Problem Chosen	2022	Team Control Number
C	MCM/ICM Summary Sheet	2217563

To be the gold: A Trading Strategy Combining Returns and Risks

Summary

Assets price prediction and auto-trading model is becoming an essential part in traders' work. Therefore, our team is asked to develop a model that uses only the past stream of daily prices to date to determine the trading strategy in each day.

For the first task, we develop a **Three-Stage prediction model** innovatively. According to the volume of accessible data, we use **Random Walk**, **Grey Model** and **Random Forest Regression** in sequence for time series prediction**.** The results show that the MAE of gold and bitcoin calculated by our model are 16.17 and 968.09 respectively. Then, **a decision model** is constructed to **maximize the return** and **minimize the risk**. Referring to **Markowitz's Mean-Variance theory,** we regard the variance of an asset's price over the past 30 days as its risk. Also, we incorporate a **risk aversion factor** to improve our model. We find that the larger the factor is, the less profitable our strategy is.

For the second part, we analyze the advantages of our strategy. **As for the prediction precision, our model outperforms others** like XGBoost, SVR and ARMA in all evaluation indicators. Also, we calculate the MAE when choosing a various range of parameters in our model, showing that the parameters we choose in task 1 are just suitable. **As for the daily strategy,** we draw a picture to visualize the daily position and daily value of cash, gold and bitcoin. Results show that our strategy can easily follow the trend of asset prices and avoid the extreme fluctuation. It is lucrative while taking risks into account. **The ultimate position of our strategy is $11553.**

With regard to section three, we study the impacts of the transaction costs on our model from two aspects. Firstly, we slightly **modify the transaction cost** but keep them the same in all five years. Results show that our model is still profitable but the return is influenced, and our model is more sensitive to the transaction cost of bitcoin than which of gold. Secondly, we further consider the **tax** rate in real life. By equally transforming the tax rate into the transaction costs, we keep the costs changing **in every year** according to the value. Our model is more sensitive to transaction costs if the risk aversion factor is set at a low level.

In terms of sector four, we test the sensitivity of our Three-Stage prediction model to the sliding window. By calculating the MAE of the same model but using different sliding windows, we ensure that sliding window has little impact on our model.

Last but not least, we summarize our strategy, model, and results and write a memorandum to the trader.

Keywords: Three-Stage Prediction Model; Mean-Variance Theory; Assets Price Forecasting

Contents

1 Introduction... 3

1.1 Problem Background... 3

1.2 Restatement of the Problem.. 3

1.3 Our Work... 3

2 Assumptions and Justifications.. 5

3 Notations... 5

4 Model Preparation.. 6

5 Trading Strategy Model.. 6

5.1 Three-Stage Price Forecasting Model.. 6

5.2 Decision Model.. 10

5.3 Model Results.. 12

6 Why Our Model.. 13

6.1 Optimality of Prediction model.. 13

6.2 Optimality of Decision model.. 17

7 Discussion of Transaction costs.. 18

7.1 Trading commission... 18

7.2 Capital Gains Tax... 19

8 Sensitivity Analysis.. 20

9 Model Evaluation.. 21

9.1 Strengths.. 21

9.2 Weaknesses.. 22

10 Memorandum.. 22

References... 25

1 Introduction

1.1 Problem Background

Quantitative trading has become a common and effective way in today's financial market. It uses quantitative methods to buy or sell certain kinds of financial assets at a specific time in order to maximize people's profit return. It can also help traders make decisions about the uncertain future[1].

Quantitative trading strategy is sought after by traders in the market because of its efficiency and accuracy. Meanwhile, quantitative trading is supported by the principles of machine learning and statistical inference, so it is also the research of focus of data scientists, statisticians and computer scientists.

Gold and bitcoin are two popular assets for investors because of their high value[2]. However, due to high volatility and risk, the prices of two assets are difficult to predict, which brings challenges to trading strategy in market. This is exactly the problem that traders are committed to solving.

1.2 Restatement of the Problem

In this problem, our team is asked to develop a model to determine the trading strategy according to the data given in the attachment. More specifically, we will solve the following problems:

- Establish a model to configure the trading strategy of the day only according to the previous price trend.
- Use our model to discuss how much benefit we can ultimately get from investing $1000 starting on September 10, 2016.
- Use evidence to prove the optimality and efficiency of our model.
- Discuss how transaction costs affect our final strategy and results.
- Write a memo to inform traders of our work and results.

1.3 Our Work

According to the requirements of the topic, our work is divided into three parts: establishing daily trading strategy, analyzing the advantages of model and strategy, and exploring the impact of trading rate. Our working mode is shown in the figure below: firstly, we establish the prediction model and trading strategy for problem 1, and then analyze the optimality of the model from the perspective of prediction accuracy, trading profit and trading rate. The specific work contents are as follows:

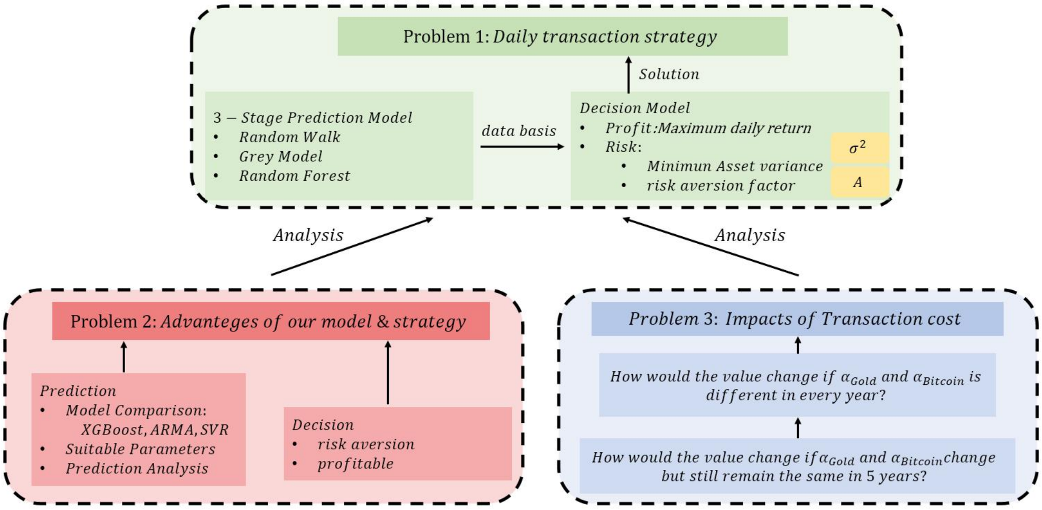

Figure 1. Model Overview

1. When establishing the price prediction model, we first put forward the Three-Stage Prediction Model. According to the amount of known information, we can judge the current stage: lack of data stage, have a small amount of data stage and have some data stage. According to the different characteristics of the three stages, we use Random Walk、 Grey Model and Random Forest Regression to predict the prices of the two assets in the next day, so as to provide a data basis for the Decision Model.

In the Decision Model, we refer to Markowitz's Mean-variance model, comprehensively consider maximizing returns and minimizing risks, and introduce the degree of risk aversion to simulate the investment decisions of investors with different preferences. Finally, the daily trading process is simulated on the whole data length to complete the construction of trading strategy.

2. When analyzing the optimality of the model, we consider the prediction results and decision results respectively. When evaluating the optimality of prediction, we compare the performance score of Random Forest Regression with that of other machine learning models, and find that the accuracy of the model is high.

At the same time, by calculating the MAE predicted by the model under different characteristics, it can be seen that the parameters are at the optimal level. In addition, we carefully selected the parameters: by drawing the position change and value change images of each asset on each trading day, we can see that our model can well capture the change trend, smooth the price curve, and obtain rich reports on the premise of fully considering the risk.

