Problem Chosen	2022	Team Control Number
C	MCM/ICM Summary Sheet	2211335

Summary

A good portfolio strategy is particularly important in financial markets full of uncertainties these days. It can not only to help us avoid risk, but also improve returns. Aiming at proposing the best strategies for traders, we establish a **Quantitative Investment Decision Model** in this paper.

Initially, since that we can only use the price data up to the day we give strategy, we need to preprocess the raw data. After unifying the format of data, we fill the missing data considering weekends and **"Holiday Effect"** through **LOWESS** and **X-13ARIMA-SEATS method**. After that, we begin our feature engineering, collecting 27 features, removing 13 highly-related features and leaving 14 features for gold and bitcoin separately. Then, we can devise our model based on these indicators.

For problem (a), to propose the best daily trading strategies and find out how much our initial $1,000 worth on 9/10/2021, we devise a Quantitative Investment Decision Model consisting of the following three submodels:

- **Assets Yield Forecast Model**: We first make prediction about the price and yield of gold and bitcoin based on the **Long-Short Term Memory (LSTM)** algorithm, so as to work out the total returns of our portfolio strategies and help verify our models.
- **Portfolio Optimization Model:** A good strategy must take both returns and risks into account. What's more, frequent trading brings unnecessary loss of service charges. Therefore, we establish the model based on **Mean-Variance Model** considering the transaction costs. Through this model, the optimal portfolio proportion is [0.2718573, 0.4759442, 0.2521985] in 9/10/2021. And at that day, our initial $1,000 worth **11,581.63 dollars**, that's more than a tenfold increase.
- **Traffic Light Signal Model:** In financial markets, judging the potential of a product can help us avoid risks and increase profits. To further optimize the model and give the best strategies, we use "traffic light" to measure the potential of assets intuitively based on **Logit** model. Then we work out that in 9/10/2021, the best portfolio proportion is [0.4192103, 0.2379721, 0.3428176], and our initial 1,000 dollars further appreciate to **13,614.33 dollars**.

For problem (b), we prove that our strategy is the best from two aspects: The prediction model is the best and the returns of portfolio are highest. We first use Mean-Square Forecast Error (MSFE) to verify that our prediction model is highly accurate. Then, we prove our portfolio model earns the highest returns by comparing it to other traditional portfolio models in terms of metrics Cumulative Yield, Sharpe Ratio, Annual Yield, and Max Drawdown Ratio. In these ways, we confirm our strategy reasonably.

For problem (c), we make sensitivity analysis on transaction costs based on a three-dimensional histogram we plotted. We find out that there is a negative correlation between returns and gold transaction costs, while returns are less sensitive to trading bitcoin.

Finally, considering the strengths and weaknesses of the model, we pointed out the direction of improvement in our future work.

Keywords: LSTM, LOWESS, X-13ARIMA-SEATS method, Mean-Variance Model

Contents

1 Introduction

1.1 Problem Background

Recently, people are paying more and more attention to finance and investment. When it comes to their goal, it is often to maximize their overall returns in the financial market.

Figure 1: Bitcoin, Gold or Cash?

Financial market is a complex system with all kinds of uncertainties. The price of financial assets is influenced by many factors, such as monetary and fiscal policies. Therefore, Investors are taking huge risks while gaining benefits. According to the data, the price of bitcoin reached nearly $20,000 in December 2017. At the end of 2017, the total value of the entire digital currency market dramatically reached $572.48 billion, with the cumulative growth of 3,028% over the year. However, since 2018, as governments around the world have strengthened the control of digital currency, while the security of the digital currency market has raised much concerns, the digital currency market see a sharp drop in market prices. It was only when the price slumped to nearly $6,000 that the digital currency market gradually began to improve.[1]

Thus, proper portfolio in a complex and uncertain financial system seems very important. The appropriate portfolio strategy can adjust the proportion of assets in portfolio according to actual situation, so as to avoid risks and raise returns.

1.2 Problem Restatement

Now we will begin with $1000 on 9/11/2016, to investigate a portfolio including cash, gold, and bitcoin [C,G,B] in U.S. dollars, troy ounces, and bitcoins respectively. Combining the conditions presented in the problem statement and the given datasets, we need to address the following issues:

- Devise a model simulating the best daily portfolio strategy. Then work out how much the initial $1,000 is going to be worth on 9/10/2021.
- Prove that the strategy we propose is optimal.
- Analyze the sensitivity of strategies to transaction costs.
- Find correlations among transaction costs, strategy and results.
- Write a memorandum to the trader, sharing our strategy, model and results.

1.3 Our Works

In order to present our ideas and work intuitively, we draw a flow chart shown in figure3.

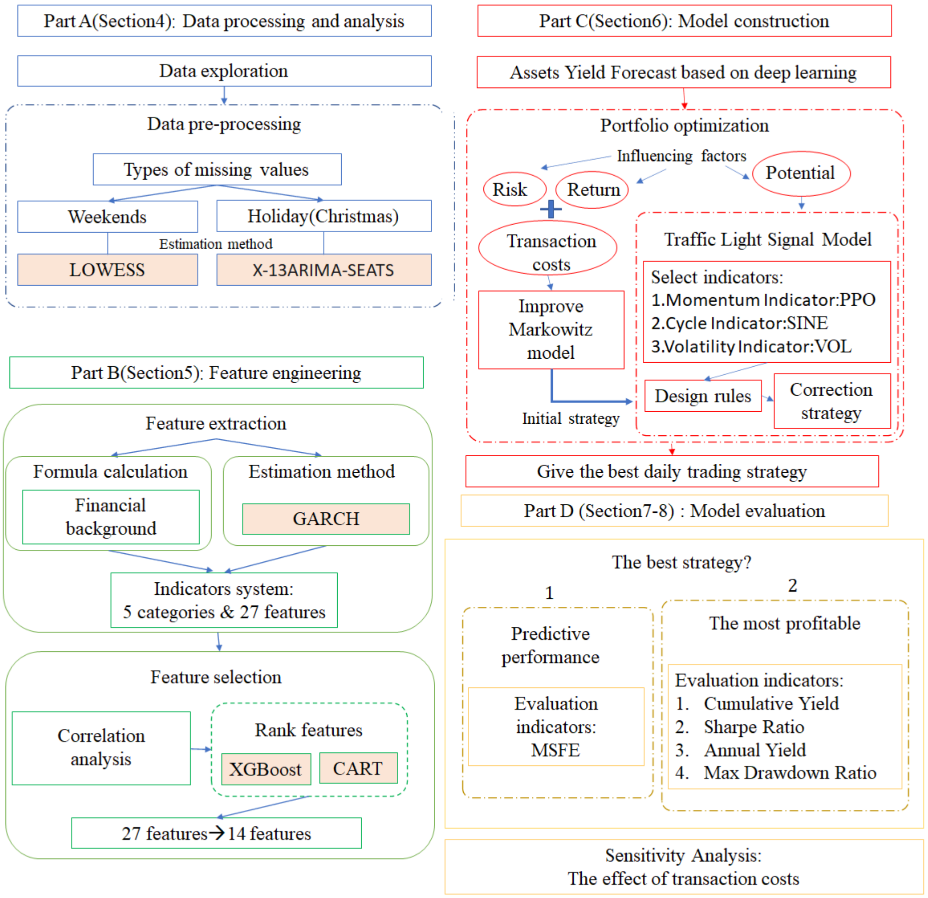


Figure 2: Overview of our work.

2 Assumptions and Justifications

We make the following basic assumptions in order to eliminate the complexity and simplify the problem. Each of our assumptions is justified and is consistent with the basic facts.

Assumption1 : We assume that there is no causality between gold market developments and bitcoin market developments.

Assumption2 : We assume that there are no other relevant random variable changes that affect the time series analysis.

Assumption3 : We assume that past trends of things will extend to the future.

Other assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate locations.

3 Notations

4 Data Pre-processing and Analysis

In this section, for purpose of extracting valid information while reducing the impact of interference messages, pretreatment to the raw data is adopted.

