Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
C		2227720

The Art of Trading

Bitcoin is an innovative investment tool that has been noticed in recent years, and the market shares of bitcoin have increased. While gold is also another popular investment target that has lasted many years. Many people believe that investing in gold can against inflation and others like to try another taste, so they invest in bitcoin, which can generate higher profit for them. For investors who hold a meticulous attitude towards the market, they choose to hold cash and wait for future investments. Therefore, we want to create a model which can get the **best-expected return** with a **proper transaction cost**. Based on this situation, several models are established: Model I: Random Forest Classification Predicting Model; Model II: Logistic Regression Model; Model III: Mean-Variance Optimizing Forecast.

Keywords:

Contents

1 Introduction

1.1 Problem Background

In recent years, bitcoin and gold are two popular assets for investment and investors always use these two assets to develop their trading strategies or hedge the risks. Some investors believe that the financial market will be worse so they long gold or hold the cash, while some investors want to get a higher profit, so they buy bitcoins. Bitcoin can be traded at any time since it is connected to the blockchain, but gold must be traded on a period of specific time.

(a) Bitcoin

（b） Gold

Figure 1: Target assets: (a) **Bitcoin**: A digital currency that has not been controlled by any governments or banks; (b) **Gold**: Under the current market system, gold is a kind of currency that can be bought or sold. Gold has a price, and the price fluctuates relative to other currencies, such as the US dollar, Japanese yen, etc.

1.2 Restatement of the Problem

Traders want to use an initial $1000 investment to maximize their profits through investing in bitcoin, gold, and cash, and by combing the information of the sources, the problems can be expressed as follows:

- Create a model that gives the best strategy based only on price data up to that day.
- Based on the model, optimizes the position that allocate to bitcoin, gold, and cash with the cost of commission.
- Back-test the model with the dataset and test the sensitivity and robustness of the transaction cost.
- Consider the results of the model and submit at most two pages memorandum to the trader.

1.3 Literature Review

This problem is mainly connected to how to allocate with the position with cash, gold, and bitcoin [C, G, B] and many researchers have already studied this question. Reuter believes that cryptocurrency (like Bitcoin) will replace the position of gold [1]. It claims that bitcoin has $700 billion in market capitalization, and gold only has 2.6 trillion. Cryptocurrency has occupied the 20% of the market. While the author Marco Quiroz-Gutierrez also explains that bitcoin has the character of anti-inflation and institutional investors would engage in more bitcoin trades [2]. Therefore, we can conclude the Pros and Cons for cash, bitcoin, and gold.

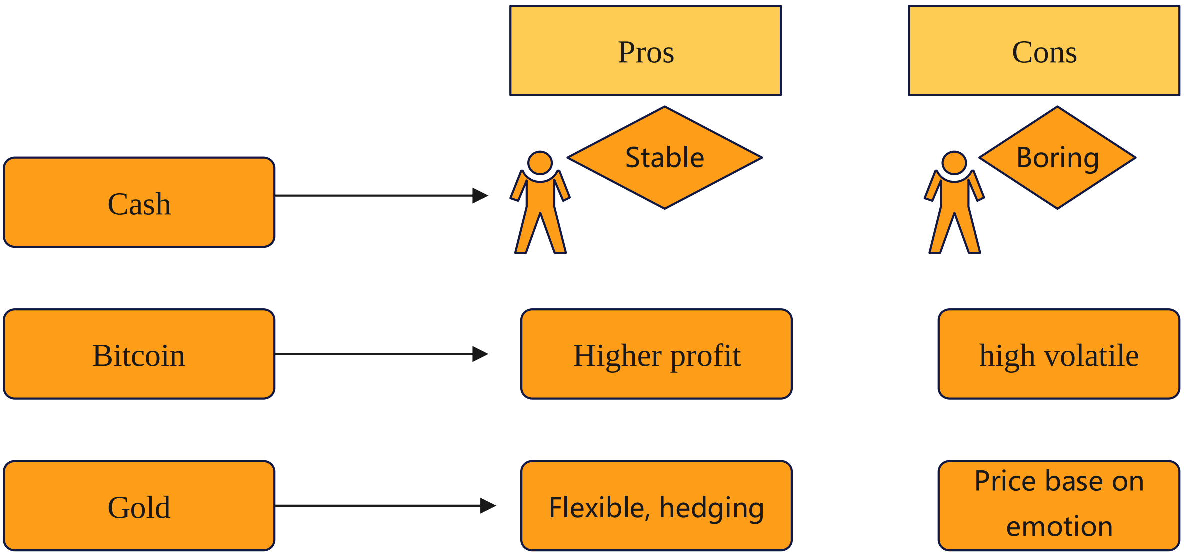

Figure 1: Literature Review Framework

1.4 Approach

To optimize the portfolio, we divide our process as the following:

1) Organize the data and visualize its statistical performance initially
2) Running Random Forest Classification Predicting Model, Logistic regression model, and Mean-variance Optimization Forecasting model to help predict the trends of investment strategy from historical price of bitcoin and gold
3) Get

2 Our Assumption and Model Overview

Since the trading strategies always are affected by many factors, we want to simplify the problem and the followings are our assumptions:

- **Assumption 1:** Assume that we only make transactions when both the gold and bitcoin's price are valid.