3. When considering the sensitivity of the model to transaction cost, we select the medium risk aversion level and consider the sensitivity of the model to transaction cost. Firstly, we change the two transaction commissions appropriately to get the change degree and direction of the final income. Secondly, we consider the impact of capital gains tax rate in reality, add the annual capital gains tax to the transaction commissions of that year, comprehensively consider the impact of the two factors, and quantify the final income of investors with different risk aversion.

2 Assumptions and Justifications

- **We assume that the changes of dollar inflation during the 5-year trading period can be ignored.**
  In this task, we are required to use only the attached past price data to develop the trading strategy. A key measure of our strategy's superiority is the final value of the asset portfolio, which is calculated in dollars obviously. If the value of the dollar fluctuates, it is difficult to assess the strategy portfolio. So we think the inflation can be ignored.

- **We assume that the transaction process is always successful and unhindered by external obstacles.**
  We develop the strategy based on past price data, including the real price of the day. After that we will carry out the day trading according to the plan. Since we are concerned with the end result of the trade, we do not need to consider trade failure in our model. That is, we don't have to care about the transaction process.

- **We assume that traders are always averse to the trading risks in the market**
  In real life, most normal people are averse to the risk of the unknown things. This is why many researchers conduct extensive research on financial risk. So it is necessary for us to avoid market risks. We fully consider the impact of risk when making strategy.

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description	Unit
C_t	Quantity of dollars at time t	USD
G_t	Quantity of gold at time t	Ounce
B_t	Quantity of bitcoins at time t	-
P_t^G	Gold Price at time t	USD
$\hat{P}_t^G$	Forecast price of gold at time t	USD
σ^2	Portfolio variance	-
A	Degree of risk aversion	-
α_{gold}	Transaction Commission of Gold	
α_{bit}	Transaction Commission of Bitcoin	

4 Model Preparation

The annex gives us the real price data of gold and bitcoin from September 2016 to September 2021 respectively. However, we found 10 vacant bars of gold price data in the dataset (e.g. 23/12/16). This could be the result of a recording error. Considering the continuity of gold prices, for the missing values, we use the average of the two valid price data before and after as a filler to complete our dataset.

5 Trading Strategy Model

According to the question, we can only use the price data before the decision date to predict the future price direction, not as God to use all data of 5 years to do the construction of the model. This poses a challenge for our decision forecasting.

In order to better predict the future price, we first design the Three-Stage Price Forecasting Model. According to the time difference between the decision date and the start date (2016.09.11), it can be divided into three stages. Random Walk, Grey Model and machine learning methods are used for each stage. Then, we design a Decision Model for strategy selection, determine the objective function and constraint conditions. According to the results of the price forecasting model, we can get the next day's asset trading strategy. Finally, using the data set given in the attachment, we calculate the final value of $1000 on September 10, 2021.

5.1 Three-Stage Price Forecasting Model

In similar problems such as gold price forecasting, bitcoin price forecasting, and stock market forecasting, people often use LSTM neural network models[3] , ARMA time series models[4] and other methods[5][6] to explore the trend of future price. In our task, we are going to develop a strategy every day from DAY1. Therefore, a single method is not suitable for an effective forecasting.

For example, when we are at DAY 10, the too little information makes it hard to support any advanced algorithm such as machine learning. While at DAY 1000, past prices may show sufficient information, which may support neural network techniques. Therefore, one inspiration is: using different methods at different stages. We will discuss it further below.

5.1.1 Stage Division

We are asked to trade from September 11, 2016 to September 10, 2021. The total number of days is 1826. We believe that the information volume brought by past prices is the basis for our choice of model. However, we do not know how much information is available to support the most suitable model. So we divide the total trading time in a ratio of 1:9 based on experience. For the first 10% of the time, we call the first 30 days **Stage I** while the rest of it is **Stage II.** For the latter 90% of the time, we call it **Stage III.** In this kind of division, we can use different methods for different stages.

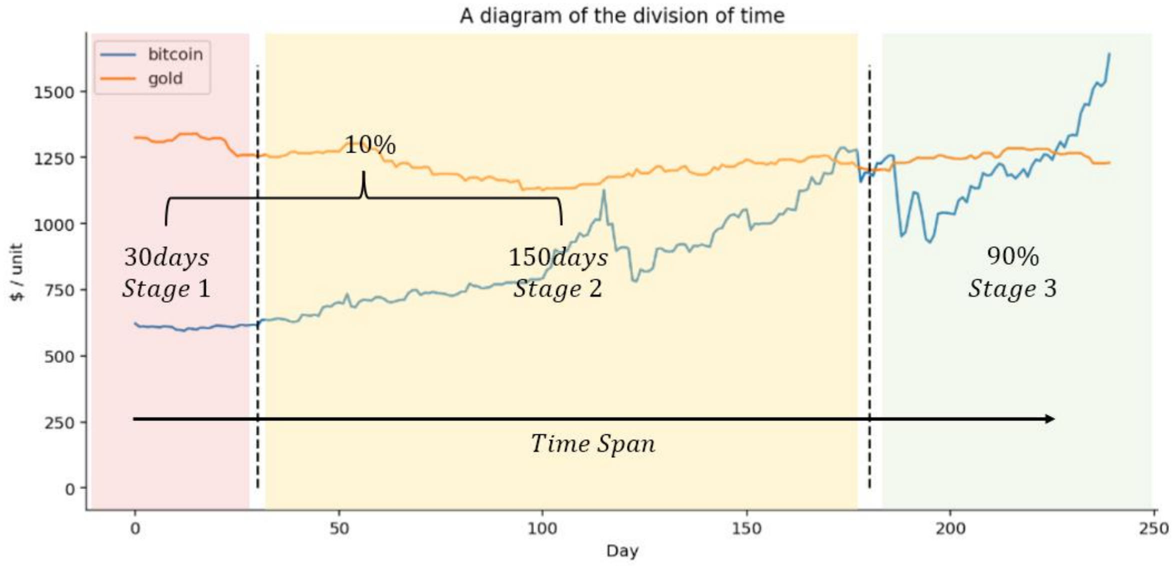

Figure 2. Division of time

5.1.2 Stage I: Random walk

At this stage, it is difficult for us to use some algorithms and models to make predictions because of the small amount of data we can refer to. So, we use **Random Walk** method to solve it.

In the continuous sequences to which this method applies, each sequence is the value of the previous sequence increased by **a randomly perturbed value**, so we can call it "Walking". The sequence values generated in this way have a certain degree of randomness which matches the volatility characteristics of assets such as gold and bitcoin.

More specifically, our operation for the price forecast of two assets during this stage is shown below:

- When we are at DAY1, the forecasting price of DAY2 $\hat{P}_2$ equals to the true price of DAY1 with a perturbation ε following a normal distribution. That is:

$$\hat{P}_2 = P_1 * (\varepsilon + 1),\ \varepsilon \sim N(0, 0.01) \tag{1}$$

- When we are at DAY t, calculate the standard deviation σ of Sequence $[P_1, P_2, \cdots P_t]$. Then we Add $-\sigma$ or $+\sigma$ randomly to today's true price P_t. The result is the forecasting price of DAY $t+1$ $\hat{P}_{t+1}$.
- When we are at weekends or holidays, we do not make predictions on gold because there is on gold trading on that day.

5.1.3 Stage II: Grey Model

Grey Model (GM for short) can develop fuzzy long-term descriptions of development of things by building grey differential prediction model **with a small amount of incomplete information.** According to the basic principles of Grey Model,

- The nature of incomplete information is absolute
- The basic law of the gray model is that the system information is incomplete

- The purpose of the gray model is to fully exploit and utilize the available information

This is in line with our expectations for Stage II. Therefore, we choose Grey Model as the base method of this stage.

The core equation of GM is as follows as shown below[7].