Table 1: Notations.

Symbol	Definition
Y_t	the asset yield on term t
μ	mean vector of yield
Σ	covariance matrix of yield
P_G	the price of gold at moment t
P_B	the price of bitcoin at moment t

4.1 Data Exploration

The originally data sets include two files, $LBMA - GOLD.csv$ and $BCHAIN - MKPRU.csv$, providing the daily price of gold and bitcoin respectively from 9/11/2016 to 9/10/2021.

At the same time, the Figure1 and Figure2 in 2022_MCM_$Problem$_$C.pdf$ show the trend of gold and bitcoin in terms of daily price respectively. The figures present lots of information.

- **gold**

  It is noticable that the figure for gold was fluctuating during the period from September 2016 to nearly June 2019. Then the price of gold climbed steadily and peaked at over 2000 U.S. dollars per troy ounce in about August 2020. It is said that gold has become an important haven when some serious global emergencies happen, including public health events. Therefore, the COVID-19 pandemic is the direct cause of the gold price increase. After that, from August 2020 to September 2021, the figure was undulating with a small slide.

- **bitcoin**

  When it comes to the price of bitcoin, it is clear that the price of bitcoin remained virtually unchanged at an extremely low price from September 2016 to July 2017.After that, the figure dramatically rose to approximately 20,000 U.S. dollars per bitcoin. Then, the price cut in half at the beginning of 2018, and remained stable during 2018 and 2019. According to the background information in section 1.1, governments around the world have strengthened the control of digital currency, while the security of the digital currency market has raised much concerns in 2018, resulting the slump of bitcoin price. Apart from this, 2020 and 2021 saw a dramatically increase from about 5,000 dollars to over 60,000 dollars since that many companies bought it to hedge against inflation. In the second half of 2021, there is a wild fluctuation in price.

4.2 Data Pre-processing

The data preprocessing in this paper can be divided into the following steps:

Step1. Format Unification

Since the formats of dates are different in files, some are ”month/day/year” while the other shown as ”year-month-day”, we unify all the date into the format of ”year-month-day” in order to simplify the follow-up work. Then, we number the data in the two files separately.

Step2. Missing value handling

There are two different kinds of missing value in datasets. One is because of the weekends, and just shown in the dataset of gold, while the other is missing maybe owning to ”Holiday Effect”. Therefore, we use different methods to handle them.

- Missing value for weekends

  For fluctuating data, simple linear regression often brings large deviations. However, **Locally Weighted Scatterplot Smoothing (LOWESS)**, can better solve this problem. **LOWESS** considers the influence of near and far time period compared with simple average treatment, giving greater weight to data near t, and less weight to data farther away. Moreover, the attenuation degree of weight can be controlled by the window width.

  Since the price of gold given in the dataset is fluctuating, **LOWESS** will be more intuitive and persuading in preprocessing it.

  Assuming that we need to predict the value in $\hat{T}_t$, we first select a series of time points near the moment t. Then, we can fit a polynomial using the following weighted least square method.

$$\hat{T}_t = \sum_{i=1}^{t} K\left(\frac{i-t}{h}\right)\left(X_i - \beta_0 - \beta_1 i - \cdots - \beta_d i^d\right)^2 \tag{1}$$

  where d is the order. Let d be 2. And $K(.)$ is the kernel function, usually taking the density function of the standard normal distribution

- Other missing value

  There are 10 missing value in LBMA-GOLD.csv except for weekends. They are

Table 2: Missing dates.

Date	
2016-12-23	2018-12-31
2016-12-30	2019-12-24
2017-12-22	2019-12-31
2017-12-29	2020-12-24
2018-12-24	2020-12-31

  We are surprised to find that these days are just coincide with Christmas. It seems that the lack of data is most likely due to the impact of Christmas. Missing values affected by holidays cannot be filled by conventional methods because of ”Holiday Effect”.[5]

  Holiday Effect refers to the impact on the global financial market resulting from global festival. For example, price can rise or fall sharply before big holidays such as Christmas in western and the Spring Festival in China. Therefore, we need to find a way to properly handle sequences that are susceptible to the holiday effect.

  In 2009, the U.S. Census Bureau officially launched **X-13ARIMA-SEATS method** based on X-11 method. Since that the X-11 method is a centralized moving average, it has difficulty in processing data at both ends of the sequence. However, **X-13ARIMA - SEATS**, introducing *regARIMA* to predict the extension of the sequence, can address this problem to some extent. This is also the latest adjustment method that can handle sequences with festival effect well in terms of fitting.[4]

In order to fill the missing value more precisely, We utilize the **X-13ARIMA-SEATS** algorithm in R package ”Seasonal”. The results are shown in table3.

Table 3: Filling results.

Date	USD(PM)	Date	USD(PM)
2016-12-23	1131.35	2018-12-31	1279
2016-12-30	1145.9	2019-12-24	1482.1
2017-12-22	1264.55	2019-12-31	1514.75
2017-12-29	1291	2020-12-24	1875
2018-12-24	1258.15	2020-12-31	1887.6

Finally, the price of gold in weekends and 10 missing items are filled. Now we have both 1824 items for gold and bitcoin.

5 Feature Engineering

Since we have finished our data-preprocessing, we now need to select meaningful features in order to send them into the algorithm and model for training.

5.1 Feature Collection

Given that we only have data regarding price, to be exactly the close price, we choose 27 features calculated only based on the close price from Internet.[6]

The selected features can be divided into five categories: **Basic Indicators, Moving Average Indicators, Momentum Indicators, Cycle Indicators and Volatility Indicators**, a total of 26 features shown in table4.

At the same time, by observing the time series chart of gold and bitcoin, it can be found that

1. The amplitude of the series fluctuation is not stable.
2. A large fluctuation is followed by another large fluctuation
3. Sequence ranges react differently to price increases or decreases sharply, and are more sensitive to price declines sharply

According to the reference book,[5] we can judge that the sequence volatility is not constant and there is a volatility cluster and leverage effect based on the above three points. S-ince that Volatility can well measure how much a sequence fluctuates,according to the above characteristics, we select the **GARCH** model to fit the volatility. The steps are as followed:

Step1. ARCH effect was tested to determine whether the sequence had conditional heteroscedasticity first. This test can be implemented by **ljung-box**, which is commonly used in financial time series.[5]In the LJung-box test, the P values of the gold sequence and the Bitcoin sequence are both less than 2.210^{-16}, and the null hypothesis is rejected at the level of 0.01, indicating that the gold sequence and the Bitcoin sequence have significant ARCH effect.

Step2. On the premise that the sequence has ARCH effect, we can obtain the estimation of volatility through GARCH(1,1). The GARCH(1,1) formula is as follows.

$$\sigma_t^2 = \omega + \alpha_1 Z_t^2 + \beta_1 \sigma_{t-1}^2, \quad Z_t = \sigma_t \epsilon_t, \epsilon_t \sim^{IID} N(0,1), \tag{2}$$

where σ represents volatility rate, Z_t represents new interest series, ϵ_t denotes residual series and ω, α, β denotes the parameters to be estimated.

Due to the importance of volatility in financial time series, we have also added it to the table4, building a system with 27 indicators.

Table 4: Indicators system.