  Justification: Specifically, since bitcoin can be traded every day, but gold is only traded on days the market is open, some value of the price in the merged data frame is NaN. However, if we are going to use the merged dataset, we are going to drop the null values. This leads to the assumption that we only make transactions on the days when gold and bitcoin both can be traded.

- **Assumption 2:** ignore any political policy in the market that may dramatically affect the price fluctuation of bitcoin and gold.

  Justification: We believe that policy is an unstable factor that can severely affect our models.

- **Assumption 3:** Assume that there is no worldwide economic crisis that may affect the price of gold or bitcoin from 2016.9 to 2021.9.

  Justification: Economic crisis will lead to a big fluctuation in the price of bitcoin and gold and other dummy variables will appear.

- **Assumption 4:** Assume that there are some inter-relationships between the price of gold and bitcoin.

  Justification: Such that we can predict the rate of return of gold using some of the information of bitcoin.

- **Assumption 5:** There exists a potential linear relationship between the prior price of bitcoin and gold between the current price of bitcoin and gold.

  Justification: This assumption is just for the logistic regression model.

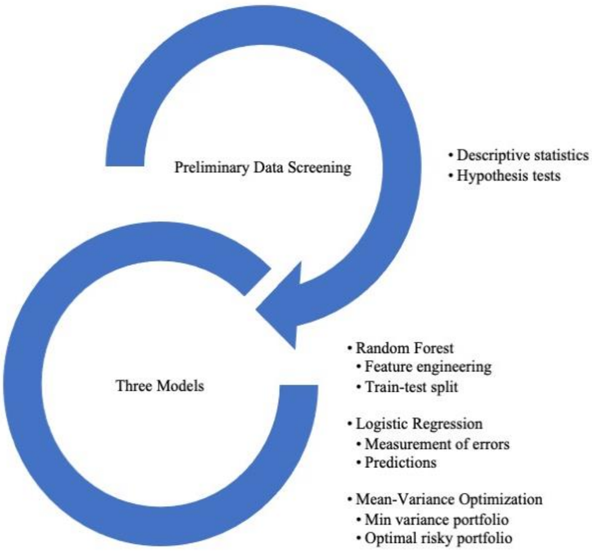

Figure 2: Model Overview

3 Model Preparation

3.1 Notation

Table 1: Notations

Symbol	Description
Y_i-gold	ith return of gold
Y_i-bit	ith return of bitcoin
$\widehat{Y_i}$-gold	Predicted ith return of gold
$\hat{Y}_i$-bit	Predicted ith return of bitcoin
X$std_$bit	Standard deviation of bitcoin
X$std_$gold	Standard deviation of gold
Wb/Wg	Weight of bitcoin/gold

Table 2: Other Symbols

Symbol	Description
MAE	Mean absolute Error
RMSE	Root Mean Squared error
MSE	Mean Squared error

Notes: we will also discuss other variables in the specific parts.

3.2 Data Overview

3.2.1 Data Collection

Since we are going to sacrifice some datasets during training the data, we are going to collect the relevant gold price and bitcoin price from the same source as given in the problem.

Table 3: Data Collection

Database	Data Sources	Data Type
Bitcoin	NASDAQ	Float
Gold	London Bullion Market Association	Float

3.2.2 Data Screening

In this problem, we mainly use the **Expected Return, Standard Deviation, Skewness, and Kurtosis** of these two assets to see the preliminary result. We also have calculated these indicators of the two databases.

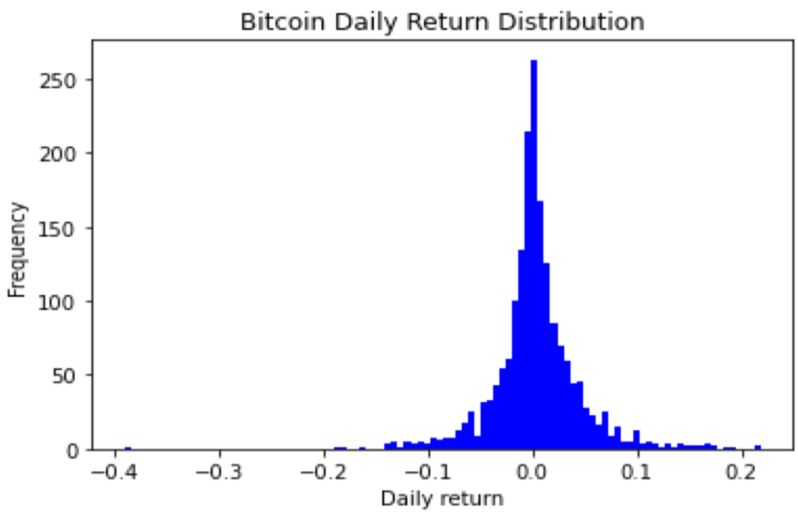

Figure 3: Distribution of Bitcoin

First, it is important to standardize the data and get the annual expected return of bitcoin is 0.808785 and the annual standard deviation is 0.65597. The skewness and kurtosis of bitcoin are -0.135 and 7.945 which is leptokurtic with fat tails.