Let the number of days of the reference time series be n , we can get continuous sequence:

$$x^{(0)}=(X^{(0)}(1),X^{(0)}(2),X^{(0)}(3),\cdots,X^{(0)}(n)) \tag{1}$$

Do an accumulation to generate the (1-AGO) sequence:

$$x^{(1)}=(X^{(1)}(1),X^{(1)}(2),X^{(1)}(3),\cdots,X^{(1)}(n)) \tag{2}$$

Where

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i)\ \ k=1,2,\cdots,n \tag{3}$$

Let $Z^{(1)}$ be the generated sequence of Immediate neighborhood mean of $X^{(1)}$

$$Z^{(1)}=(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)) \tag{4}$$

Where

$$z^{(1)}(k)\ =\ 0.5x^{(1)}(k)\ +\ 0.5x^{(1)}(k-1) \tag{5}$$

Model the grey differential equation of GM(1,1) as

$$x^{(0)}(k)\ +\ az^{(1)}(k)\ =\ b \tag{6}$$

Where a is Development factor and b is amount of grey effect.

Establishing the shadow equation for the gray differential equation:

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b \tag{7}$$

The solution is:

$$\hat{x}^{(1)}(t)=\left(x^{(1)}(0)-\frac{b}{a}\right)e^{-at}+\frac{b}{a} \tag{8}$$

Then the time response series of the grey differential equation is

$$\hat{x}^{(1)}(k+1)=\left(x^{(1)}(0)-\frac{b}{a}\right)e^{-at}+\frac{b}{a} \tag{9}$$

Set $x^{(0)}(1)\ =\ x^{(1)}(0)$, through accumulation and reduction, we get the final prediction equation:

$$\hat{x}^{(0)}(k+1)\ =\ \hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)\ =\ \left[x^{(0)}(1)-\frac{b}{a}\right](1-e^{a})e^{-ak} \tag{10}$$

Where $k=1,2,...,n-1$

5.1.4 Stage III: Random Forest Regression

In this stage, we believe that past price data at this time is more informative. So we started to consider using some advanced algorithms.

Random Forest Regression model is powerful and accurate. It is widely used in the field of asset forecasting and has achieved good results. Jigar Patel explore Artificial Neural Network, Support Vector Machine (SVM), Random Forest(RF), naive-Bayes and the ability of their portfolio model to predict the movement of financial assets[8] . The result shows **Random Forest outperforms** other three prediction models on overall performance.

So we plan to choose Random Forest Regression model at this stage to predict the price of gold and bitcoin of the next day.

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble model combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.

The algorithm is as follows:

- Pick at random k data points from the training set.
- Build a decision tree associated to these k data points.
- Choose the number N of trees you want to build and repeat steps 1 and 2.
- For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

In our task, we believe that data that are too early have no impact on our predictions today. For example, when we are at DAY1000 and we want to predict the price of DAY1001, the price of DAY10 has little connection with the prediction. At the same time, our past prices are a strong time series data. So we introduced the concept of **sliding windows.**

At DAYt, we take the previous n days price data as a sliding window. We put the data from sliding window into the model. Then we can predict prices for the next m days.

Our operation is shown in the figure below.

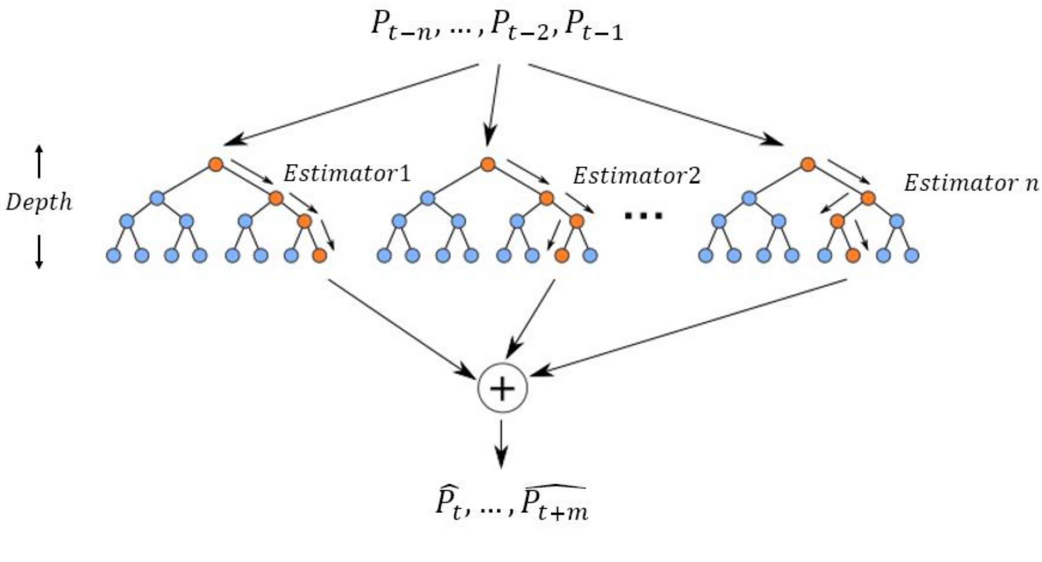

Figure 3. Diagram of RFR

The $[P_{t-n}, \cdots P_{t-1}]$ in the graph is a sliding windows with length n.

5.2 Decision Model

Suppose it is now DAY t . In order to achieve the "best" trading strategy, we need to forecast the next day's prices and plan our asset trades. First, we calculate tomorrow's predicted price $\hat{P}_{t+1}^{G}$ and $\hat{P}_{t+1}^{B}$ based on the Three-Stage prediction model. Second, we build a nonlinear programming model based on **Risk** and **Return.** The specific process is as follows.

5.2.1 Determination of Decision Variables

Our initial portfolio today is $[C_t, G_t, B_t]$. And The given price today is P_t^G and P_t^B . To reach our goal, we make the asset portfolio become $[C_{t+1}, G_{t+1}, B_{t+1}]$ by buying and selling gold, bitcoin. Therefore, our decision variables are the dollar, gold and bitcoin holdings obtained at the end of today's allocation, that is C_{t+1}, G_{t+1} and B_{t+1}.

5.2.2 Determination of Optimization Goal

- **In term of Return**

First, a rational person will always seek to maximize his or her own interests. **Therefore, our goal is to make the portfolio of assets after today's trading session can be maximized at the metric of tomorrow's predicted price**. That is

$$\max\ C_{t+1} + G_{t+1} * \hat{P}_{t+1}^{G} + B_{t+1} * \hat{P}_{t+1}^{B} \tag{1}$$

- **In term of Risk**

In investment activity, risk and return follow each other: high-yielding assets tend to be accompanied by high risk, and low-risk assets usually result in lower levels of return.

According to the well-known Markowitz portfolio theory, diversification maximizes investment returns while minimizing risk. **We introduce the "mean-variance model" in this theory to quantify the risk of an investment.**

The specific operations are as follows.

1. **Calculate the portfolio weights:**

Think of gold and bitcoin as two risky assets and holdings in U.S. dollars as a risk-free asset. According to the mean-variance model, we can use diversification strategies to minimize risk and maximize return.

For the three assets, USD ,Gold and Bitcoin, we assign different weights to them: w_C , w_G , w_B indicating the percentage of our investment in the different assets. The weights satisfy the following equation:

$$w_C = \frac{C_{t+1}}{C_{t+1} + B_{t+1}\hat{P}_{t+1}^{B} + G_{t+1}\hat{P}_{t+1}^{G}} \tag{2}$$

$$w_G = \frac{G_{t+1}\hat{P}_{t+1}^{G}}{C_{t+1} + B_{t+1}\hat{P}_{t+1}^{B} + G_{t+1}\hat{P}_{t+1}^{G}} \tag{3}$$

$$w_B = \frac{B_{t+1}\hat{P}_{t+1}^{B}}{C_{t+1} + B_{t+1}\hat{P}_{t+1}^{B} + G_{t+1}\hat{P}_{t+1}^{G}} \tag{4}$$

$$w_{C+} w_G + w_B = 1 \tag{5}$$

2. **Quantify portfolio risk:**

Since **the risk of each asset can be measured by the variance**, we use the historical price data at each moment t to calculate the variance of each asset and **the covariance of the risky asset** σ_{BG}.