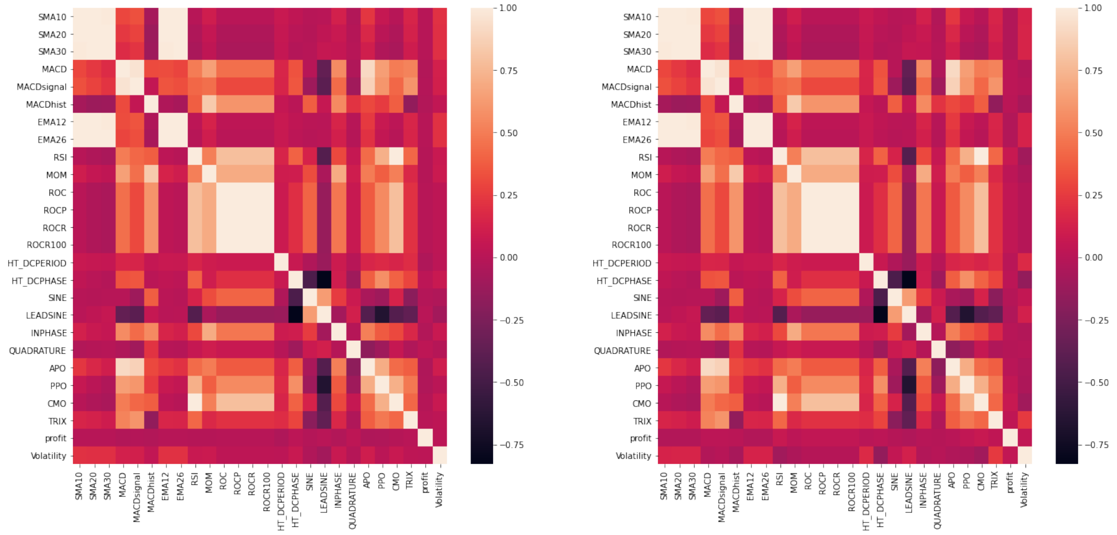
categories	features	definition
Basic Indicators	USD(PM)	The daily price
	SRR	Simple rate of return
Moving Average Indicators	SMA10	Simple moving average over 10 periods
	SMA20	Simple moving average over 20 time periods
	SMA30	Simple moving average over 30 time periods
	EMA12	Exponential moving average weighted 12
	EMA26	Exponential moving average weighted 26
Momentum Indicators	RSI	Smooth profit to loss ratio
	MOM	An upward trend value
	PPO	Percentage Price Oscillator
	APO	Absolute Price Oscillator
	CMO	Chande Momentum Oscillator
	TRIX	The one-day rate of three times smoothed EMA
	MACD	Moving Average Convergence Divergence
	MACDsignal	Moving Average Convergence Divergence signal
	MACDhist	Moving Average Convergence Divergence hist
	ROC	The rate at which prices change on a given day and on a given day before
	ROCP	Percentage of ROC change rate
	ROCP	ROC rate of change
	ROCR100	ROC rate of change in 100 ratio
Cycle Indicators	HT_DCPERIOD	Hilbert transform - dominant period
	HT_DCPHAS	Hilbert transform - Dominant cycle stage
	SINE	Hilbert transform - sine wave
	LEADSINE	Hilbert transform - lead sine wave
	INPHASE	Hilbert transform - The components of the Hilbert transform vector
	QUADRAT	Hilbert transform - The components of the Hilbert transform vector
Volatility Indicators	VOL	Volatility rate

- **Basic Indicators:** The basic indicator in this paper is the price and simple rate of return. All the other indicators are auxiliary indicators based on it.
- **Moving Average Indicators:** Moving averages are one of the most widely used tools for analyzing time series data. They can help traders identify current trends and judge future trends more reasonable.
- **Momentum Indicators:** These kinds of indicators help analyze short or medium term price volatility.

- **Cycle Indicators:** The operation of many cyclical industries has obvious ”cycles”, with obvious peaks and troughs. And the ups and downs of the industry are closely linked with the economic cycle. Judging the cycle can help traders avoid the trough.
- **Volatility Indicator:** Volatility indicators can reflect price changes, meanwhile helping investors measure market risk.

5.2 Feature Analysis

Next, we will draw thermal maps to test the correlation of extracted features shown in figure3.



(a) gold (b) bitcoin

Figure 3: Features’ thermal map.

Since that the lighter the color, the higher the correlation between features, showing the correlation coefficient almost 1; while the deeper the color, the less relevant the features are with the correlation coefficient nearly -1, it is clear that most correlation between features is weak and there is less overlap information between them. Therefore, we can judge that our feature extraction is very effective.

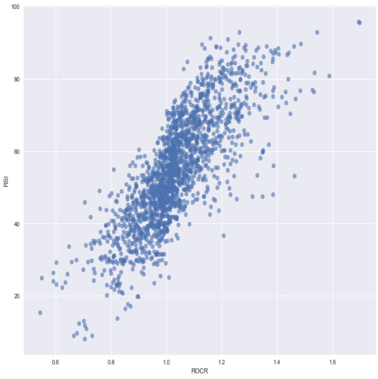


Figure 4: correlation between MCAD and EMA12.

However, there are still some features correlated with each other tightly, showing almost black in the thermal map. For example, the correlation between MCAD and EMA12 are very strong shown in figure4. Therefore, to reduce redundancy, we need to reduce the dimensions of these 27 features by removing highly relevant features.

5.3 Feature Selection

A good combination of data characteristics require not much to make the model perform well. However, the redundant features will increase the computational load of the model, resulting in the consumption of unnecessary computing resources and training time. There are many features that could be used to describe an object or an event, but in most cases, only a few attributes are needed to make a correct judgment. [1] Therefore, we now need to extract some of the most typical features to simplify the training.

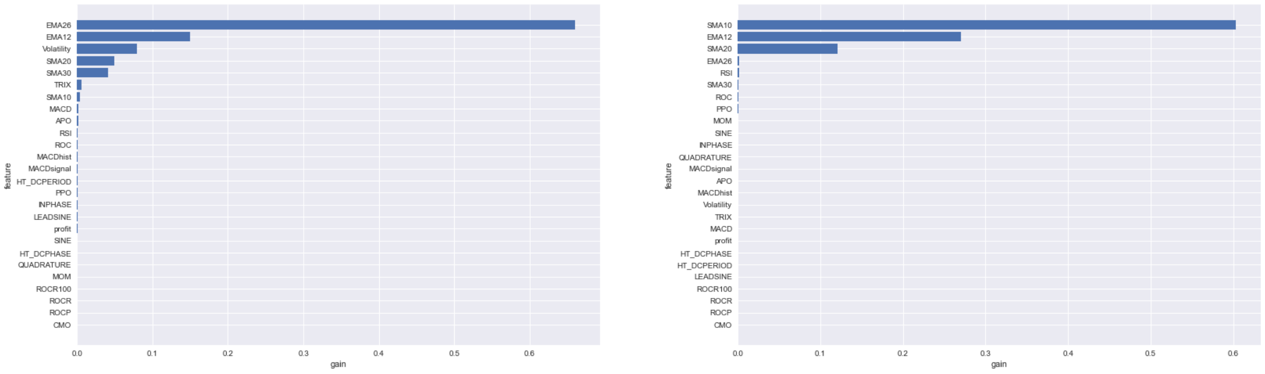
In this part, **XGBoost** algorithm is selected to extract features. **XGBoost** is one of the Boosting method that belongs to **ensemble learning**. The idea of **ensemble learning**, is to form a strong classifier by constructing several weak classifiers.

The core idea of **XGBoost** is simple: keep generating tree models, and keep growing trees through feature splitting. Every time it adds a tree, it learns a new function to fit the predicted residuals of the previous tree. Therefore, each tree is connected to each other in series. Finally, the prediction result is the sum of score of leaf nodes corresponding to K trees. Meanwhile, the Tree model used by **XGBoost** is the **Classification and Regression Tree (CART Regression Tree)**, which can integrate several CART Regression trees as weak classifiers.[1]

We first extract features with high correlation. The highly correlated features in gold thermal map and bitcoin thermal map are the same, they are 'MACD', 'EMA12', 'ROCP', 'MACDsignal', 'ROCR', 'RSI', 'CMO', 'ROC', 'SMA20', 'SMA10', 'EMA26', 'SMA30', 'ROCR100', and 'APO'.

Of course, we also need to preserve the most iconic features. Therefore, we rank these features regarding importance, preserving two features with the highest scores and removing others.

To rank the importance of features, **XGBoost** uses the split point of each feature in a single decision tree to improve the quantity of performance measures, so as to calculate the importance of feature attributes, which are weighted and recorded by nodes. Finally, the weighted sum of the results of a feature across all trees is averaged to get the final importance score.



