Problem Chosen	2022	Team Control Number
C	MCM/ICM Summary Sheet	2221608

Investing is More than Gambling

Summary

For market traders, investment is one of the fundamental trading strategies. It includes money income or exchange of any other wealth that can be measured in terms of the value of money in order to achieve future value increase. Meanwhile, how to maximize the output of the investment has also been the primary concern in economics.[1]

Faced with two popular financial products — gold and bitcoin, our group aims to build a mathematical model that integrates financial knowledge and algorithms, which enables us to witness the birth of a new 'man of wealth'.

Before succeeded in building up a fairly good model, we took a lot of detours by trying LSTM(Long Short-Term Memory network). We summed up our lessons and expounded why it does not suit well.

From the very start, with little data to suggest both risk and potential profit, we compared two distinct choices. One is to hold on for a couple of days, while two is to invest randomly according to random walk theory.

As time goes by, the amount of historical data for reference keeps increasing. Then we are able to adopt ARIMA(Autoregressive Integrated Moving Average) model and apriori algorithm. They are both used to assist the moving average criterion, which has the best effect among them when working individually.

To examine the effect of our model, besides the three regular control groups, we introduced two more with different assumptions and matched algorithm. Group one bases on a God-like perspective, which means we know the whole process from the beginning to make full consideration. DP(Dynamic Programming) is capable to achieve such global optimum required upon. Group two suppose that we can forsee the price changes the next day to support our decisions. Greedy algorithm meets this demand perfectly.

Finally, we changed the transaction commissions of Bitcoin and gold respectively, re-conduct the investment simulation, compare the simulation results with the previous results, and analyze the sensitivity of the transaction strategy to the transaction cost.

Keywords: Moving average; ARIMA; Apriori algorithm; LSTM; Random walk theory; Greedy algorithm; Dynamic programming.

Contents

1 Introduction

1.1 Background

Bitcoin is a kind of digital virtual currency, which was first officially proposed by Nakamoto in 2009. On January 12th, 2009, world's first bitcoin transaction took place. Only until 2017, a large number of investors from the United States, Japan, South Korea and some other countries poured into the bitcoin market, making the price of bitcoin soar continuously. Bitcoin became a hot topic overnight.[2]

Nowadays, bitcoin is often regarded as a global safe-haven asset like gold. In the era of global turmoil, even gold has become unreliable and might be confiscated like India. Some people have begun to think of the feasibility and probability for bitcoin to replace gold, as in addition to the reserve capacity of gold, bitcoin also has some advantages over gold, such as low handling fee, rapid transfer ability, decentralization, etc. With the increasing popularityand the decreasing volatility of bitcoin, the status of gold has been gradually under threaten.

Meanwhile, gold enjoys a long history, almost as long as the length of human civilization, which has stood the test of countless tests throughout time. Bitcoin, by contrast, is far younger, not to say that its value has experienced a leap. With no entity to back it up, it is hard to foresee whether bitcoin will still be around in a decade, a century or even a millennium like gold.

Therefore, for investors to make as much profit as possible, it is necessary to take various factors into consideration, gain an insight into the market, make full use of algorithm tools and ensure every single investment reasonable.

1.2 Question restatement

In the set situation, we are asked to play the role of traders to develop a model suitable for investment in a market with two common finantial products. Price data accurate down to each market day over the 5-year-long period of the transaction together with the commission for separated product's each transaction are provided, while daily decisions should be made only on the basis of the day's and previous data. At the start all 1000 dollars we have are cash in hand.

- Section A

  Check the amount of money that can be earned over the five-year period of investment relying on the model and strategy.

- Section B

  Describe the theories and superiorities of the model. Using the result to prove itself.

- Section C

  Determine the model's sensitivity when it comes to varying transaction costs. Find out whether the strategies and results changes frequently according to the costs.

- Section D

  Design a memorandum with discription of strategy, model and results to persuade the trader on trusting the model.

1.3 Our work

Our main goal is to predict the value trend of Bitcoin and gold, explore the correlation between data, make reasonable investment planning based on prediction and association rules, try our best to get the most return, and then explore the sensitivity of transaction strategy to transaction cost.