We believe that premature price data is not informative for current trading. Therefore, we now use only historical prices within 30 days from today to calculate the volatility variance of our assets. Based on the variance and covariance of each asset, the portfolio variance σ is calculated as follows:

$$\sigma^2 = w_B^2 \sigma_B^2 + 2 w_B w_G \sigma_{BG} + w_G^2 \sigma_G^2 \tag{6}$$

3. **Consider individual investment preferences:**

Since each investor has a different level of risk aversion, **we introduce** A **as coefficient of variance effect to differentiate the level of risk aversion of individuals.**

In general, we can quantify the degree of risk aversion as a specific value, expressed as an integer between 0 and 10 based on the degree of preference for risky assets.

In our task, we classify investors into 5 classes according to their risk aversion: 2, 4, 6, 8, and 10. That is:

$$A \in [2, 4, 6, 8, 10] \tag{7}$$

A: variable measures people's degree of risk aversion.

As the value increases, the risk aversion increases and the trade is gradually influenced by the variance of the asset portfolio.

To sum up, our another goal is :

$$\min\ A\sigma^2 \tag{8}$$

- **Combining Return and Risk**

Our final goal is:

$$\max\ C_{t+1} + G_{t+1} * \hat{P}_{t+1}^{G} + B_{t+1} * \hat{P}_{t+1}^{B} - A\sigma^2 \tag{9}$$

5.2.3 Determination of Constraints

- **Constraint I**

We trade through our portfolio of assets in hand on day t to get our expected asset holding outcome by **paying a certain transaction cost.** For this, we have the following constraint:

$$\begin{gathered}(C_{t+1} - C_t) + (G_{t+1} - G_t) P_t^G + (B_{t+1} - B_t) P_t^B = \\ -|G_{t+1} - G_t| * P_t^G * \alpha_{gold} - |B_{t+1} - B_t| * P_t^B * \alpha_{bitcoin}\end{gathered} \tag{1}$$

Where α_{gold} is the transaction costs of gold, $\alpha_{bitcoin}$ is the transaction costs of bitcoin.

- **Constraint II**

Since gold is not traded on trading days, we have the following constraint:

$$G_{t+1} = TG_{t+1} + (1 - T) G_t \tag{2}$$

When the current day is trading day, $T=1$, else $T=0$.

- **Constraint II**

  The number of assets is non-negative.

$$G_{t+1} \geqslant 0 \tag{3}$$

$$C_{t+1} \geqslant 0 \tag{4}$$

$$B_{t+1} \geqslant 0 \tag{5}$$

5.2.4 General Format

To sum up, our Decision Model at DAYt is:

$$\max\ C_{t+1} + G_{t+1} * \hat{P}^{G}_{t+1} + B_{t+1} * \hat{P}^{B}_{t+1} - A\sigma^2$$

$$\begin{cases} (C_{t+1}-C_t)+(G_{t+1}-G_t)P_t^G+(B_{t+1}-B_t)P_t^B = \\ -|G_{t+1}-G_t| * P_t^G * \alpha_{gold} - |B_{t+1}-B_t| * P_t^B * \alpha_{bitcoin} \\ \sigma^2 = w_B^2\sigma_B^2 + 2w_B w_G \sigma_{BG} + w_G^2\sigma_G^2 \\ w_B = \dfrac{B_{t+1}\hat{P}^B_{t+1}}{C_{t+1}+B_{t+1}\hat{P}^B_{t+1}+G_{t+1}\hat{P}^G_{t+1}} \\ w_G = \dfrac{G_{t+1}\hat{P}^G_{t+1}}{C_{t+1}+B_{t+1}\hat{P}^B_{t+1}+G_{t+1}\hat{P}^G_{t+1}} \\ \hat{P}^G_{t+1} = Predict_G([P_1^G,P_2^G,\cdots,P_t^G]) \\ \hat{P}^B_{t+1} = Predict_B([P_1^B,P_2^B,\cdots,P_t^B]) \\ C_{t+1},G_{t+1},B_{t+1} \geqslant 0 \end{cases} \tag{1}$$

Where $[P_1^G,P_2^G,\cdots,P_t^G]$ is the historical price series; $Predict_G$ and $Predict_B$ is the symbolic representation of prediction models; $\sigma_B,\sigma_G,\sigma_{BG}$ can be calculated by historical price series.

5.3 Model Results

In this part, we will start to simulate real trading scenarios from September 11, 2016 and formulate daily trading strategies based on the above two models. The specific process is described as follows.

- **Start:** When we are at DAY1, Our initial portfolio is $[C_0,G_0,B_0]=[1000,0,0]$.
- **Step1:** Enter the Three-Stage model. First we determine the stage of the day. Then we consider the price and solve the prediction model. We can get the predicted value of next day: $\hat{P}^G_{t+1}$ and $\hat{P}^B_{t+1}$.
- **Step2:** Enter Decision Model. Calculate the target asset portfolio for the day .That is $[C_{t+1},G_{t+1},B_{t+1}]$, i.e. the initial portfolio of the next day
- **Step3:** Trade according to the plan to reach our trading goals.

- **Step4:** When the time came the next day, $t = t + 1$. Back to **Step1**

In Grey Model, we use historical data **within 5 days** to predict price of DAY $t + 1$. In Random Forest Regression, **we set the size of the Slide_Window to 30** according to experience. Also, we set the **depth** of the tree model to be 2 and **n_estimators** to be 10 . In the process of solving the decision model, we use **SciPy in Python** to solving nonlinear programming.

Our calculated final asset values for different levels of risk aversion are shown in the table below.

Table 2. The Final Value

A	**2**	**4**	**6**	**8**	**10**
Final Value （USD）	1370480	77976	11553	6299	4572

6 Why Our Model

In this section, we will discuss why our model gives the best strategy. We can divide our model into two parts: the prediction part and the decision part. Different solutions for each part can lead to different results. We will discuss the optimality of the model of each part. Due to space limitations, we conduct the following discussion with a risk aversion level of 6.

6.1 Optimality of Forecasting model

6.1.1 Comparison with prediction quality of other Models

Let's review our prediction model. We have divided the trading period into three stages according to the estimated information volume. At the first stage, we borrow the idea from Random Walk. At the second stage, we select the Grey Model. And at the third stage, we use Random Forest Regressor, the machine learning method.

In the financial markets, there are also other useful methods of price forecasting such as **XGBR, SVR, ARMA** and so on. They all have different application situations. In the following, we will select different methods we mentioned above and apply them at **StageⅢ** respectively. After that, we use **MSE, MAE,RMSE** and **final asset value** to show the performance.

SVR: SVR is based on SVM algorithm, mainly by constructing a linear decision function in a high-dimensional space after dimensionality enhancement to achieve linear regression.

XGBR: XGBR is based on XGBoost algorithm, which is an improved algorithm based on gradient lifting decision tree. It can effectively build enhancement tree and run in parallel.

ARMA: It is an important method for studying time series, consisting of an autoregressive model (AR model for short) and a moving average model (MA model for short).

MSE: Mean Squared Error. The formula is:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \tag{1}$$

MAE: Mean Absolute Error. The formula is:

$$\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{2}$$

RMSE: Root Mean Square Error. The formula is:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \tag{3}$$

Where n is the number of total samples, y_i is the true value, $\hat{y}_i$ is the predicted value.