(a) gold

(b) bitcoin

Figure 5: Features importance rank

We rank these highly correlated features' importance of gold and bitcoin respectively, the results are shown in figure5.

It is clear that EMA26 and EMA12 get the highest score in gold while SMA10 and EMA12 in bitcoin. Therefore, we reserve features EMA26 and EMA12 for gold and SMA10 and EMA12 for bitcoin, and remove the less important features.

Finally, the features we select are shown in table 5. We have removed 13 features for gold and bitcoin respectively, both leaving 14 features.

Table 5: Final Features.

asset	features
gold	SMA20, MACDhist, EMA26, MOM, HT_DCPERIOD, HT_DCPHASE, SINE, LEADSINE, INPHASE, QUADRATURE, PPO, TRIX, SRR, VOL
bitcoin	SMA10, MADChist, EMA12, MOM, HT_DCPERIOD, HT_DCPHASE, SINE, LEADSINE, INPHASE, QUADRATURE, PPO, TRIX, SRR, VOL

6 Quantitative Investment Decision Model

The key to making good investment decisions is to make predictions taking full advantage of information available, while the key to making accurate predictions is to exploring patterns from historical data.

In this section, aiming at providing the best daily trading strategy based only on up-to-day price data, we first build **Assets Yield Forecast Model** based on **Long-Short Term Memory (LSTM)**, predicting the price and yield while finding out how much our initial investment turn out to be worth. After that, we can set up **Portfolio Optimization Model** to get the portfolio every day. And finally, we devise a **Traffic Light Signal Model** to give the best strategies intuitively and help traders analyse by themselves.

6.1 Assets Yield Forecast Model

People tend to make investment decisions based on yield. Now we have the price data, so we define the formula of assets yield as follows:

$$Y_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{3}$$

where Y_t denotes the asset yield on term t, P_t denotes the asset price on term t. Then, we can predict not only the price but also the yield, thus making it more convenient to make decisions.

According to the requirement, we could predict the price of each day only based on **price data up to that day**. However, If all the data are put into the model for prediction, future information will be included in the process of parameter estimation, that is, the parameters will be affected by future, which will not meet the requirements. Therefore, inspired by the **Backtesting**, our prediction model adopts the strategy of **step by step prediction**.

6.1.1 The Structure of LSTM Model

We first divide to dataset into two parts, training samples and test samples. Since there are 1824 items in our dataset, we use the first 1000 items as a training set and the rest as a test set.

And now, we use **Long-Short Term Memory model** to transform the time series prediction problem into a supervised learning problem.

Long-Short Term Memory (LSTM) model, an improved algorithm of Recurrent Neural Network (RNN), shows strong applicability in processing financial time series data. L-STM has introduced a gate mechanism, which can control the accumulation rate of pre-information while choosing to forget part of the accumulated information. Therefore, it largely solves the problems of gradient disappearance and gradient explosion in simple RNN.

LSTM adds three logic control units on the basis of RNN: Input gate, Output gate and Forget gate. These three LCU are connected to a multiplicative element separately. By setting the connection weight between the memory unit and other parts of the neural network, the input and output of data flow as well as the state of memory cells are controlled. The specific concept figure is as follows.

The description of specific component in figure7(a) is as follows:

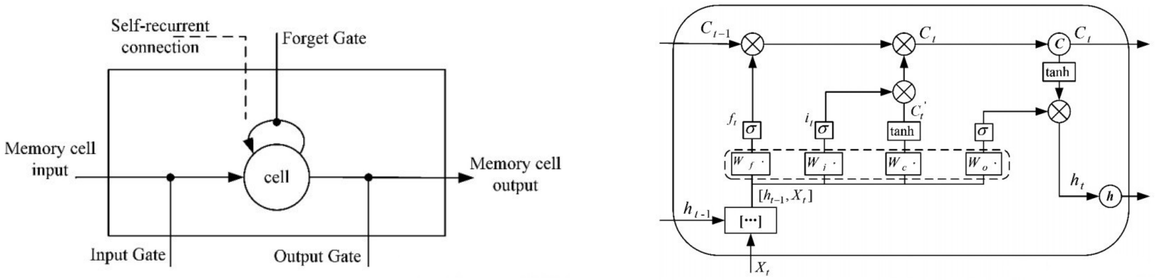
- **input gate:** Control whether information flows in, denoted as i_t.
- **output gate:** Controls whether the information of the memory cell at the current moment flows into the current hidden state h_t, denoted as o_t.
- **forget gate:** Control whether the information of the memory cell at the last moment is accumulated into the memory cell at the current moment, denoted as f_t
- **cell:** Memory unit, which represents the memory of the state of neurons, enables THE LSTM unit to preserve, read, reset and update long-distance historical information, denoted as c_t.

And at moment t, the formula of LSTM model is defined as follows:

$$\begin{cases} f_t = sigmoid\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \\ i_t = sigmoid\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ o_t = sigmoid\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \\ \tilde{c}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \\ c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \\ h_t = o_t \times \tanh\left(c_t\right) \end{cases} \tag{4}$$

According to the formula description, the details of LSTM are shown as figure7(b).

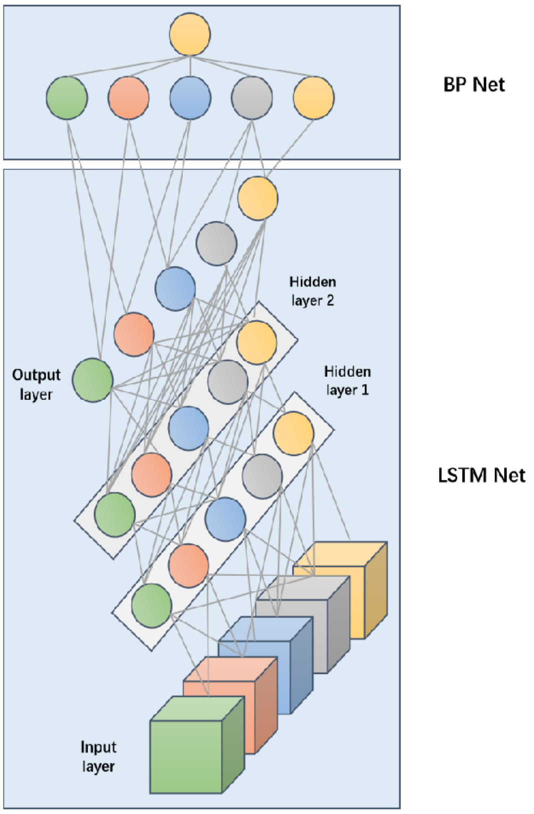
In the training process of LSTM neural network, the data features at moment t are firstly input to the input layer, and the results are output through the excitation function. Then the output results, the output of the hidden layer at $t-1$ and the information stored in the cell at $t-1$ are input into the nodes of the LSTM structure. Then the data is output to the next hidden layer or output layer through the processing of input gate, output gate, forget gate and cell unit. Finally, the results of LSTM structure nodes are output to the neurons in the output layer, while the back propagation error is calculated and each weight is updated.



(a) concept figure

(b) detail figure

Figure 6: LSTM Structure Figure.



(a) structure

Figure 7: LSTM Structure Figure.

6.1.2 Training Based on LSTM

In this section, we use LSTM to transform the time series prediction problem into a supervised learning problem. As we all know, the essence of time series prediction is to predict the time series value of $t + 1$ moment according to the observation data of the first t moment, which can be transformed into supervised learning problem in machine learning. That is, train a prediction model by using the previous samples to predict the new input samples.

The values of the hyperparameters set in the model are shown in the table 6.

After training our model based on the above hyperparameters, we get the loss function graph shown in figure8. Meanwhile, the light color with low transparency is the confidence interval, and the solid line shows the downward trend.

