\ M_A+Cash\times(1-\alpha_A), & f>Cash \text{ and } \frac{M_B}{M_A}>\frac{q}{p},\\ M_A, & \frac{M_B}{M_A}\leqslant\frac{q}{p},\end{cases}\tag{5.7}$$

$$M_B=\begin{cases}M_B+f\times(1-\alpha_B) & f\leqslant Cash\, and\,\frac{M_B}{M_A}\leqslant\frac{q}{p},\\ M_B+Cash\times(1-\alpha_B) & ,f>Cash\, and\,\frac{M_A}{M_B}\leqslant\frac{q}{p},\\ M_B, & \frac{M_B}{M_A}>\frac{q}{p},\end{cases}\tag{5.8}$$

$$Cash=\begin{cases}Cash-f, & f\leqslant Cash,\\ 0, & f>Cash.\end{cases}\tag{5.9}$$

   The second part is to buy or sell gold and bitcoin so that their value ratio is close to p : q. The formulas are below.

$$f=M_B\times p-M_A\times q,\quad Cash=0\,,\tag{5.10}$$

$$M_A=\begin{cases}M_A+Cash\times p\times(1-\alpha_A), & Cash>0,\\ M_A+f\times(1-\alpha_A), & Cash=0\, and\, f\geqslant 0,\\ M_A+\frac{f}{1-\alpha_A}, & Cash=0\, and\, f<0,\end{cases}\tag{5.11}$$

$$M_B = \begin{cases} M_B + Cash \times q \times (1-\alpha_B)\,, & Cash > 0, \\ M_B - \frac{f}{1-\alpha_B}, & Cash = 0\, and\, f \geqslant 0, \\ M_B - f \times (1-\alpha_B)\,, & Cash = 0\, and\, f < 0, \end{cases} \tag{5.12}$$

$$Cash = 0\,. \tag{5.13}$$

4) Repeat steps 1), 2) and 3) on each day of the investment until September 10, 2021 updated to the last gold, Bitcoin and cash's value M_A, M_B and Cash. The final Value is as follows:

$$Value = M_A + M_B + Cash\,. \tag{5.14}$$

Holding the principal of $1000, the daily change of asset value during the investment process is shown in Figure 9.

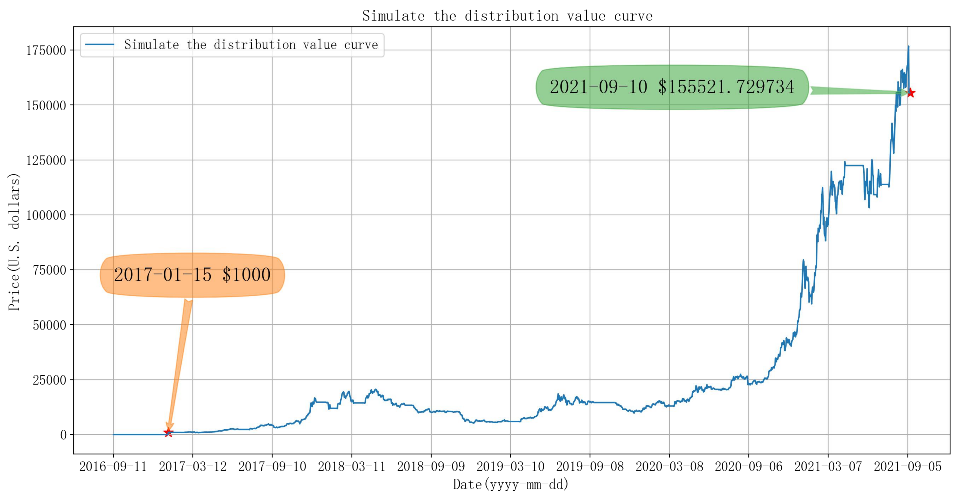

Figure 9: Simulated asset allocation value curve

In the end, we end up with a value of $155521.73 on September 10, 2021 for the initial $1000.

6 Evidences to Prove Our Strategy is The Best

6.1 Evidence 1 : Random Forest

We choose a random forest classification method to predict whether purchasing or not. That's because the Random Forest has effectively run on large data sets, is available to internal generated an unbiased estimate of the error, can evaluate the importance of each characteristic in classification problem and for the default value it will also be able to obtain good results. By observing the daily closing trend of gold and Bitcoin and considering the influence of gold's closed period, we decide to choose the classification method of Random Forest.