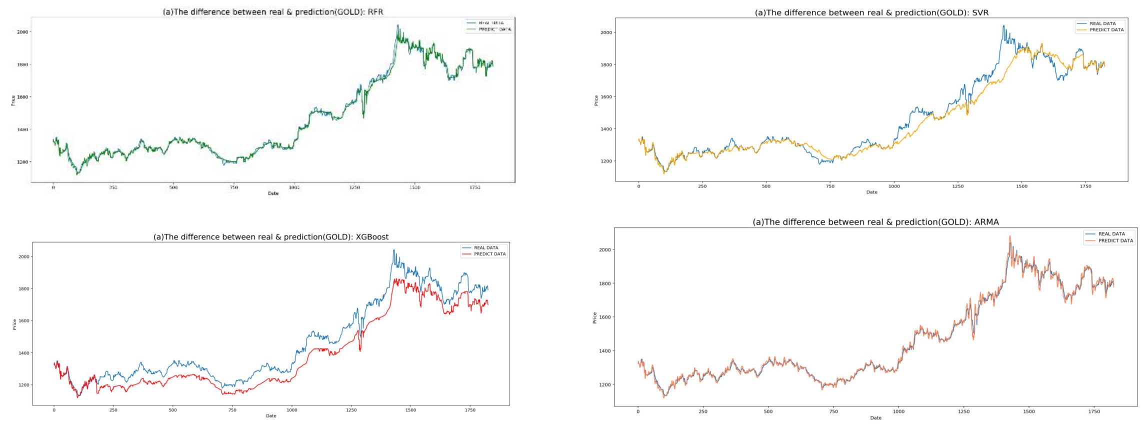

Figure 4. Performance on Gold

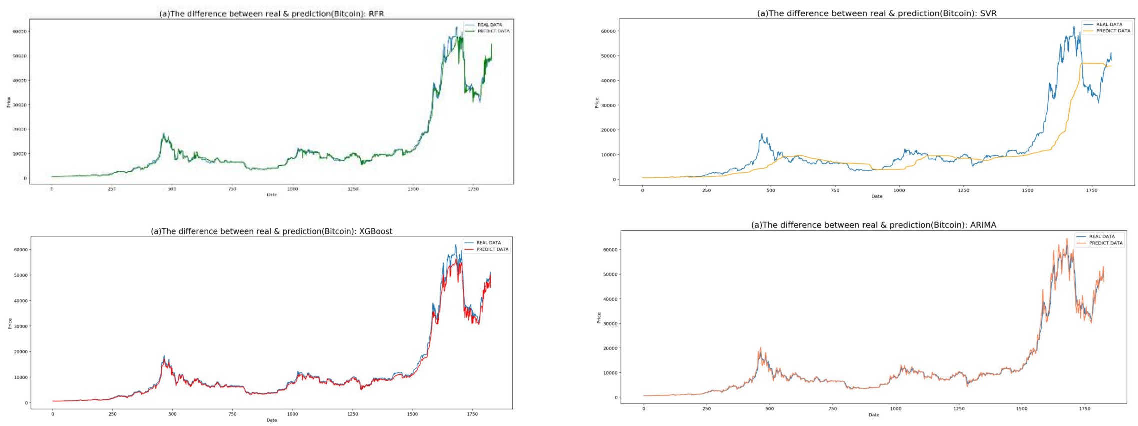

Figure 5. Performance on Bitcoin

Table 3. Performance on Gold

Model Name	MSE	MAE	RMSE
RFR	590.7	16.17	24.3

SVR	3888.94	38.74	62.36
ARMA	615.80	18.34	24.82
XGBR	7834.24	83.53	89.07

Table 4. Performance on Bitcoin

Model Name	MSE	MAE	RMSE
RFR	3421003.75	968.09	1254.75
SVR	74255386.50	4619.75	8617.15
ARMA	3941063.17	1043.83	1985.21
XGBR	3597834.97	1047.73	1927.59

As can be seen from the images and tables, the RFR model shows the best result, proving we choose the correct model.

6.1.2 Comparison with self parameters

We draw the calculation results of gold price and bitcoin price in one day under different depths and estimator values as well as under the medium risk aversion level. When depth=2 and estimator =10, the model has a good performance. When the parameter value is smaller than this values, MAE is too large and the prediction result is not added. When the above values continue to increase, MAE decreases little, which is easy to cause overfitting. Therefore, our parameter selection is correct.

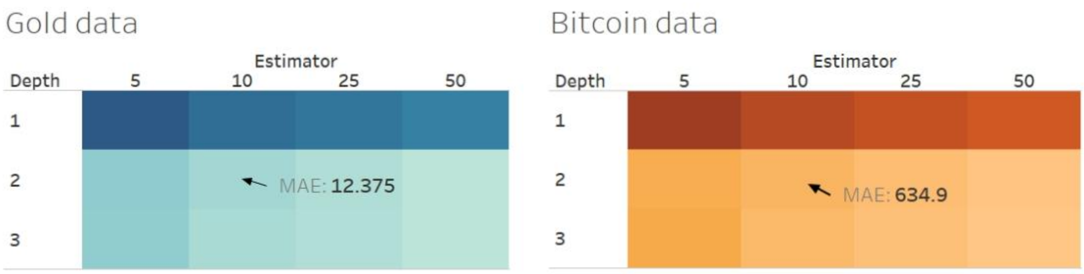

Figure 6. MAE of Gold and Bitcoin under different parameters in RFR

6.1.3 Further analysis of our predictive models

Observe the forecast result for gold and bitcoin（**Figure 7. and Figure 8.**）:

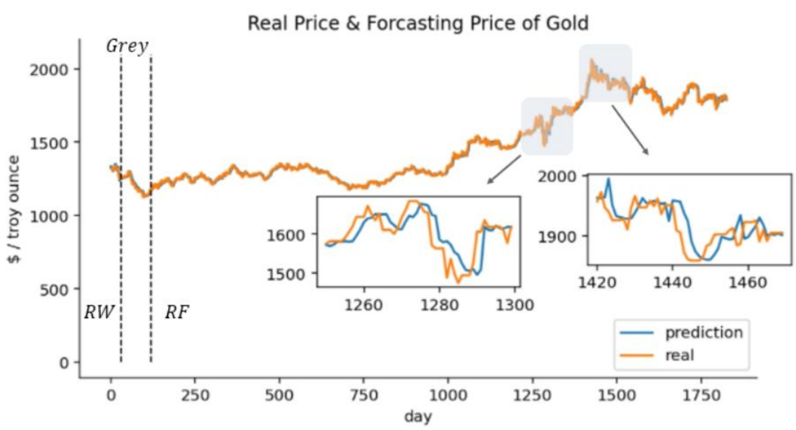

Figure 7. Real and Predicted Value of Gold Price

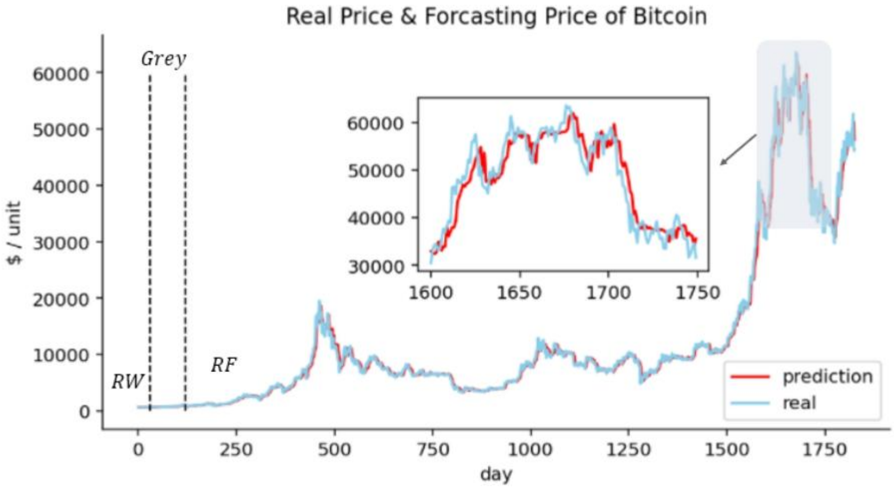

Figure 8. Real and Predicted Value of Bitcoin Price

- In our Three-Stage Price Forecasting Model, Random Walk and Grey Model were set up reasonably. For both gold and bitcoin, the curve between 0 and 180 is flat and easy to predict, so it can produce good results. After obtaining enough data, we turn to the Random Forest Regression for prediction.
- For gold data, the overall trend is relatively stable, the variance is small, the asset risk is small, and the prediction accuracy is high. Furthermore, There are two obvious mutation points in gold price between 1250-1300 stage and 1420-1470 stage, which have been marked in **Figure 7.**
- For bitcoin data, the overall trend changes greatly, the variance is large, the asset risk is also large, and the prediction accuracy is lower than that of gold data. However, the model can still capture the price movement trend well in most time periods. Bitcoin data has multiple mutation points, among which the most severe mutation occurred between 1600-1750 and has been marked in **Figure 8.**
- By observing and predicting the overall value and the value at the mutation stage, the model has the following characteristics and advantages in prediction:
  - When the data is relatively flat, the model can capture the change trend well.
  - When data surge or collapse, although the model cannot predict the same as the real value, it still can give a rise or fall "forecast". Investors base on the "preview" to make decisions -- such as the trailer rises in advance to buy assets -- forecast decline to sell assets in advance, thereby gaining profits or avoid losses;
  - When the actual data consecutive alternating rise and fall in a short period of time, the model can be well smoothed, such performance has two points -- one is to avoid frequent trading, resulting in high transaction costs; Second, risk aversion. For example, in the case of multiple consecutive rise before they fall, the prediction is falling, according to the planning equation, we will immediately sell assets, no longer buy during prediction falling, until the model predicted rises again to buy, save transaction cost, and avoid the "a little rise followed by huge loss".

6.2 Optimality of Decision model

Based on the predicted value, daily position portfolio and position value can be calculated by programming equations (**Figure 9, Figure 10**) on 9/10/2021, the total value of assets was $11553.19.

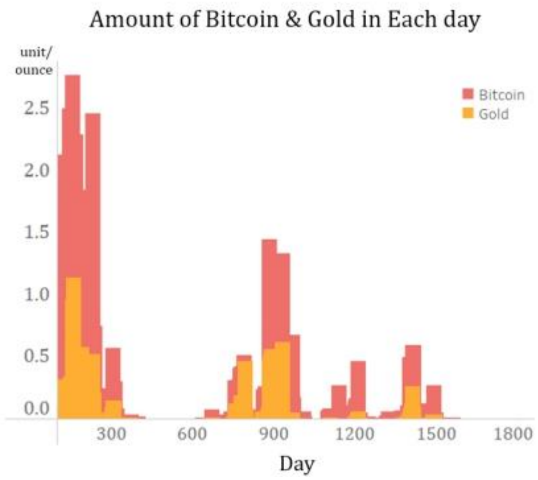

Figure 9. Position of Bitcoin and Gold Each Day

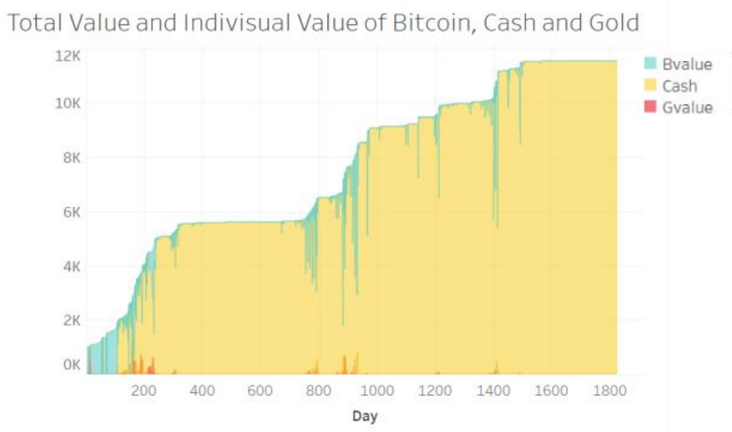

Figure 10 . Total and Individual Value of Bitcoin, Cash and Gold

Observe daily changes in bitcoin and gold:

- In the first 374 days, the model automatically buys bitcoin and gold, and the asset value also shows a rapid increase from the initial $1,000 to $5,648.61, achieving an annual return of nearly 500%.
- Then, at 374 to 600 days, the model detects large fluctuations in bitcoin, that is, large variance in the first 30 days, and stops trading in bitcoin. The price trend of gold is expected to be relatively flat, and the slight increase in gold is not enough to cover the cost of trading gold, so we do not trade gold. As a result, the model hardly trades in the second year due to a combination of benefits and risks.
- During the following 600 to 1507 days, the trend of bitcoin gradually flattens out. According to the variance in the first 30 days, the model believes that bitcoin transactions can be carried out and bitcoin can be bought. Meanwhile, gold rises, so a certain amount

of gold is also bought. The value of assets also showed a rapid growth, from \$5,648.61 to \$11553.19, realizing the doubling of assets.

- According to the past portfolio variance, the model judges the abnormal fluctuations of bitcoin again after 1507 days. Taking into account the sharp increase in transaction risks, the model stops trading bitcoin, while gold also shows a downward trend, so the asset position is completely emptied.
- To sum up, the strategy constructed by our model has the following characteristics and superiority:
  - The strategy is constructed with full consideration of asset risk. When the variance of an asset in the past 30 days is detected to be large, the asset will be sold to avoid excessive risk of the position. Therefore, this model is very suitable for the financial risk management requirements of enterprises or individuals.
  - Models are sensitive to rising or falling trends when asset movements are relatively flat. While the strategy sometimes fails to make a profit following this trend, it is a great guarantee that it will not incur a loss.
  - The structure of the strategy product is very considerable, in the gentle rise of assets to achieve a great return.

7 Discussion of Transaction costs

When a gold or bitcoin transaction takes place, people usually have to pay a percentage of cash, which is the **transaction commission.** In addition to this, investors in many countries are required to pay a certain amount of capital gains tax on their investments due to the taxation system of stocks. That is, the government is required to collect a percentage of the investor's net income in taxes.

Therefore, when considering the impact of transaction costs on investment decisions, we will consider both of these factors together and explore the optimal investment decisions of investors.

7.1 Trading commission

The trading commission is usually a percentage of the transaction amount. In this question, the transaction commission for gold is $\alpha_{gold} = 1\%$, and the transaction cost for bitcoin is $\alpha_{bit} = 2\%$. We consider fluctuating the trading commission rates proportionally with a moderate level of risk aversion ($A = 6$). We can get the magnitude and direction of the change in the final total value of each portfolio.

Table 5. Fluctuation of Final Value

Bitcoin \ Gold	-10%	-1%	0	1%	10%
-10%	0.07	-0.27	-0.30	-0.24	-0.02
-1%	-0.08	0.07	-0.03	-0.03	0.05
0	-0.05	0.00	0.00	0.01	0.17
1%	-0.02	0.03	0.04	0.04	0.12
10%	-0.07	-0.03	0.53	0.57	0.64

From the above data, it can be seen that the impact of gold and bitcoin trading commissions on the final investment return has the same trend. Both show a positive change in relationship. The specific analysis is as follows.

- When the changes of α_{gold} and $\alpha_{bitcoin}$ are within 1% , the sensitivity of the final value of the asset portfolio is small, which is within 10%. Combined with the real investment situation, we speculate that **higher transaction costs for investors and higher opportunity costs due to a small increase in trading commissions.** Investors will therefore adopt a more cautious investment strategy, forgoing aggressive investment opportunities and avoiding potential risks. This leads to an increase in the total value at the end of the period.
- When the changes of α_{gold} and $\alpha_{bitcoin}$ are around 10% , The final value of assets can change by up to 64%. This indicates that: when trading commissions increase significantly, the cost of investment increases and investors will reduce the number of trades and choose investments with higher returns. Therefore the total value at the end of the period increases.
- The final product of assets is more sensitive to bitcoin transaction fees $\alpha_{Bitcoin}$. As can be seen from the figure, the change of bitcoin transaction commission will cause a great change in the final product. This is inseparable from the price characteristics of Bitcoin itself: the price of Bitcoin has a wide range of changes, fast speed and poor stability, but it also has many speculative opportunities. Therefore, investors can make reasonable decisions, so that the return is greater than the trading commission.