$$\text{Skewness} \approx \left(\frac{1}{n}\right) \frac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{s^3}.$$

Skewness is a quantitative measure of skew (lack of symmetry), a synonym of skew. It is computed as the average cubed deviation from the mean standardized by dividing by the standard deviation cubed

$$K_E \approx \left[\left(\frac{1}{n}\right) \frac{\sum_{i=1}^{n} (X_i - \overline{X})^4}{s^4} \right] - 3.$$

Kurtosis is a measure of the combined weight of the tails of distribution relative to the rest of the distribution.

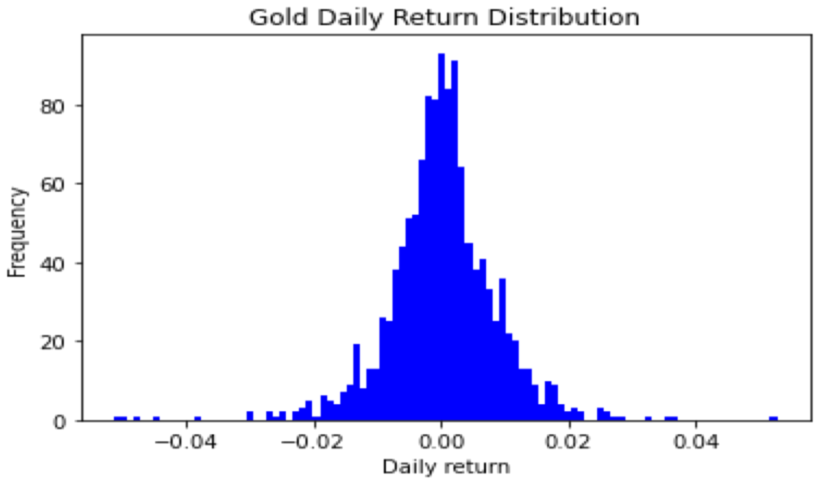

Figure 4: Distribution of Gold

Then, the annual expected return and standard deviation of gold are 0.0498 and 0.1367. The skewness and kurtosis are -0.3376 and 5.1630. Also, after analyzing the sharp ratios of the two assets (1.1802 for bitcoin and 0.1112 for gold), we conclude that our portfolio should have more weight of bitcoin since a unit of investment would generate more profit than gold.

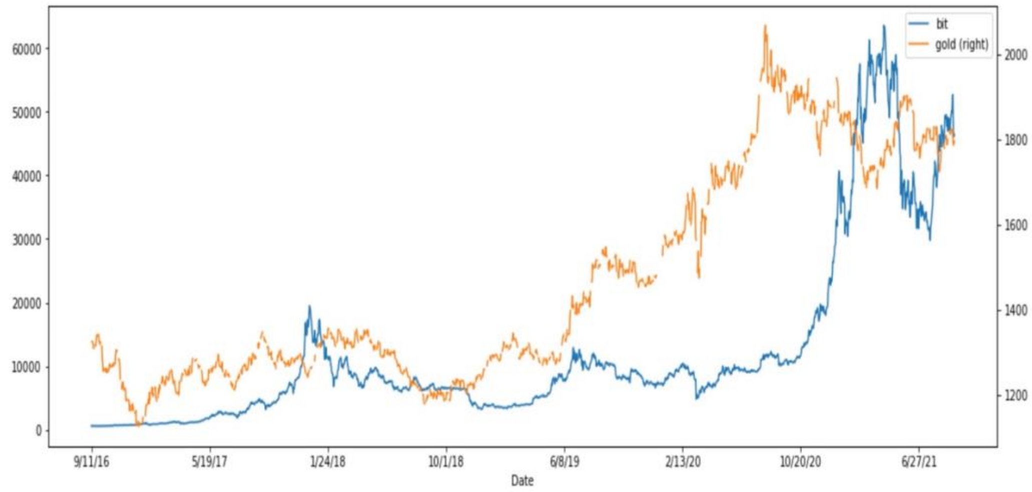

Figure 5: Time Series of Two Assets' price

We generated a graph combining the change of price of bitcoin and gold. We can observe unlike bitcoin, the graph of the price of gold is not a continuous line. This is because bitcoin can be traded every day, but gold is only traded on days the market is open. We can see that the price of gold increased dramatically between 6/8/2019 to around 10/20/2020, and fluctuated between 10/20/2020 and 9/11/2021, which is an interesting trend worth utilizing in our trading strategy model. Compared to the price of gold, the price of bitcoin seems more volatile. Starting from around 10/20/2020, the price of bitcoin started to increase and then fluctuated more volatile than gold.