We define the evaluation index P:

$$P_t = \exp(\sum_{i=0}^{N} bin_t \times returns_t)\ , \tag{6.1}$$

Of course, we actually use other models to make predictions under the same conditions and the prediction evaluation results of each model are shown in Figure 10 and 12.

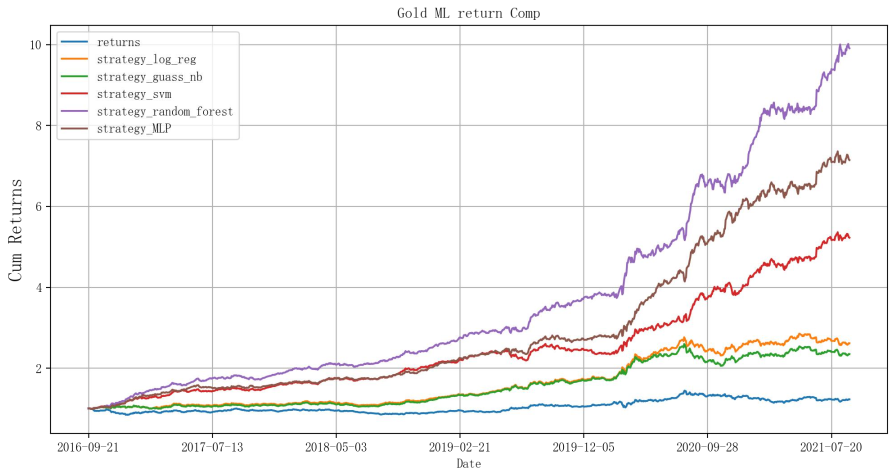

Figure 10: Gold ML return comp

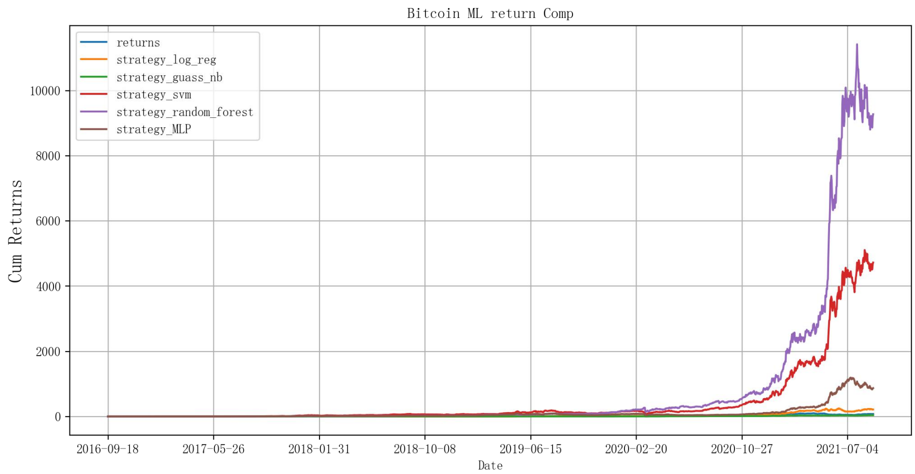

Figure 11: Bitcoin ML return comp

The superposition and sum results of the daily growth rate corresponding to the data with prediction label 1 are used as the evaluation index of model excellence. After comprehensive comparison of multiple classification models such as Random Forest, support vector machine (SVM), neural network (MLPClassifier), GaussianNB and LogisticRegression, it can be seen that Random Forest has excellent performance in both sequences. so we choose it as the final model. And as the classification model part of the combined prediction model, it provides the basis for the optimality of the strategies.

6.2 Evidence 2 : Our New Quantitative investment Strategies

We establish the model of Random Forest to obtain daily forecast whether to buy and volatility (threshold) Settings. Through the portfolio theory, Sharpe Ratio, Markowitz model and effective boundary and other theories prove that it is a better decision making method.

Subsequently, we conduct a planning modeling of strategy optimization for our model with the goal of maximizing the number of transaction return at the end of the trading period. We take the daily trading volume of gold and Bitcoin as our daily decision-making variables and add random factors to simulate random decision-making. In addition, we simulate the results through matlab algorithm and repeat simulations as follows. It can be found that Random Forest, volatility (threshold) set at 0.1408 and effective boundary theory strategy adopted by us are the strategies with the highest return at the end of the transaction, namely, the optimal strategy with the best optimization effect.