7.2 Capital Gains Tax

In addition to transaction commissions, capital gains taxes make up the second part of transaction costs. In each market, investment income is taxed at a rate known as capital gains tax. Similarly, investment gains in gold and bitcoin are subject to a percentage of capital gains tax. For different amounts of income, the tax rate is also different. Generally, the higher the income, the greater the tax rate. For the two assets in this question, gold and bitcoin, we have found the applicable tax rates as shown in the table below:

Table 6. Capital gains tax rate in USA[8]

Tax rate	Taxable income bracket
10%	\$0 to \$9,950
12%	\$9,951 to \$40,525
22%	\$40,526 to \$86,375
24%	\$86,376 to \$164,925
32%	\$164,926 to \$209,425
35%	\$209,426 to \$523,600
37%	\$523,601 or more

According to the long-term capital gains tax policy, we are supposed to pay a certain amount of tax on each transaction's gains and deduct a portion of the tax on losses. For the sake of calculation, we assume that the capital gains tax at the end of each year takes the same deductible balance as the initial amount for the second year. After deducting capital gains tax, the solution model can obtain the final returns under different risk aversion degrees, as shown in the following figure:

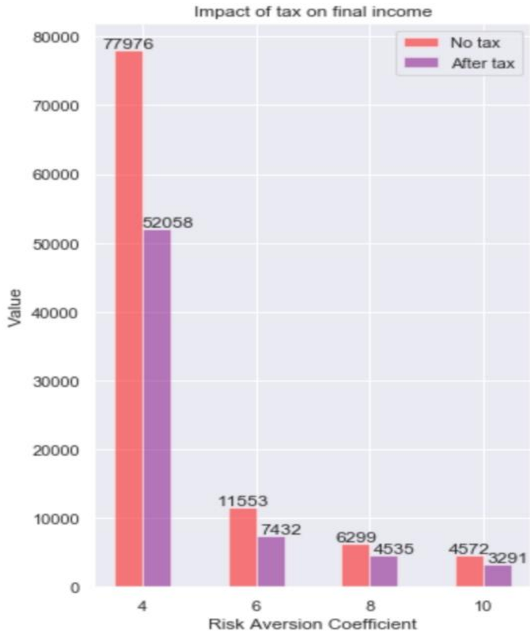

Figure 11. After-Tax Gains under Different Degree of Risk Aversion

It can be seen that investment strategies with different risk aversion degrees are affected differently by capital gains tax:

- Investors with low risk aversion prefer to invest in high-risk assets and have higher original returns, so they are greatly affected by capital gains tax;
- Investors with higher risk aversion are less affected by capital gains tax due to their conservative investment strategies and lower raw returns.

To sum up, investment strategies will be jointly affected by transaction commission and capital gains tax, resulting in certain transaction costs. On the one hand, it is influenced by portfolio: the risk assets and the transaction commission between portfolio will change investors' investment decision and investment profit; On the other hand, restricted by relevant policies, such as capital gains tax will reduce the returns of investors who prefer risk investment more.

8 Sensitivity Analysis

When we use the RFR model to predict, we select Slide_Window=30 according to the empirical value. Next, we analyze the prediction results of the model in different Slide_Window and observe the influence of Slide_Window change on the predicted value.

Figure 12.(a) and Figure 12.(c) respectively represent the difference between the predicted value of gold and bitcoin and the actual value of the last 500 Slide_Window periods of **30,180,270,360.**

- Whatever Slide_Window is, the variation trend of the difference between the actual value and the predicted value is basically the same, and the difference values under the 4 Windows basically coincide.
- Whatever Slide_Window is, the predicted value of gold is generally smaller than the real value.
- Whatever Slide_Window is, the difference between the predicted value and the actual value of bitcoin has a large variance, which is consistent with the conclusion that bitcoin has a large risk (variance) and a low prediction accuracy.

Figure 12.(b) and Figure 12.(d) respectively represent MAE and RMSE predicted by the model for gold and bitcoin when Slide_Window=30/180/270/360.

- Whatever Slide_Window is, there is no big difference between MAE and RMSE predicted by the model.
- With the decrease of Slide_Window, MAE and RMSE tend to increase, indicating that Slide_Window=30 is a reasonable choice for the model to reduce the prediction error.

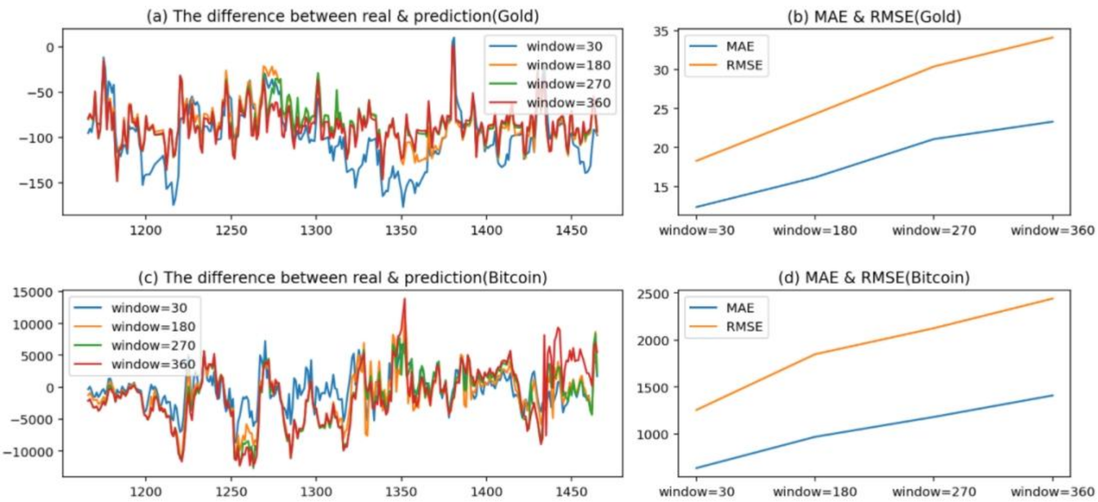

Figure 12. Sensitive Analysis

9 Model Evaluation

9.1 Strengths

- In the price forecasting, we create the mixed Three-Stage forecasting model that takes full account of the inadequate information volume of past days. We are able to combine small sample prediction methods and machine learning methods to ensure the accuracy of our models. This is innovation of our model.
- In the trading strategy model, we are influenced by the Markowitz portfolio model. We introduce the objectives of both return maximization and risk minimization. Based on this, we give separate decision strategies for different levels of risk aversion.
- We make a sufficient evaluation of our model including different model comparison,

parameter comparison, etc. The results fully demonstrate the advantages of our model.

9.2 Weaknesses

- Due to the availability of price information only, our forecasts of gold and bitcoin price trends may be subject to error.
- Although we consider both benefits and risks, the information we can rely on is limited. We idealized some cases, which may differ from the real situation.

10 Memorandum

Dear Sir/Madam,

Asset price forecasting has always been a hot topic. In order to better predict the price movements of bitcoin and gold and automatically trade profits, we propose a highly profitable model that can automatically trade. It's simple enough to consider past prices; It is comprehensive enough to measure benefits and risks, meet the risk management needs of the company, and take transaction commissions into account.

In a nutshell, the automated trading model can be divided into three steps:

- Forecast tomorrow’s price based on the price of the previous period.
- Establish a plan equation on the forecast price that automatically calculates the appropriate amount of bitcoin, gold and cash to be held under a trade-off of return and risk.
- If the most held position is different from the current position, it will trade accordingly.