3.3 Hypothesis Test Procedure

From the graphs above, we can see that most data is clustered around point zero, so we want to make a hypothesis test to confirm the previous descriptive statistic results.

One sample t-test:
Null hypothesis: The daily return is equal to zero

Two samples mean t-test:
Null hypothesis: The daily return of bitcoin and gold assets is equal to zero

Chi-square t-test:
Null hypothesis: The daily volatility of bitcoin is equal to zero

Table 4: Hypothesis testing statistics

Type of hypothesis tests	P-value of Bitcoin	P-value of Gold
One sample mean t test	0.00088174	0.41661043
Two samples mean t-test	0.00246768	0.00246768
Chi-square test for one sample	0.00	0.00

Table 5: Hypothesis testing results

Type of hypothesis tests	Bitcoin	Gold
One sample mean t test	Reject null	Cannot reject the null
Two sample mean t test	Reject null	Reject Null
Chi-square test for one sample	Reject null	Reject Null

From the results we get from doing the hypothesis tests, we can have sufficient evidence under 5% significance level that the daily total volatility of bitcoin and gold is not equal to 0, and the return performance of bitcoin and gold is not equal, which means both are a risky asset and we can take a strategy by diversification investment, which approximately follows the same trend from the descriptive statistics shown previously.

3.4 Exploratory Data Analysis (EDA)

EDA analysis would help us to check the integrity of the process which discovers patterns, spots anomalies, tests hypotheses, and checks assumptions with the help of summary statistics and graphical representations. Also, EDA can visualize the data and we can get the following chart with the tendency of both change of bitcoin and gold.