1) Monte Carlo simulation of volatility (threshold).

   Through random sowing in the threshold range from 0 to 0.5, we found through multiple simulations that the volatility (threshold) at 0.1408 could achieve the ultimate maximum return.

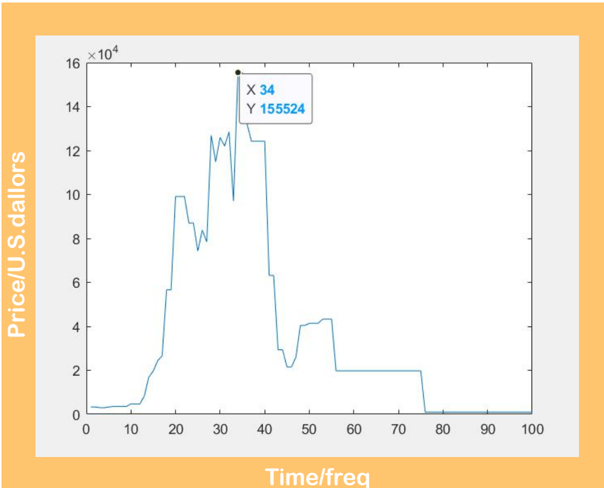

Figure 12: Search for the best volatility (threshold)

   It can be seen from Figure 12 that the volatility (threshold) is highest at about 34. That is 0.1408.

2) Random perturbation of the predicted growth rate in the range of [-10%,10%] (reflecting Random Forest optimization).

   We apply random perturbation in the range of [-10%,10%] to the predicted growth rate, and carry out simulation for 50 times through the optimal strategy defined later. The result shows that the optimal strategy is still the Random Forest strategy.

   From Figure 13 can be seen that with the increase of disturbance, the highest return of the obtained results under each disturbance is not as much as that of our strategy. And as the disturbance increases, the volatility of returns increases, indicating that the risk is getting higher and higher.

3) Randomly select purchase, hold and sell of gold and Bitcoin daily investment strategies (reflecting the optimal Selection of Markowitz model).

   We add random factor probability selection (10%,20%,50%,100%) to make random decision (gold, Bitcoin's trading volume) of gold or Bitcoin. The simulation results are as follows.

   It can be observed in Figure 14 that with the increase of random decision probability, our final income is decreasing. By adding random decision factors, our new quantitative

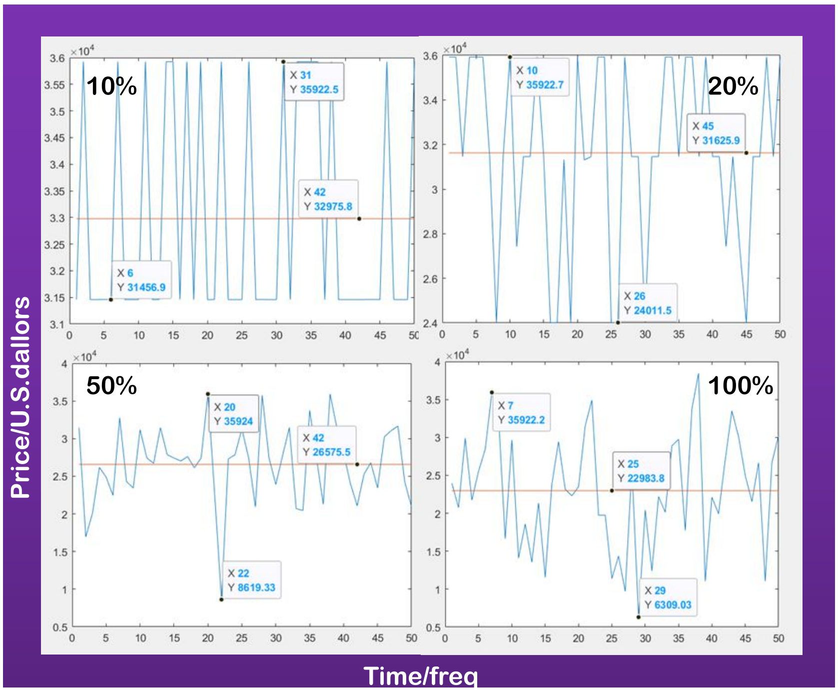

Figure 13: The results of random perturbation of [- 10%, 10%] for the predicted growth rates of 10%, 20%, 50% and 100% respectively

investment strategies are weakened and the effect is also reduced, which further proves that our new quantitative investment strategies are optimal.