Table 6: Setting of hyperparameters.

hyperparameters	meaning	value
n_{in}	Lag period	1
n_{out}	Advanced prediction number	1
n_{vars}	Number of features	15(1 prediction target with 14 features)
n_{neuron}	Number of hidden layer neurons in LSTM	120
n_{batch}	Batch sizewhich is the number of samples selected for a training session	72
n_{epoch}	Training times of the model	100
$repeats$	Number of models being trained	5

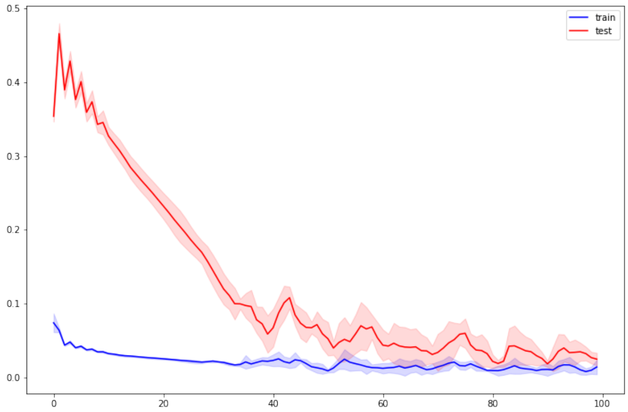


Figure 8: Loss function graph.

It is clear that the loss function tends to converge after 40 training sessions with small fluctuation, showing that the performance of our model is very good.

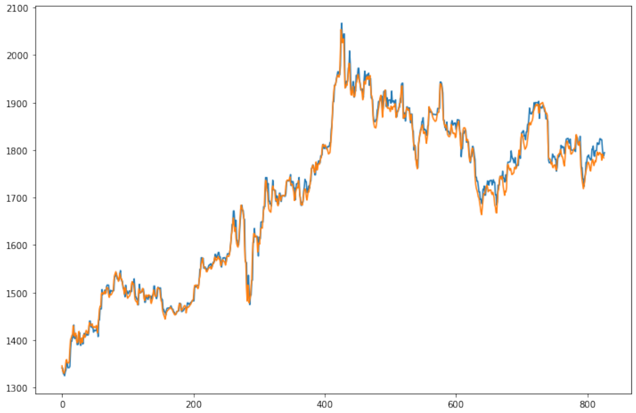


Figure 9: Yield prediction of gold.

Then, we apply our model to predict the yield of gold and bitcoin respectively. Considering we already have the real data, we let our predictive results compare with the real value, thus enabling us to judge the effectiveness of our model. We show the graph of gold for example in figure9.

In this figure, blue line represents real value while the yellow on represents predictive value. It can be seen from the picture that the two curves almost overlap with each other. Therefore, we can judge that our model is effectiveness with high accuracy.

6.2 Portfolio Optimization Model

So far, we have predicted the price and yield of gold and Bitcoin respectively. However, it is not the case that the higher the yield, the higher the return for the trade in reality. It also needs to take the **transaction costs** and **risks** into account. Since that investing in a single asset may have a high risk, and most investors in the market are risk-averse, preferring stable returns and trying to avoid high returns under high risks. Therefore, it is very essential to build a suitable portfolio.

In this paper, a portfolio of virtual currency Bitcoin, real asset gold and risk-free asset cash is constructed first. And then, we will optimize it for risk and transaction costs.

6.2.1 Portfolio Theory

In 1952, Markowitz put forward the modern portfolio theory, which was the first time that scholar used mathematical methods to study the asset allocation in portfolio.[7] In Markowitzs portfolio theory, returns and risks are characterized by means and variances, thus enabling the build of Markowitz Mean-Variance Model – a model with dual goal of maximizing returns and minimizing risk. This model is widely used in investment selection and portfolio allocation, and its applicability and effectiveness have been proved in practice.[2]

However, the transaction costs havent been taken into account in the model. In fact, frequent asset purchases or redemptions can lead to a high commission fee, reducing net profits. Therefore, in order to maximize the traders returns, we devise a Portfolio Optimization Model considering transaction costs based on Markowitz Mean-Variance Model.

6.2.2 Construction of Portfolio Optimization Model

We now hold a portfolio consisting of cash, gold, and bitcoin $[C, G, B]$ in U.S. dollars, troy ounces, and bitcoins respectively. And the commission for each transaction costs $\alpha\%$ of the amount traded, we let α_G represents the cost of gold and α_B for bitcoin. Let $\boldsymbol{tc}$ represents the transaction costs, we have

$$\boldsymbol{tc} = \begin{bmatrix} 0 \\ -0.01P_G \\ -0.02P_B \end{bmatrix} \tag{5}$$

where P_G denotes the price of gold and P_B denotes the price of bitcoin. And we can use yield to represent price through formula (3).

Then, aiming at maximizing returns and minimizing risks, we can determine the optimal weight of the portfolio through the following goal programming model.

$$\max_{w} \boldsymbol{\mu}^T (\boldsymbol{\omega} + \boldsymbol{tc}) - \frac{\gamma}{2} (\boldsymbol{\omega} + \boldsymbol{tc})^T \cdot \boldsymbol{\Sigma} \cdot (\boldsymbol{\omega} + \boldsymbol{tc}) \tag{6}$$

$$s.t. (1\,1\,1)\boldsymbol{\omega} = 1$$

Where γ is investor risk aversion coefficient showing as a constant, w represents the total returns. $\boldsymbol{\omega}$ is a 3×1 weight vector. As for $\boldsymbol{\mu}$, it is a vector representing mean yield for assets, while $\boldsymbol{\Sigma}$ is the covariance matrix.

After running the model, we get daily portfolio strategy of the training set. We select a period showing the strategy as an example in figure10.

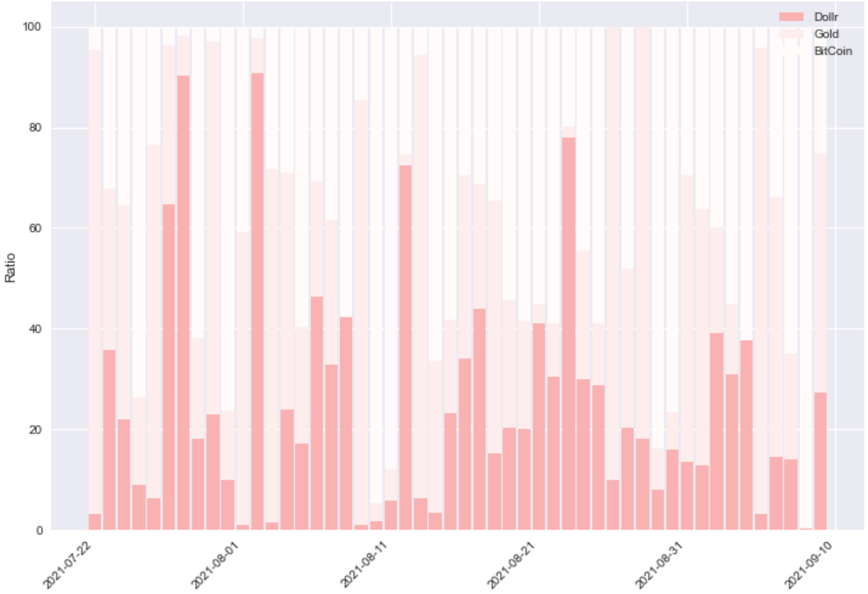


Figure 10: Daily portfolio strategy.