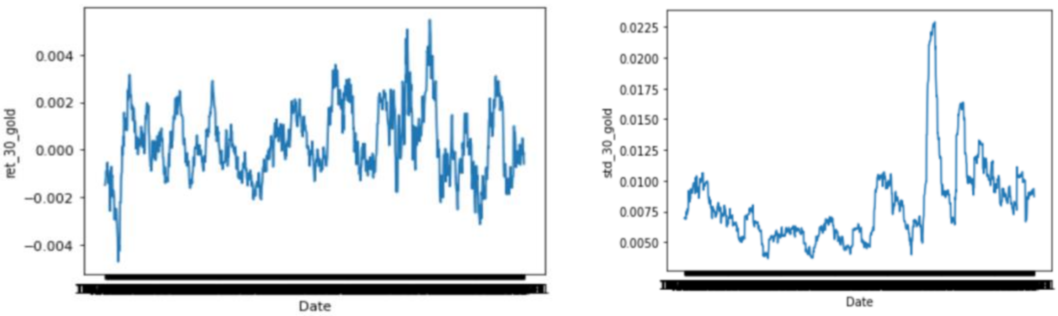

Figure 6: 30-day rolling percent change of price change of bitcoin/gold

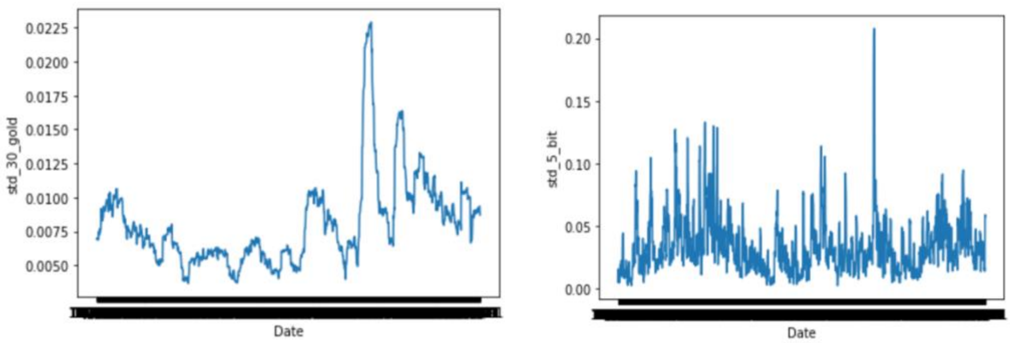

Figure 7: 30-day rolling standard deviation of price change of bitcoin/gold

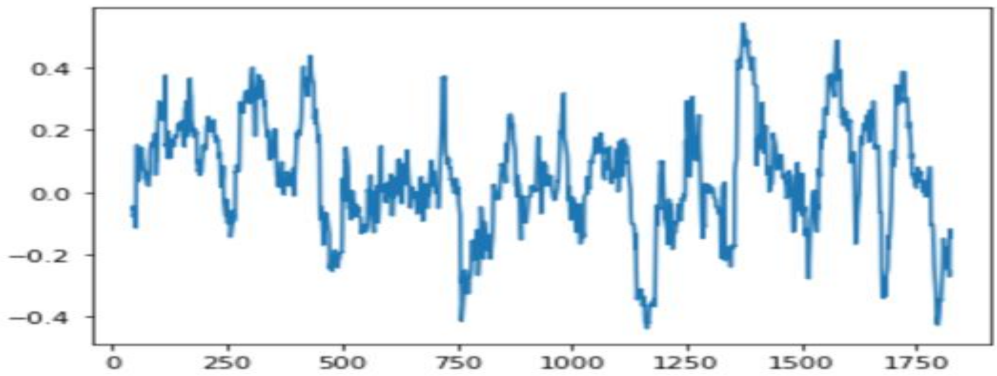

Figure 8: 30-day rolling correlation between the price change of bitcoin and price change of gold

4. Model I: Random Forest Classification Predicting Model

4.1 Model Establishment

In this model, we are trying to select several features from our price datasets and find out if there is a relationship between these features and the ultimate prices. We are going to construct a random forest model to solve this problem. The first step is the data collection, since we are going to sacrifice some datasets during training the data, we are going to collect the relevant gold price and bitcoin price from the same source as given in the problem. For Gold, we collected the price from 2016-1-4 to 2016-10-1 (Source: London Bullion Market Association). For Bitcoin, we collected the price also from 2016-1-4 to 2016-10-1(Source: NASDAQ). We merge two price datasets of bitcoin and gold into one table from the time series effect.

4.2 Test Model

4.2.1 Feature Engineering and Train-test split

For this dataset, since we only have the price for the data, we are going to use the rolling window with different days and different statistical features as our predictors. From the frame of random forest, it shows a strategy return of bitcoin and gold, and we want to find out the optimal percentage to invest from 1000 dollar original with the known expected return. Specifically, we first compute the return of each day and then compute the other features. Since the Rolling window method could be easily interfered with by the error from previous results, and by considering that the fluctuation and the value of the previous prices may affect the future ones, we include the rolling standard deviation and rolling mean of the previous 30 days and 5 day's prices as our feature. We have done 8 rolling windows to test different statistical features like standard deviation, mean, max, min of returns in different periods. After we prepare feature engineering, we start running the model of machine learning. We make a classification of outputs by setting 1(buy bitcoin), when the return of bitcoin > return of gold; -1(buy gold) when the return of bitcoin < return of gold

	precision	recall	f1-score	support
-1	0.99	0.95	0.97	156
1	0.96	0.99	0.97	194
accuracy			0.97	350
macro avg	0.97	0.97	0.97	350
weighted avg	0.97	0.97	0.97	350

Figure 9: Classification Report Metrics

We now split the dataset into 75% Training dataset and 25% for the Testing dataset by splitting the X and y into train and test datasets. With the internal package of Random Forest Classifier by python, when testing the machine level, we get a relatively high 97.14% Correct Prediction.

From the frame of random forest, it shows a strategy return of bitcoin and gold, and we want to find out the optimal percentage to invest from 1000 dollar original with the known expected return. The following charts are shown the predicted cumulative return and daily return of both bitcoin and gold based on this model.