7 Sensitivity Analysis

7.1 The Sensitivity of Our Strategies to Transaction Costs

We consider the sensitivity of our strategy to transaction costs fluctuations, which can be divided into the following two situations:

1) In the case that the single transaction cost of gold or Bitcoin is changed and other parameters remain unchanged, the influence of the sensitivity of the strategy to the transaction cost on the trend of the total amount at the end of the 5-year trading period is observed.

The Figure 15 shows when the change range of transaction fees is $\pm 20\%$, the change trend of our different transaction fees on our strategic transaction returns in five years. The red one is our situation when there is no change, and the gray one is the trend line after the change. It can be seen that within this change range, the trend of the red gray line is roughly the same, that is, our transaction decision-making has not been greatly affected. The final transaction return is only a slight change in capital due to the gradual accumulation of handling fees. This clearly shows that our model has good robustness.

And the Figure 16 and 17 show the sensitivity of our strategy to the transaction cost of bitcoin and the change curve of the final result. It can be observed that as the transaction cost of Bitcoin increases, the sensitivity increases for a short period before $\alpha_{bitcoin}$ reaches 0.008(with -20% fluctuation), and then decreases as a whole. The final value fluctuate within a narrow range until $\alpha_{bitcoin}$ reached 0.023(+15%), after which it falls sharply.

It can be concluded that the sensitivity of our strategy to transaction cost is in a certain