This model has many advantages.

First of all, the model has high accuracy in predicting the future without considering the amount of data. Our prediction model is divided into three stages. When the amount of data is small, random walk and grey prediction will be automatically selected. These two models can obtain more accurate prediction results under small samples. When the amount of data is large enough, a more accurate random forest model will be used to further enhance the accuracy of prediction. Looking back at bitcoin and gold transactions over the past five years, our model has the following features and strengths:

- Compared with other models commonly used today, our model has the smallest prediction error (green curve in the figure) and is fast in calculation.
- When the data is relatively flat, the model can capture the change trend well.
- While models cannot predict exactly the same value as reality, they can still give a "forecast" of rise or fall when there is a spike or sharp decline. You can make decisions based on this "forecast" -- for example, buy assets early when the forecast goes up, sell assets early when the forecast goes down, and make a profit or avoid a loss.
- When the actual data rise and fall repeatedly in a short period of time, the model can be smoothed well, which has two advantages -- one is to avoid frequent trading, which leads to high transaction costs; Second, it avoids risks. For example, in the case of many consecutive rises and then falls, the model's forecast is a continuous decline, the automatic trading model will suspend trading until the model forecasts the rise again to

buy, which can reasonably avoid the large trading loss in the case of "small rises and huge falls" for many times.

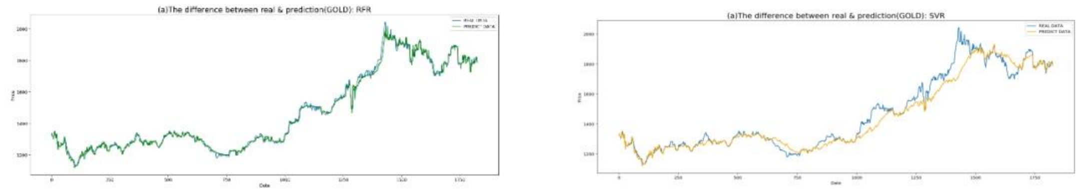

Figure 13. Actual Value and Predicted Value in Our Model

Secondly, the automatic trading model fully considers the risks that trading may bring. In the process of solving the optimal asset, we add the risk coefficient, which can be set artificially, and it adjusts the risk of holding an asset, that is, the standard deviation of an asset in the past 30 days, and the penalty. Within this constraint, you can tailor the model to your risk management needs. Looking back over the past 5 years of bitcoin and gold transactions, our model has the following features and advantages when the coefficient of risk aversion is set at a reasonable level:

The strategy is constructed with full consideration of asset risk. When the variance of an asset in the past 30 days is detected to be large, the asset is sold to avoid excessive risk of the position.

- Models are sensitive to rising or falling trends when asset movements are relatively flat. While the strategy sometimes fails to make a profit following this trend, it is a great guarantee that it will not incur a loss.
- The strategic return rate of the structure is very substantial, achieving great gains in the period of flat rising assets. Within 5 years, the asset price changed from \$1000 to \$11553.19, achieving a ten-fold increase.

Finally, the parameter is settable. In the model, we set a series of parameters:

A: Represents the risk aversion level of the enterprise. The higher the risk aversion level, the higher the enterprise's requirements for risk management. If you want to trade more conservatively in the future, you can increase the value of A. In this way, automated trading models do not buy assets that are volatile from the current period of time; If you want to increase your risk position, you can lower the value of A. In this way, automated trading models can receive products with higher volatility.

α : represents the transaction cost of an asset. In our model, you can adjust transaction costs to real values at will and change them at any time; The model also takes into account capital gains taxes, which you can choose to amortize into your daily transaction costs.

window_size: represents how many past periods of data are used for prediction. You can determine whether past data is predictive based on your rich experience, and then adjust this parameter at any time to make the model meet your needs.

Using the automated trading model is also very simple, you just set the parameters and the model automatically starts trading. As shown in the figure below, the model reports the number and value level of assets on a daily basis for your monitoring.

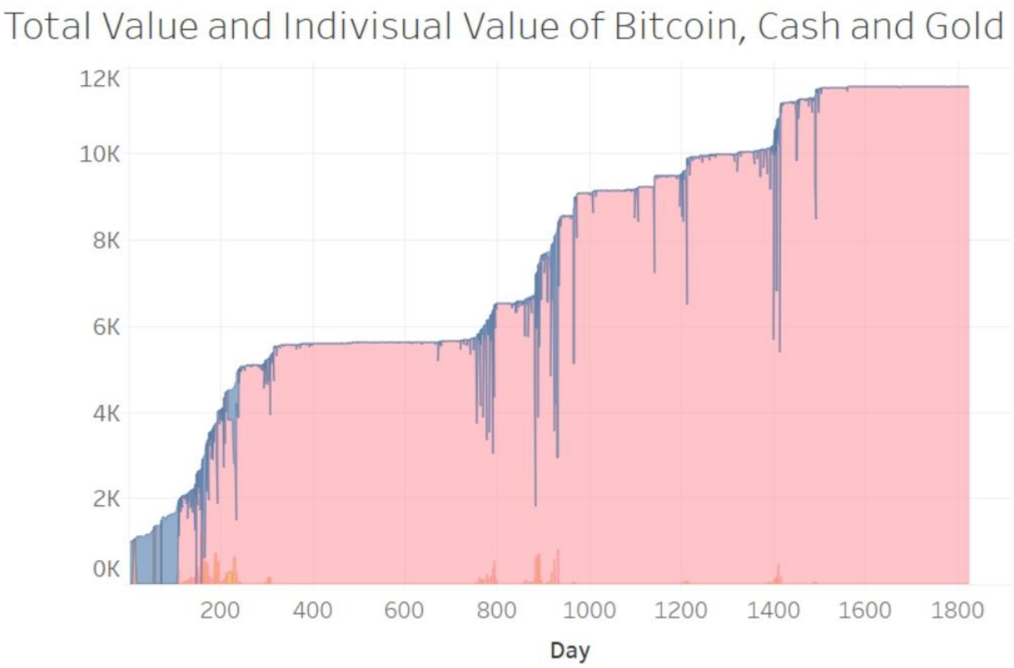

Figure 14. Total Value and Individual Value

We hope the automated trading model will be a powerful tool for you. If you have any questions about this model, please feel free to contact me and I will be happy to answer them.

Yours Sincerely,
Team # 2217563

References

[1] Liu Lijun, Liang guopeng, Wang Haitao Research on quantitative trading strategy system model [J] Shanghai business, 2021 (09): 158-159

[2] Wang Hongtao, Li Fan Analysis of bitcoin price bubbles based on VEC model [J]. China market, 2022 (03): 1-2 DOI:10.13939/j.cnki. zgsc. 2022.03.001.

[3] Zhang Ning, Fang Jingwen, Zhao Yuxuan Bitcoin based hybrid price prediction model [LSTM. J] Computer science, 2021,48 (S2): 39-45

[4] Zhang Yanhui, Yin Xiangyu, Yan Liang Combined forecasting model based on arma-ao-garch [J] Market weekly, 2021,34 (12): 155-157

[5] Ding Lei, Guo Wanshan Research on gold price prediction based on arima-garch hybrid model [J] Journal of Xuchang University, 2019,38 (06): 124-129

[6] Jing Zhigang, Shi Guoliang Prediction of gold price based on ls-svm-arima combined model based on wavelet analysis [J] Gold, 2017,38 (05): 5-8 + 14

[7] Deng Julong. Basic methods of gray systems [M]. Huazhong University of Technology Press, 1987.

[8] Jigar Patel, Sahil Shah, Priyank Thakkar, K Kotecha,"Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", Expert Systems with Applications, Volume 42, Issue 1, 2015, Pages 259-268, ISSN 0957-4174

[9] https://www.nerdwallet.com/article/taxes/federal-income-tax-brackets