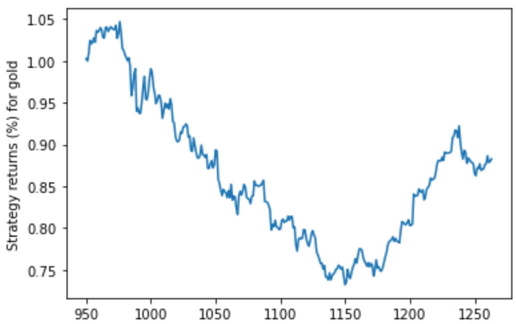

Figure 10: Cumulative return of Bitcoin

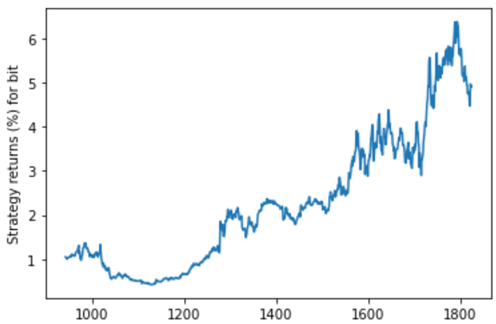

Figure 11: Cumulative return of Gold

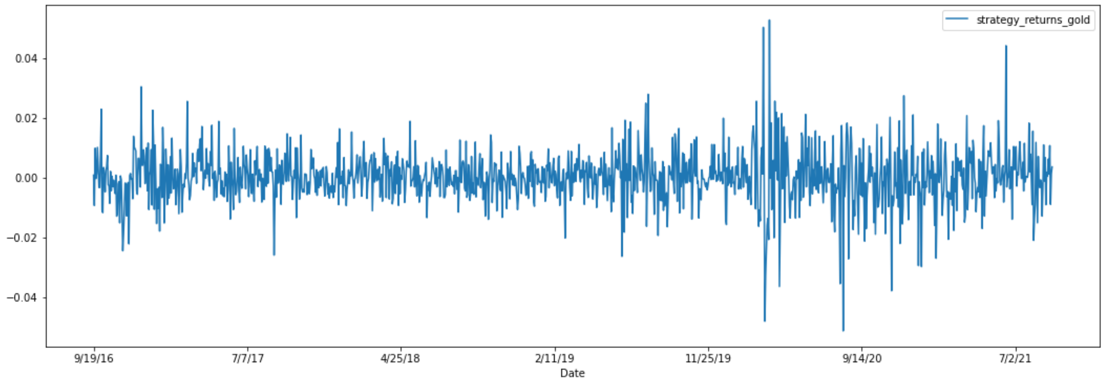

Figure 12: Daily return of Bitcoin

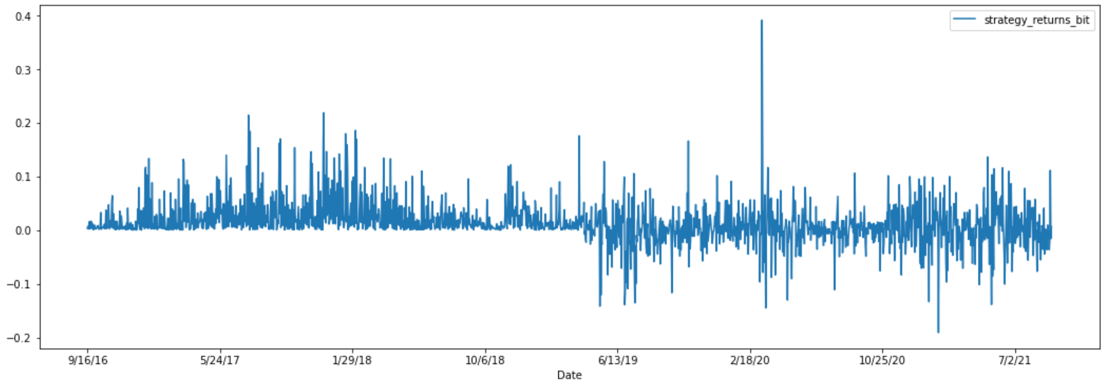

Figure 13: Daily return of Gold

4.2.2 Model Back-test

Based on our assumption, we are going to buy bitcoin when the return of bitcoin – the return of gold > 0.05; buy 50%-50% bitcoin and gold when -0.05 < return of bitcoin – the return of gold < 0.05; buy only gold when the return of bitcoin – the return of gold < -0.05

Assumption: weight of bitcoin and gold is 50% separately.

	Date	strategy_returns_bit	strategy_returns_gold	decision	commission
145	2016-09-12	-0.002050	0.000717	Half-N-Half	happen
146	2016-09-13	-0.003437	-0.001435	Half-N-Half	not happen
147	2016-09-14	-0.002562	0.008284	Gold	happen
148	2016-09-15	-0.002081	-0.001869	Half-N-Half	happen
149	2016-09-16	-0.001773	-0.004968	Half-N-Half	not happen

Therefore, we get our final result. If we add the commission fee in our calculation, then the condition should be (return of bitcoin - alpha2) - (return of gold - alpha1) > 0.05; buy 50%-50% bitcoin and gold when -0.05 < (return of bitcoin - alpha2) -

(return of gold - alpha1) < 0.05; buy only gold when (return of bitcoin - alpha2) - (return of gold - alpha1) < -0.05

Additionally, our return of each option will be (return of bitcoin) for bitcoin; (return of gold) for gold; and (return of bitcoin + return of gold)/2 for half and a half when commission does not happen.

(Return of bitcoin - alpha2) for bitcoin; (return of gold - alpha1) for gold; (return of bitcoin - alpha2 + return of gold - alpha1)/2 for half and a half when commission happens as indicated in the data frame.

Since the value of alpha1, alpha2 did not specify, the formal code cannot be formulated.

5 Model II: Logistic regression

In this part, we test the linear relationship between past performance and simulate a predicted future return of these two assets, and we finally decide an investment strategy based on the predictive results. We split the dataset into 75% Training dataset and 25% for Testing dataset.