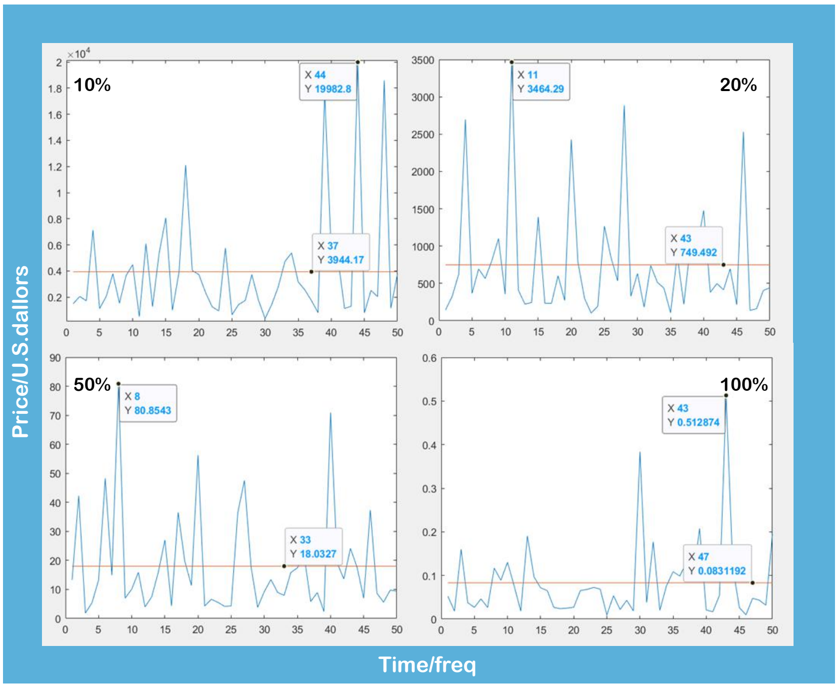

Figure 14: The result diagram of random decision is inserted

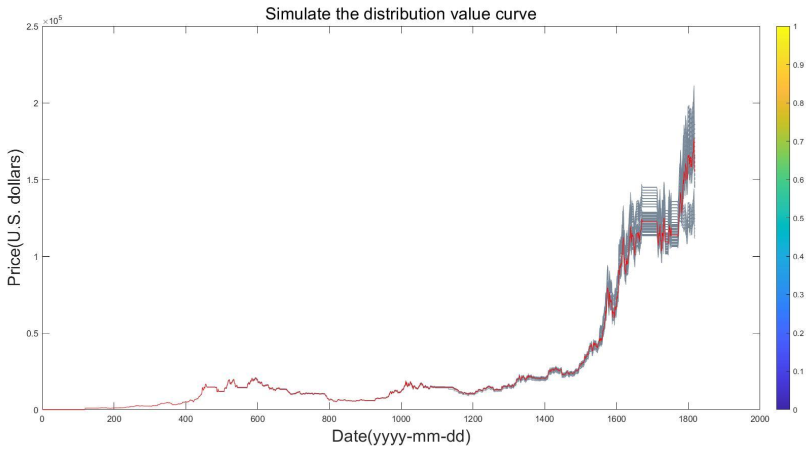

Figure 15: Simulating asset allocation curve

acceptable range and the overall trend is declining with the increase of transaction cost, but maintain a high final value. It shows that our model has a certain robustness, of course, when the fee exceeds the scope, our strategy still needs to be adjusted and changed in time.

The sensitivity of the transaction cost of gold has a similar property, which is not graphed due to space, but the corresponding situation can be seen below.

2) In the case of both gold and Bitcoin transaction costs change and other parameters remain constant, observe their impact on trading returns.

By the Figure 18 and 19, we can see that gold and Bitcoin in charge fluctuation have relatively similar behaviors. With the loss of the procedure rate, the sensitivity increase slightly, but all in the acceptable range or transaction costs under a certain degree of volatility (-20%, 20%). The deal returns still has higher ultimate value, which proves our model stability is higher.

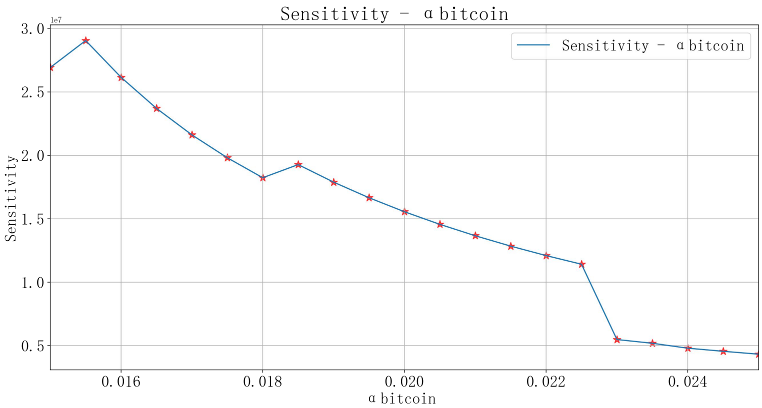

Figure 16: Relationship between transaction cost $\alpha_{bitcoin}$ and sensitivity

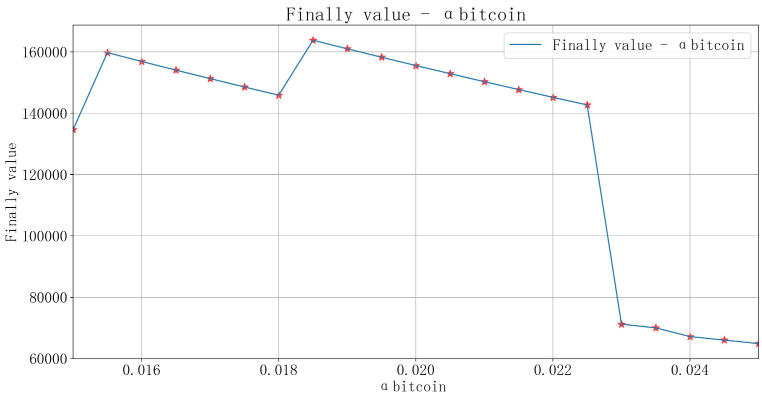

Figure 17: Relationship between transaction cost $\alpha_{bitcoin}$ and final value

7.2 Transaction Costs to Our Strategies and Results

Through the sensitivity analysis of the above our strategy to transaction costs, it can be seen that transaction costs play a certain decisive role in our strategies. When the transaction costs do not exceed the limited range, our strategies can be maintained. When the transaction costs do exceed the limited range, we need to make adjustments and improvements, otherwise, the final value will drop sharply.

It can be concluded that when transaction costs change significantly, our strategies and final result are severely affected. At this time, it is necessary to adjust the decision variables in the strategy such as volatility (threshold), so as to obtain better returns.