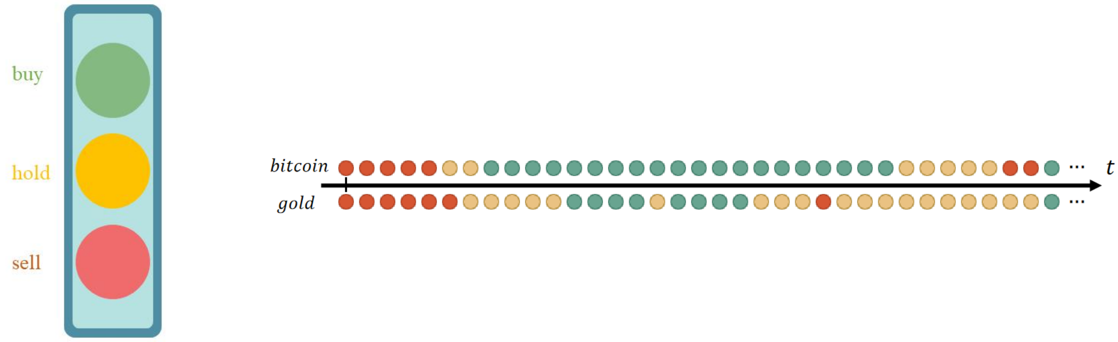
Through the figure 1, the price series of gold and bitcoin, we find that when the prices of gold or currency have a downward trend, the portfolio given by the model will become more conservative. For example, during the period from 2021-08-11 to 21, prices of gold and bitcoin are on a downward trend, while the cash weight in portfolio occupying bigger and bigger. Based on this model, the optimal portfolio proportion is $[0.2718573, 0.4759442, 0.2521985]$ in 9/10/2021. And at that day, our initial 1,000 dollars worth **11,581.63** dollars, that's more than a tenfold increase.

6.3 Traffic Light Signal Model

In financial markets, the value of assets is affected by many factors. For example, our world advocates sustainable development these days, so energy-saving and environmental friendly industries such as new energy vehicles may have a good prospect. When this phenomenon reflected in the financial market, it means that stocks of related companies have great potential. Therefore, the price is likely to rise sharply in the future. So if we can measure the potential of asset, the portfolio strategy will be even better.

In our paper, aiming at giving the best strategy and proposing the portfolio strategies more intuitively that every trader can analyse by themselves, we also build a **Traffic Light Signal Model** to measure the potential.In this section, we will demonstrate the judgement criteria.

In order to reflect the potential more intuitively, we use **traffic lights** to represent it. When we pass the crossroads, we can pass when the light is green, and we must stop when the light is red, we use green light to indicate low risk while red light indicating high risk. Therefore, green light further reflect that its a bull market now while red light showing that you are likely to lose a lot with your assets in a bear market. What's more, when the light is yellow, it would be wiser for traders to hold on their assets in hand.



(a) traffic light

(b) signal in series

Figure 11: Color signal.

Now we have proposed several features belonging to five categories already. In order to better indicate potential, we select features that are capable of measuring risks first, since that we have difficulties measuring potential only based on the previous data while the risks can better reflect potential. The higher the risks, the less potential the asset. Among the five categories,

- Momentum Indicators reflects the acceleration, deceleration and inertia in the process of fluctuation, which can reflect the change trend of the fluctuation degree;
- Cycle Indicators can reflect the law of price change;
- Volatility Indicators, can reflect the amplitude.

Therefore, these three categories can well indicate risks.

Since that there are many features that belong to these three categories, we choose the feature with the highest importance score to represent the categories respectively, **PPO** in Momentum Indicators, **SINE** in Cycle Indicators and **VOL** in Volatility Indicators.

To better measure the risks, we need to define the zones of the three colors. In our paper, we use **Logit** model to give the zones. **Logit** model, it's a discrete choice model and a common method of statistical empirical analysis with fast solving speed and convenient application. We will get the zones through the following process:

- **Input:** Time series before moment t of the three indicators **PPO**, **SINE** and **VOL**.
- **Classification:** rise or fall from moment $t-1$ to moment t.
- **Output:** The probability of rise or fall at moment $t+1$ through formula ().

$$P = \frac{\exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{1-\exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right)} \tag{7}$$

where $\boldsymbol{x}'$ is a matrix of the three indicators' time series and $\boldsymbol{\beta}$ is a parameter matrix.

Then, we can get the zones allocation based on the probability shown in table7. Based on these zones, we can use traffic lights to represent daily risk in a time series. We pick one of them shown in figure11(b).

But what can we do after seeing the color signal? Assuming we hold three assets cash, gold and bitcoin of the same amount now. Then we will give the principles under this circumstances as follows.

Table 7: Color Signal Zones.

assets	zone		
	red	yellow	green
gold	[0,0,3499)	[0.3499,0.3690)	[0.3690,1)
bitcoin	[0,0.5196)	[0.5196,0.5574)	[0.5574,1)

- **Green light:** Put a third of your cash in asset with green lights.

  For example, the price of gold is in the green light zone now, so we can put one third of cash to buy gold. Green light means that it is a good time to investigate or buy more. However, it doesn't mean there is definitely no potential at all. To be conservative, we choose an appropriate investment.

- **Red light:** Sell half of your assets into cash.

  It means that once the price of gold or bitcon lies in the red light zone, we need to sell half of it in exchange for cash. Red light demonstrated that if we hold on to the asset, there is a high probability of a loss. Therefore, it is wise for us to sell a proper amount of it. Considering bottom out, we still keep half of the asset.

- **Yellow light:** Maintain the portfolio allocation holding in hand.

  Yellow light means that the market is not very optimistic or negative now. To avoid unnecessary transaction costs, it's better to hold the asset.

Based on this this model, we are surprised to find that our returns has increased with the warning of the risk. In 9/10/2021, the best portfolio proportion is **[0.4192103, 0.2379721, 0.3428176]**, and our initial 1,000 dollars appreciate to **13,614.33** dollars, more than the return based on the Portfolio Optimization Model we established . Considering we only hold two goals, one for the biggest returns and one for the lowest risks, there may be something we haven't taken in to account. So the result is reasonable and exhilarating.

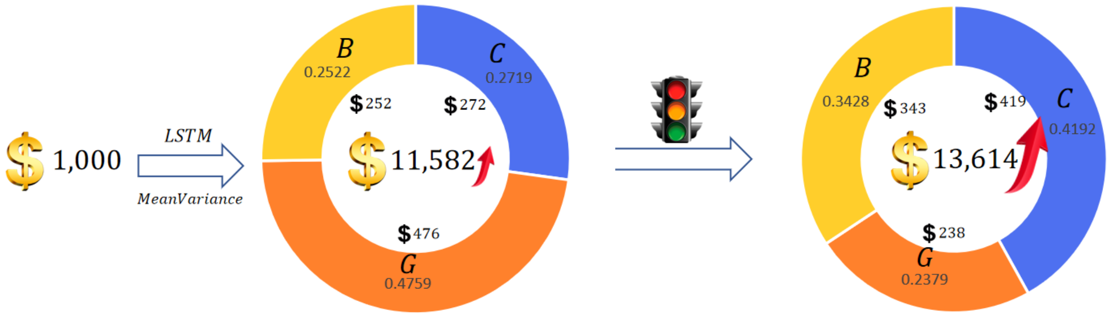


Figure 12: How much our initial cash worth?

7 Model Verification

In this section, we will provide evidence to prove that the strategies we proposed are the best. We prove it from two perspectives:

- **The prediction model is optimal.**

The result of investment portfolio is greatly affected by the forecast result, so the accuracy of forecast is very important.

- **The returns of portfolio are optimal.**

  People tend to invest with the goal of maximizing returns. Therefore, the best strategy is always the one that maximizes the return of the portfolio.

Then we will discuss these two aspects respectively.

7.1 Prediction Model Verification

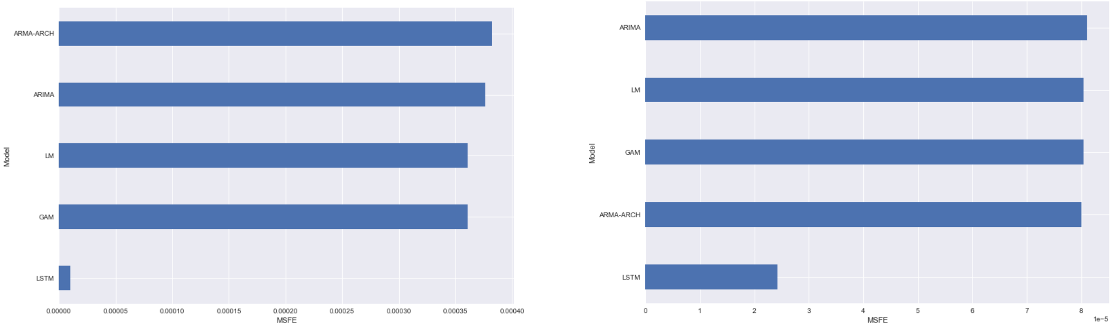
An optimal prediction model, means that it has the lowest error and the highest accuracy. Therefore, we use Mean-Square Forecast Error(MSFE) to measure the accuracy of our model. The formula is shown as follows:

$$MSFE = \frac{V_{predict} - V_{real}}{L} \tag{8}$$

where $V_{predict}$ denotes the predicted value of one stage, V_{real} denotes the real value of one stage and L denotes the length of this prediction interval.