Table: Logistic regression results

Linear regression indictor	**Bitcoin**	**Gold**
$Xstd\_5$	0.09	1.37e-02
$Xret\_5$	1.02	9.57e-01
$Xstd\_30$	-0.08	-1.06e-03
$Xret\_30$	-0.03	-6.47e-04
Beta	-0.00020	-8.16e-05

Table: Logistic regression results

Symbol	**Bitcoin**	**Gold**
MAE	0.030	0.007
MSE	0.002	9.02
RMSE	0.040	0.010

Explanation:

$Xstd\_5$-Standard deviation of return of bitcoin in a 5-day rolling window

$Xret\_5$-Mean of return of bitcoin in a 5-day rolling window

$Xstd\_30$-Standard deviation of return of bitcoin in a 30-day rolling window

$Xret\_30$-Mean of return of bitcoin in a 30-day rolling window
MAE- Mean absolute Error
RMSE- Root Mean Squared error
MSE- Mean Squared error

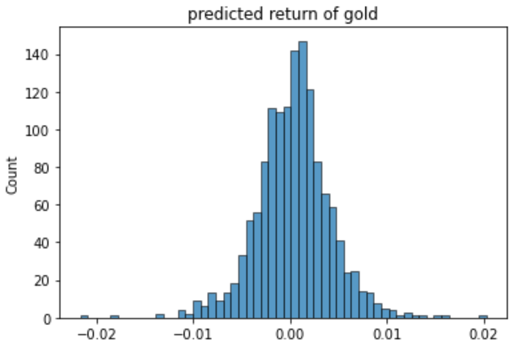

Figure: Histogram of predictions of gold

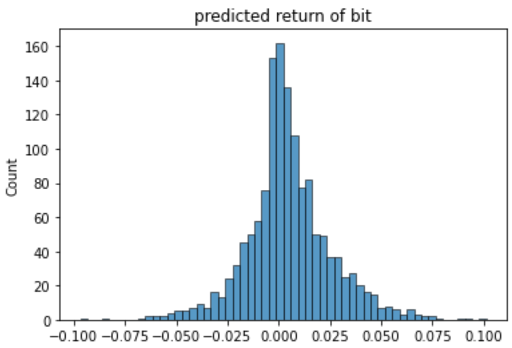

Figure: Histogram of predictions of gold

Therefore, we get our final result. If we add the commission fee in our calculation, then the condition should be by bitcoin when (return of bitcoin - alpha2) > (return of gold - alpha1) ; buy gold when (return of bitcoin - alpha2) < (return of gold - alpha1)

Additionally, our return of each option will be (return of bitcoin) for bitcoin; (return of gold) for gold; when commission does not happen; (return of bitcoin - alpha2) for bitcoin; (return of gold - alpha1) for gold when commission happens as indicated in the data frame. Since the value of alpha1, alpha2 did not specify, the formal code cannot be formulated.

6 Model III: Mean-variance Optimization Forecasting

6.1 Basic Indicators

We First test the correlation with the portfolio and get the correlation matrix which shows a 0.044087 correlation between bitcoin and gold.

	USD (PM)	Value
USD (PM)	1.000000	0.044087
Value	0.044087	1.000000

Figure: Correlation Matrix

Then we can generate the variance of the portfolio which is 0.001939

```
#Portfolio Variance
w = {'USD (PM)': 0.1, 'Value': 0.9}
port_var = cov_matrix.mul(w, axis=0).mul(w, axis=1).sum().sum()
port_var
```

```
0.0019388974638322601
```

Figure: Portfolio Variance

Then, according to Modern Portfolio theory, which is also a mean-variance theory, allows the user to maximize returns for a given risk level. MPT assumes that all investors are risk-averse, i.e., if there is a choice between low risk and high-risk portfolios with the same returns, an investor will choose one with low risk. MPT encourages the diversification of assets. It says that a high variance asset A if combined with diverse assets B and C, where A, B, and C have little to no correlation, can give us a portfolio with low variance on returns.

$$Portfolio\ Variance\ of\ N\ Assets:\ \sigma_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{i,j} = \sum_{i=1}^{N} x_i^2\sigma_i^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{i,j}$$

We import the bitcoin and gold's data,
Portfolio Optimization
Efficient frontier

$$min\{\mu w^T Vw - (1-\mu)\bar{r}^T w$$

$$\text{subject to } \sum_{i=1}^{n} wi = 1,\ \ wi \geq 0, i = 1,2,\cdots\cdots,\ n$$

$$\text{w=}[w1w2\cdots wn]^T, \bar{r} = [\bar{r_1}, \bar{r_2}, \cdots\cdots \bar{r_n}]^T$$

$$\text{r=}[r_1, r_2, \cdots\cdots\ r_n]^T, \bar{r} = E\{r\}$$

$$V=E\{(r-\bar{r})(r-\bar{r})^T\}$$

N is the number of risky portfolios, w_i is the portfolio weight that the asset allocation invested in security i, $\bar{r}$ is the expected security return, r is the actual security return, V is the variance of the expected return of security, and μ is the risk aversion factor of investor with the interval between $0 \leq \mu \leq 1$. In the optimization equation above, $\bar{r}^T w$ is the expected portfolio return, $w^T Vw$ is the portfolio total risk.