8 Model and Strategy Evaluation

Strengths

- A variety of prediction methods are selected for comparison. Finally the Random Forest model with higher accuracy was selected. And we add a new threshold in the Random Forest judgment to reduce the influence of random factors on the daily price.
- Both neural network and machine learning methods are used to make the results more reliable.
- Through the combination of risk and return analysis, it can better avoid the situation that

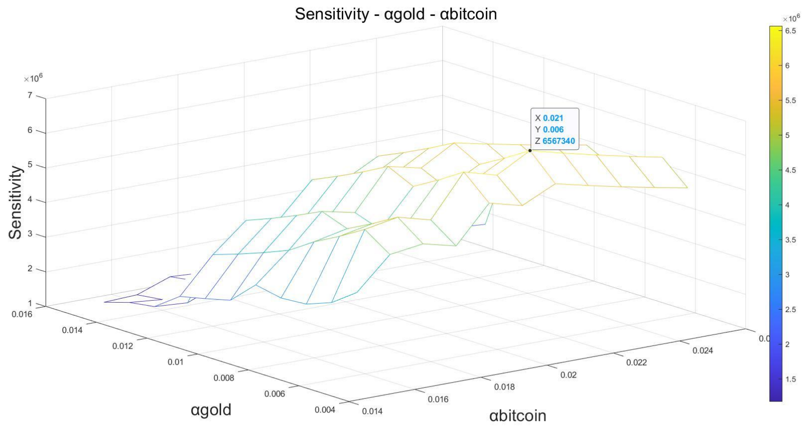

Figure 18: Relationship between transaction costs $\alpha_{bitcoin}$, α_{gold} and sensitivity

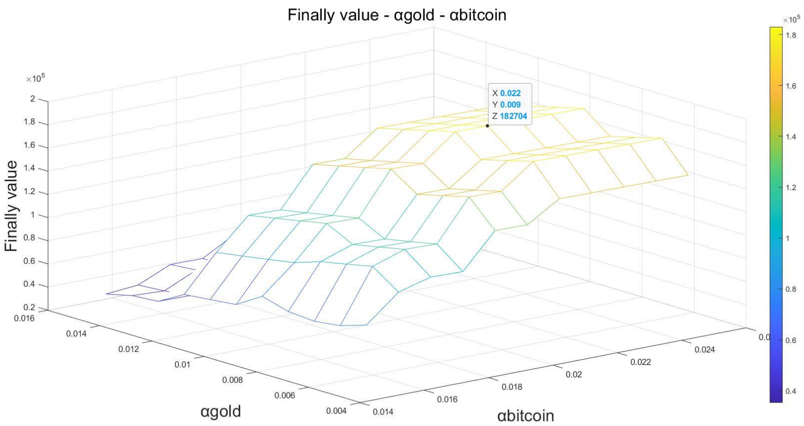

Figure 19: Relationship between transaction costs $\alpha_{bitcoin}$, α_{gold} and final value

the risk is too high and the return is easy to decrease. So it can get the trading return in 5 years has a trend of rising safely.

- When buying gold and Bitcoin, we consider the method of effective portfolio to realize the ratio strategy of gold and Bitcoin with low risk and high return.
- The parameters of the model are optimized and verified and the sensitivity of some parameters is analyzed.

Weaknesses

- The LSTM has underfitting and lag phenomenon.
- At present, the purchase and sale only forecast the data of the next few days, and the analysis of the long-term trend is not perfect.
- The analysis factor is too single and only has the value sequence of gold and Bitcoin, which is difficult to analyze from multi-dimension.
- Due to the limitation of hardware and time, the testing and tuning of neural network are not perfect.

Improvements

- Add daily trading volume, dollar price and other factors, multidimensional analysis.
- Take the Fourier transform of the sequence, transform it into the frequency domain for analysis.
- Improve the LSTM model to increase the self-attention mechanism and reduce the overfitting phenomenon and make the long-term prediction of the model to obtain the increasing and decreasing trend of the sequence
- Make quantitative analysis of the trend of increase and decrease to make the amount allocation more reasonable.

Memo

To: Trader

From: Team 2205329

Subject: The Summary of Our Trading Model, Strategy and Result

Date: Monday, February 21, 2022

We are amused to have the opportunity to build a trading model, which uses the past stream of daily prices so far to determine whether the portfolio is purchased, held or sold on a daily basis, resulting in higher returns. We have the principal of $1000 on September 11, 2016. And we are confident that our trading model can provide us with a reasonable trading plan for gold and Bitcoin for a total of five years until September 10, 2021 to achieve higher returns.