Based on this formula, we get the MSFE value of five prediction models: ARMA-ARCH, ARIMA, LM, GAM, and LSTM model we used.

The value for predicting gold yield and bitcoin yield is shown in figure13.



(a) bitcoin MSFE

(b) gold MSFE

Figure 13: MSFE of different prediction models.

It is clear that LSTM has the lowest MSFE definitely, far lower than other models. So we have verified that our prediction model is the best by showing absolutely high accuracy.

7.2 Portfolio Returns Verification

In order to prove that our portfolio model based on **LSTM** and **Traffic Light Signal Model** earns the best returns, we compare it to other traditional and widely-used models such as ARIMA - MeanVariance Model or LM - MeanVariance Model.

After training the initial portfolio by both our model and reference model, we select ”Cumulative Yield”, ”Sharpe Ratio”, ”Annual Yield”, and ”Max Drawdown Ratio” as indicators to carry out visual comparative analysis on the effect.

- **Cumulative Yield**

The trader's goal is always to maximize returns, and the cumulative yield is the best indicator to reflect returns. The higher the cumulative yield, the better our model has trained, and the better our model proved to be.

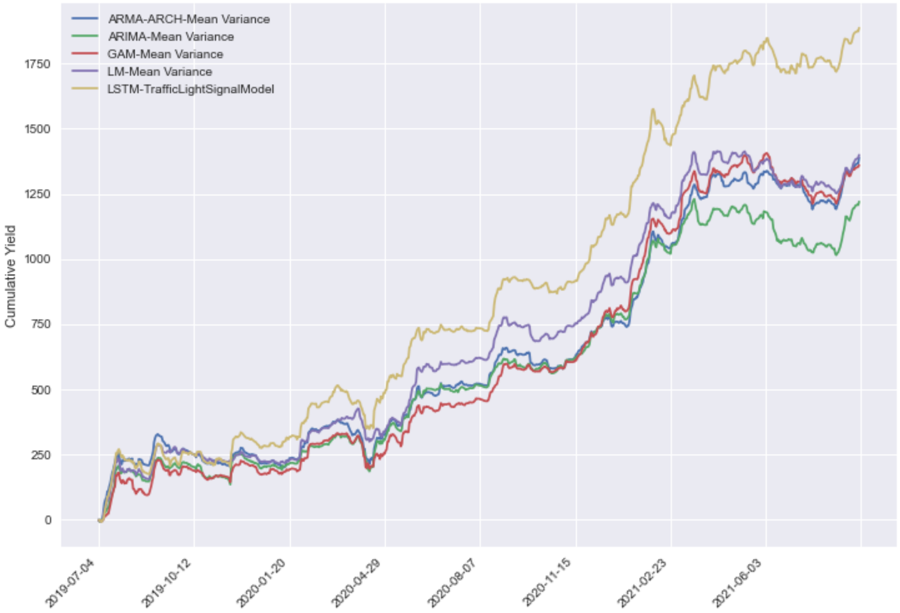


Figure 14: Cumulative Yield Graph.

By summarizing the current-time yield of each asset in the portfolio, we can get the cumulative yield of the model.

As can be seen from the graph, the trend of these models are similar, just showing differences in cumulative yield. Among all models, it is clear that the cumulative yield of LSTM model with Traffic Light Signal model is the highest one finally. And the values of other models are very close, showing still a long way to our model.

- **Sharpe Ratio**

  It is not enough to just look at the return of investment, but also consider the risk to bear. The Sharpe Ratio measures how well a trading strategy can make a return facing one unit of risk. The higher the value, the stronger the risk resistance of model, and the better the investment strategies. The formula is

$$SharpeRatio = \frac{E\left(R_p\right) - R_f}{\sigma_p} \tag{9}$$

where $E\left(R_p\right)$ denotes the expected rate of return on a portfolio, σ_p denotes the standard deviation of the portfolio and R_f denotes the risk-free rate.

As can be seen from the figure, there are not much differences among these models.

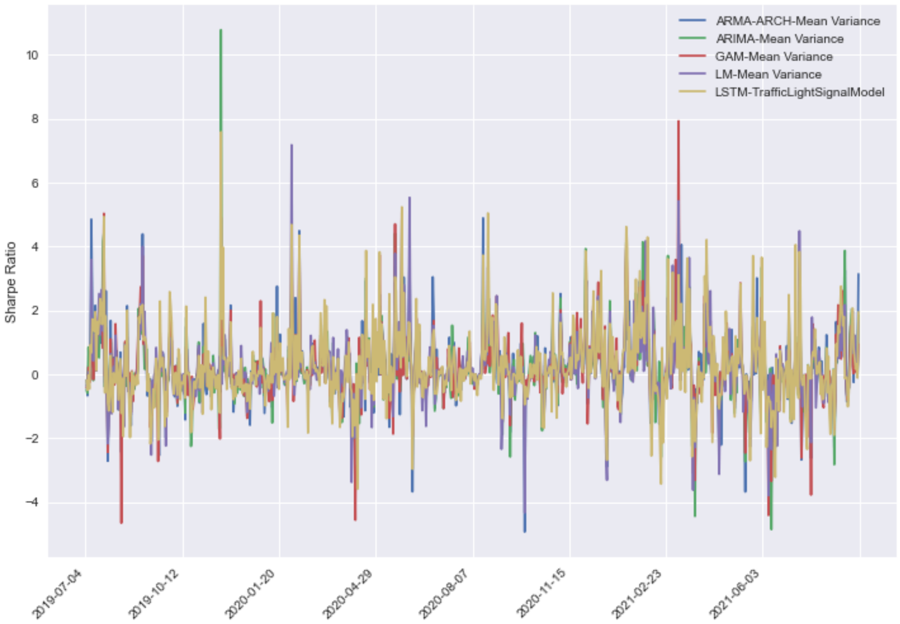


Figure 15: Sharpe Ratio Graph.

- **Annual Yield**

  Annual yield is the annualized value of the return in the current period of time, which is a common index to objectively measure investment income.[8] The formula is as followed.

$$R = \left(\frac{F}{I}\right)^{\frac{1}{n}-1} \tag{10}$$

  where I denotes our initial fund, F denotes our final assets worth, R represents the rate of interest and n represents number of period.

- **Max Drawdown Ratio** Max Drawdown Ratio is used to measure the maximum rate of loss sustained by an investment strategy during the entire trading process. The smaller the ratio is, the stronger its ability to cope with risks is. The formula is as follows.

$$Max\ Drawdown\ = \frac{Max\left(P_x - P_y\right)}{P_x} \tag{11}$$

The table8 shows the value of Annual Yield and Max Drawndown Ratio of different models. It is noticeable that both the Annual Yield and Max Drawndown Ratio of LSTM-Traffic Light Signal model rank first among all the models. Therefore, we can judge that our model can get the biggest returns.

To sum up, based on section7.1 and section7.2, we verify that our model is the optimal model, and further prove that our strategy is the best.

Table 8: Value of Annual Yield and Max Drawdown Ratio

model	Annual Yield	Max Drawdown Ratio
ARMA-ARCH-MeanVariance	0.0091	86.32
ARIMA-MeanVariance	0.0089	111.67
GAM-MeanVariance	0.0091	110.8
LM-MeanVariance	0.009	95.16
LSTM-TrafficLightSignalModel	**0.0095**	**115.21**

8 Sensitivity Analysis

In this section, we make the sensitivity analysis about the parameter α, the transaction costs, based on the portfolio optimization model we established in order to find out how this parameter affect the returns and strategies.