6.2 The Efficient Frontier

Based on the MPT and basic indicators, we can do many experiments to generate an efficient frontier.

```
Algorithm 2: Graph of the efficient frontier
Input: Four variables: r, vol, Wb, Wg (i=1,2, 3…)
Output: p (r, vol, Wb, Wg)
for i=1 to 10000 do
    Do enough large times of experiments
    Randomly select the weight of bitcoin and gold
    Graph the efficient frontier which is
    generated by different weights
    The output can be expressed p(r,vol,Wb,Wg)
end
```

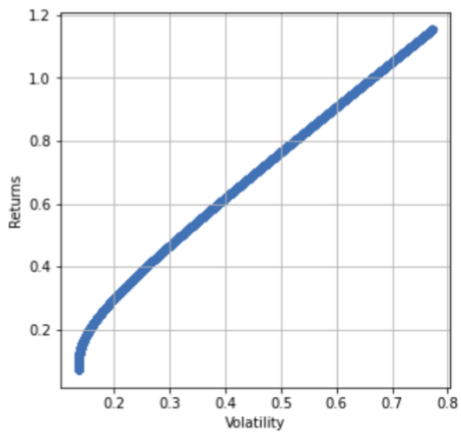

Figure: The Efficient Frontier

Alpha (α) is a term used in investing to describe an investment strategy's ability to beat the market or its "edge." In this model, we can appropriate α in the min-variance portfolio and optimal risk portfolio based on our randomly generated proportion model.

6.3 Optimal Risk Portfolio

If we want to get the highest sharp ratio, we need to set the risk-free rate as small as possible which can make enough large of the numerators and gets the highest sharp ratio.

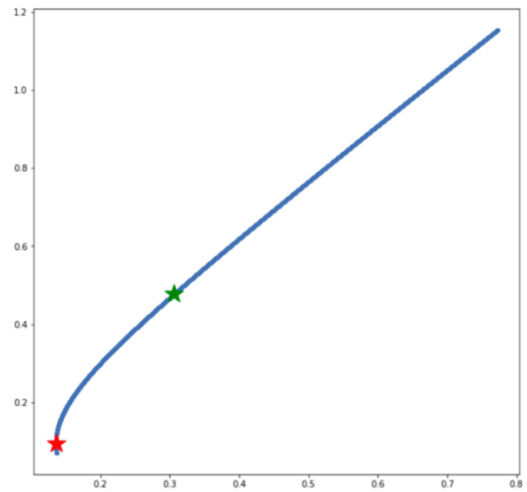

Figure: Optimal Risk Portfolio

The red point represents the smallest variance of the portfolio. If the investors want to stand the smallest risk, they will choose this point to invest. The smallest variance point has the lowest return and volatility which are 0.00947 and 0.1359. The corresponding wrights of bitcoin and gold are 0.9767 and 0.0233. Similarly, the green point represents the theoretical optimal position since it has the largest value of the first derivative of the curve.

```
Returns              0.476990
Volatility           0.306295
USD (PM) weight      0.624279
Value weight         0.375721
Name: 67693, dtype: float64
```

Figure: Weight of Optimization

The optimal portfolio constitutes a return and volatility are 0.4769 and 0.3063. The corresponding weights of bitcoin and gold are 0.6243 and 0.3757. Then, we make a comparison between the optimal risk portfolio and the min-variance portfolio.

	Optimal Portfolio	Min-Variance Portfolio
Return	0.4769	0.00947
Volatility	0.3063	0.1359
Weight of bitcoin	0.6243	0.9767
Weight of gold	0.3757	0.0233
Sharp-Ratio	1.557	0.069

Figure: Comparative Diagram

We can get the highest sharp ratio is 1.557 for the optimal risk portfolio and 0.069 for the min-variance portfolio. The difference between the return is 0.4674 and the difference in volatility is 0.1704. Therefore, one unit of the risk premium for investing in the min-variance portfolio is very small and the optimal portfolio can balance the risk and return.

9 Conclusion

9.1 Strength

- After all three models' testing, it will give us insights about the prediction of investment strategy from past data
- We have established a relationship between the proportion invested of different assets and respective volatility and return and consider many kinds of situations in line with the actual situation. Therefore, the model has excellent universality and flexibility
- We have developed a series of models and suggestions on the effectiveness of testing the performance of the main portfolio

9.2 Weakness

- The random forest model is strongly relying on past data's performance, and it could have interference by harming the overall data accuracy
- The commission fee is not a fixed amount and depends on each transaction amount during trading while given by the fixed percentage , so it can influence the factors behind applying directly into three models
- All three models can only forecast a future fixed-point performance, not a dynamic trading strategy based on the daily balance amount