Our model and strategy are divided into two parts, namely the prediction part and the asset allocation part.

1) Prediction part In this part, we build a combined prediction model based on Random Forest and Long-Short Term Memory. They use the daily closing prices of gold and Bitcoin in the first 120 days as training data and different training data composition types for training. After that, they use the new training data to update themselves every day to achieve better prediction effect.

- The model of Random Forest We take advantage of Random Forest's better classification characteristics and use the classification result (-1, 0, 1) to predict whether to buy on that day. Of course, we can also get the probability distribution corresponding to the corresponding situation. It becomes the discriminant condition in the process of asset allocation in the next part.

- The model of LSTM We use the LSTM model to predict the subsequent daily closing prices of gold and Bitcoin to obtain the expected daily returns. Using the data smoothing and denoising processing, the model still show different degrees of hysteresis. And comprehensive consideration based on the prediction algorithm is used to determine the future trend of stock market prices and finance of the efficient market hypothesis of the basic rules of contradictory reasons, we finally decide to use a few days ago data to calculate the expected return. It also provides the basis for the following asset allocation strategies.

2) Asset allocation part We first understand and learn the Sharpe Ratio, expected rate of return, Markowitz model and effective boundary and other theoretical knowledge to determine the goal to obtain the maximum profit and minimum risk of the portfolio. The following asset portfolio allocation strategies have been developed:

- According to the Random Forest model obtained (-1, 0, 1) and its probability distribution and LSTM model obtained expected return rate and gold Bitcoin procedure rate comparison preliminary according to the flow chart for the day of gold and Bitcoin processing.

- According to the combination of gold and Bitcoin with different processing conditions, the corresponding asset allocation strategy is selected in the allocation strategy table to reallocate gold, Bitcoin and cash.

- Update the corresponding value of gold, Bitcoin and funds for each period of time by using the update formula we provide for gold, Bitcoin and funds.

- Repeat the above steps until the final date for the final excellent return. Later, we prove the optimality of the combined use of our model and strategy from different aspects.They include:

- The evaluation index P is used to comprehensively compare Random Forest, SVM, MLPClassifier, GaussianNB, LogisticRegression and other classification models. The results show that the random forest performs well on both sequences.

- Programming optimal decision, adding random factors to simulate random decision. Respectively, the volatility (threshold) in the setting judgment conditions is randomly sown, the predicted growth rate is randomly disturbed in a certain range and random factors are added to randomly select the allocation strategy combination. The results of the above Random Factors all reduce the final value to some extent, which provides evidence for the optimality of our model and strategy.

As for the sensitivity of transaction cost to our model, we can get that our model can still get a higher final transaction return value when the transaction cost changes by [-20%, 20%] through program simulation. It can be concluded that our model has high model robustness in transaction costs. Of course, when the change range of transaction cost is large, we need to adjust the strategy, That is, we can achieve better returns by adjusting decision variables such as volatility (threshold at Random Forest).

References

[1] E. Livieris, N. Kiriakidou, S. Stavroyiannis, et al., An advanced CNN-LSTM model for cryptocurrency forecasting. Electronics, 2021, 10: 287.

[2] C.P. Román, J.L.C. Pérez, P.G. Estévez, Portfolio theory application with gold numismatic assets and precious metal, Cuadernos De Gestión, 2012 123-143.

[3] E. Platanakis, A. Urquhart, Should investors include bitcoin in their portfolios? A portfolio theory approach. Social Science Electronic Publishing. The British accounting review, 2020, 52: 100837.

[4] L. Khaidem, S. Saha, R. Dey, Predicting the direction of stock market prices using random forest. 301-322.

[5] E. Bouri, R. Gupta, A. Lahiani, et al., Testing for asymmetric nonlinear short-and long-run relationships between bitcoin, aggregate commodity and gold prices. Resources Policy, 2018, 57: 224-235.

[6] A. M. Ramadhani, K. N. Rang, L. Hun, et al., Bitcoin price forecasting using neural decomposition and deep learning. Journal of the Korea Industrial Information Systems Research, 2018, 23: 81-92.

[7] T. Awoke, M. Rout, L. Mohanty, et al., Bitcoin price prediction and analysis using deep learning models. Communication Software and Networks. Springer, Singapore, 2021: 631-640.

[8] S. Long, H. Pei, H. Tian, et al., Can both bitcoin and gold serve as safe-haven assets? — A comparative analysis based on the NARDL model. International Review of Financial Analysis, 2021, 78: 101914.