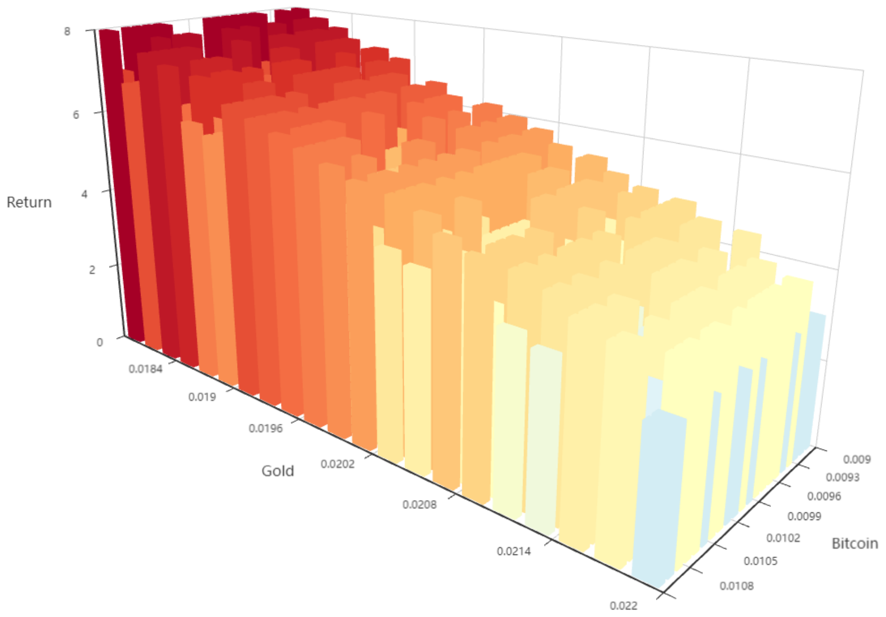


Figure 16: The impact of transaction costs on returns.

Since that the returns will be influenced by both the transaction costs of gold and bitcon, we plotted a three-dimensional histogram (figure16), showing the effect of changing parameter group $(\alpha_{gold}, \alpha_{bitcoin})$ on total returns on 10/9/2021.

We discuss the sensitivity of α_{gold} and $\alpha_{bitcoin}$ respectively, while giving the strategy adjustment measures:

- α_{gold} : As can be seen from the picture, there is a negative correlation between returns and α_{gold} overall. The lower the transaction costs, the higher the returns. So the returns are sensitive to α_{gold}. Under this circumstances, it is wise not to buy or sell the gold too often or at the time when the fees are high, thus avoiding unnecessary costs.
- $\alpha_{bitcoin}$: It is noticeable that returns are less sensitive to bitcoin, that is, the returns doesn't change a lot as the transaction costs of bitcoin trading change. Therefore, with regard to trading strategy, there is no need to consider transaction costs too much.

9 Model Evaluation and Further Discussion

9.1 Strengths

- **Variety** The features extracted in this paper are diverse, which greatly improves the effect of deep learning algorithm.
- **Velocity** The model proposed in this paper can give the portfolio strategy in time, which is particularly important in the financial market.
- **Veracity** The LSTM model used in this paper is far more accurate than other prediction models.
- **Value** The Quantitative Investment Decision Model used in this paper can get more effective strategies to increase investment returns, which has a wide range of practical application value.

9.2 Weaknesses

- We ignore the correlation between the gold and bitcoin markets, which may result in a matter of error
- We can not get more features due to the limitation of data set. Getting more features will make the deep learning algorithm better trained
- Our strategy only considers the interests of investors and ignores their own preferences.

References

[1] Song Wenzhu. Digital currency portfolio strategy research [D]. Nanjing Information Engineering University, 2019.

[2] Wang Yue. Research on stock portfolio Construction based on LSTM neural Network [D]. 2021.

[3] Wang Shuping, ZHU Yanyun, Wu Zhenxin. Analysis of seasonal Fluctuation of wheat price based on X-13A-S method [J]. Journal of management science in China, 2014, 22 (S1) : 22-26.

[4] Sax C, Eddelbuettel D (2018). ”Seasonal Adjustment by X-13ARIMA-SEATS in R.” Journal of Statistical Software, 87(11), 1-17.

[5] Ruey S. Tsay. An Introduction to Analysis of Financial Data with R [M]. Wiley, 2012.

[6] Financial indicators
https://reference.wolfram.com/language/guide/FinancialIndicators.html.zh?source=footer

[7] Markowitz H. Portfolio Selection[J]. Journal of Finance, 7(1):77-91.

[8] Jiao Yuming. Deep reinforcement learning for stock portfolio management [J]. Northwestern University,2021.

[9] He X. Virtual currency portfolio strategy design based on PLSTM-ATT [D]. Shanghai Normal University,2021.

10 Memo

Memorandum

To: Trader

From: Team 2211335

Subject: The Best Trading Strategy.

Date: February 22, 2022

Dear traders,

Proper portfolio shows particular importance in a complex and uncertain financial system these days. To better allocate assets in portfolio, quantitative research seems very essential. As response to your requirement, we are very delighted to have the opportunity to introduce our research and suggestions to you, with the hope that it may give you some insights of the future strategies.

In our research, a Quantitative Investment Decision Model consisting of three submodels for different purposes is devised. And we have verified its effectiveness comparing with other traditional models.

- **Assets Yield Forecast Model based on LSTM:** Predict the yield of different assets.
- **Portfolio Optimization Model based on Mean-Variance Model:** Give portfolio strategy considering returns, risks and transaction costs.
- **Traffic Light Signal Model based on Logit model:** Give the best strategy further taking potential into account.

Based on our models, the best strategy in 9/10/2021 is holding the portfolio with the proportion of [0.4192103, 0.2379721, 0.3428176], and our initial $1,000 appreciate to $13,614.33. It's quite inspiring.

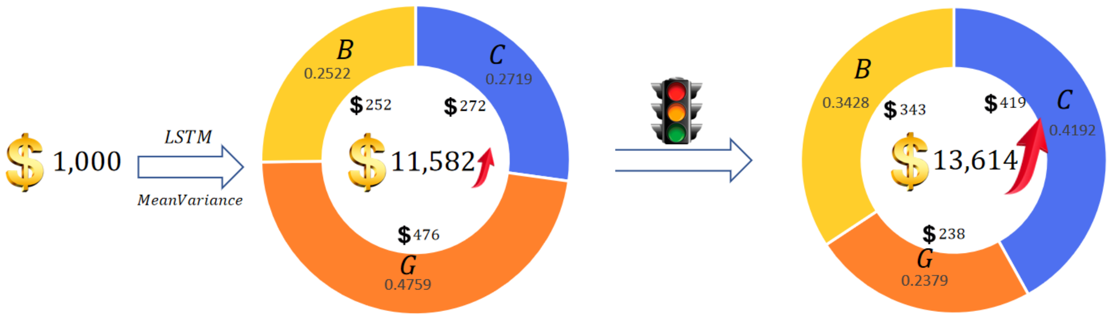


Figure 17: How much our initial cash worth?

There are still some tips for you.

- **Don't buy or sell the gold when the transaction costs are high.** In our analysis, the total returns show strong sensitivity to the transaction costs of gold trading. That is to say, the lower the transaction costs, the higher the returns. So it's wise for you to buy or sell gold when the transaction costs are low. In contrast, the total returns seem insensitive to the transaction costs of bitcoin trading. So when to trade bitcoin, it's up to you.

- **Don't buy or sell the assets too often.** According to the market rules, you need to pay a commission fee every time you buy or sell an asset. Therefore, frequent trading will bring an unnecessary expense. So how to make accurate choices of trading timing? We believe that our models will help.

We appreciate your time and consideration to give us the opportunity sharing our models and strategies to you. We sincerely hope that our study will benefit you.

Sincerely,

Team 2211335