[9] D. Lv, S. Yuan, M. Li, et al, An empirical study of machine learning algorithms for stock daily trading strategy. Mathematical Problems in Engineering, 2019, 2019:1-30.

[10] S. Ji, J. Kim, H. Im, A comparative study of bitcoin price prediction using deep learning. Mathematics, 2019, 7: 898.

Appendices

Here are simulation programmes we used in our models as follow.

Preprocess the training data.

```
def getData(df, column, train_end=90, days_before=30, days_pred=1,   \
return_all=True, generate_index=False):
    series = df[column].copy()
    data = pd.DataFrame()
    for i in range(days_before):
        data['b%d' % i] = series.tolist()[i: -days_before - days_pred + i]
    for i in range(days_pred):
        data['y%d' % i] = series.tolist()[days_before + i: - days_pred + i]
    return data, series
db_data, db_series = getData(db[100:], 'bvalue', days_before=DAYS_BEFORE,   \
days_pred=DAYS_PRED, train_end=TRAIN_END)
def Norm(a, b): return (a-np.min(b))/(np.max(b)-np.min(b))
```

Preprocess training data deep learning network (LSTM * 2 + relu + RNN).

```
class Net(nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        self.lstm = nn.LSTM(
            input_size=1,
            hidden_size=64,
            num_layers=2,
            batch_first=True)
        self.out = nn.Sequential(nn.ReLU(), nn.Linear(64,1))
    def forward(self, x):
        r_out,h1 = self.lstm(x,None)
        out = self.out(r_out[:, -4:, :])
        return out
```

Training and prediction of machine learning model.

```
models = {
    'log_reg': linear_model.LogisticRegression(),
    'guass_nb': GaussianNB(),
    'svm': SVC(),
    'random_forest': RandomForestClassifier(max_depth=10,n_estimators=100),
    'MLP':MLPClassifier(max_iter=100)}
def fit_models(data):
    mfit = {model: models[model].fit(data[cols_bin], data['direction'])   \
        for model in models.keys()}
def derive_positions(data):
    for model in models.keys():
        data['pos_'+model] = models[model].predict(data[cols_bin])
        if model == 'random_forest':
            t = trans1(models[model].predict_proba(data[cols_bin]))
            data['pp+1_'+model] = t[2]
def evaluate_strats(data):
    global strategy_rtn
```

```
    strategy_rtn = []
    for model in models.keys():
        col = 'strategy_' + model
        data[col] = data['pos_'+model]*data['returns']
        strategy_rtn.append(col)
    strategy_rtn.insert(0, 'returns')
```

Training and prediction of machine learning model. When both decide to buy at the same time, quantitative portfolio investment is used to formulate allocation strategy.

```
for j=1:N
    mengte=rand;
    Mengte=[Mengte,mengte];
    come=corrcoef(a(time-29:time),bbb(time-29:time));
    t2 = alvt*mengte+blvt*(1-mengte)
    t1 = afangchat*mengte^2+bfangchat*(1-mengte)^2
    t1 = t1+2*mengte*(1-mengte)*come(1,2)*sqrt(afangchat*bfangchat)
    t1 = sqrt(t1)
    yxedge=[yxedge;t1,t2]; end
redmark=find(yxedge==min(yxedge(:,1)));
p = Mengte(redmark); q = 1-p;
if am/bm <= p/q
    t = (p/q*bm-am)/(1-alfaa);
    if t <= mm
        mm = mm - t;
        am = am + t*(1-alfaa);
    elseif t > mm
        am = mm * (1-alfaa) + am;
        mm = 0; end
elseif am/bm > p/q
    t = (q/p*am-bm)/(1-alfab);
    if t <= mm
        mm = mm - t;
        bm = bm + t*(1-alfab);
    elseif t > mm
        bm = mm * (1-alfab) + bm;
        mm = 0;
    end
end
if mm > 0
    am = am + mm*p*(1-alfaa);
    bm = bm + mm*q*(1-alfab);
    mm = 0;
elseif mm == 0
    t = am*q - bm*p;
    if t >= 0
        am = am - t / (1-alfaa);
        bm = bm + t * (1-alfab);
    elseif t < 0
        am = am - t * (1-alfaa);
        bm = bm + t / (1-alfab);
    end
end
```