In order to solve these problems, first of all, after we found that LSTM and other machine learning models had over fitting problems in predicting the value trend, we chose to use ARIMA time series prediction model to predict the value trend.

Then we used the apriori algorithm to mine the association rules between the data, conducted investment simulation based on the moving average, changed the transaction commission of bitcoin and gold, and compared the investment results to analyze the sensitivity of the transaction strategy to the transaction cost.

Finally, in order to prove the effectiveness of our model, we assumed two circumstances, one is to accurately know the value of the bitcoin and gold the next day, the other is to accurately know daily value of the Bitcoin and gold in the five years. We used the greedy algorithm and dynamic programming respectively, and compared the investment results with the results of our model.

After the huge failure in adpoting LSTM, we were finally able to recognize what to blame. The conclution is as follows.

- **Inaccuracy of LSTM in price prediction**

  Since the 1990s, scholars from various countries have tried various methods to predict the time series of financial products, including statistical methods, econometrics models, artificial intelligence and machine learning.Among them, time series analysis methods gradually emerged after entering the 20th century. In 2003, American economist Robert.F.Engle and British economist Clive W.J.Granger used statistical methods (ARCH model) to study economic time series, and finally won the Nobel Prize in Economics. Since then, many mathematical models of statistics and econometrics have been used to predict the time series of financial products. The main research methods include: regression analysis method, moving average method, periodic change analysis method and seasonal change analysis method.[3]

  Time-series prediction is a relatively difficult class of predictive problems. Unlike common regression prediction models, the "sequence dependence" between the input variables adds complexity to the time series problem. Financial time series are characterized by high noise and nonlinearity, and neither the traditional multiple regression nor the linear regression are applicable for the analysis. Such a non-linear sequence can only be effectively described in advanced mathematical models or machine learning ways. At present, deep learning technology and big data technology are constantly applied and developed in the field of financial analysis. Our group also decided to try to get our own investment plan with deep learning.

  A neural network that can be used to handle sequence dependence is called recurrent neural network(RNN). However, the RNN is affected by short-term memory. If a sequence is long enough, it will be difficult to transfer information from earlier to later time steps.During backpropagation, the RNN faces the problem of gradient extinction.

The Long Short-Term Memory Network, first proposed by Sepp Hochreiter and Jurgen Schmidhuber in 1997, is a novel deep machine learning neural network built on the RNN. The emergence of this neural network remedies the problem of RNN existence.The control process of the LSTM is similar to the RNN, which all process the data flowing through the cells during the forward propagation, varying in the structure and operation of the cells in the LSTM. LSTM networks can be used to build large-scale recurrent neural networks to deal with complex sequence problems in machine learning with good results.

The core concept of LSTM lies in the cell state as well as the "gate" structure.The cell state is equivalent to the path of information transmission, allowing the information to be transmitted in a sequence link. We can consider it as a way of memory for neural networks. In theory, cell states are able to convey relevant information during the sequence processing process.Thus, even information from earlier time steps can be carried into cells with later time steps, overcoming the effect of short-time memory. The addition and removal of information we achieve through the "door" structure, which will learn what information to save or forget during training. Once an input sequence is received and processed, the gates in the model use S-type activation units to control whether they are activated, thereby changing the model status and adding information(memory) to the model.

One activation cell has three gates:

Forget Gate: Decide which information to discard.

Input Gate: Determine which values are entered to update the memory status.

Output Gate: Determine the output value based on the input and memory status.

Each activation unit is like a mini-state machine, and the weights of each door in the cell are obtained through training. Based on the widespread application of LSTM in time series prediction, our group tried to apply it to price predictions for gold and bitcoin and hoped to make investment decisions based on the predicted prices. After testing, we regretted to find a significant lag in the predictions trained with only historical price data.We constantly adjust the parameters and even used the idea of reinforcement learning. We even update the prediction models daily according to the new data to make new predictions. The resulting prediction is a time series where the change trend lags by about one day.[4]

Ultimately, we have reached the following conclusions: Although the LSTM neural network model can partly achieve the prediction of future price trends, only a significantly lagging prediction sequence can be obtained if the training is based only on a single closing price in the past.That is because the past closing price is just one of the many factors influencing the changes in the financial time series.Even from the perspective of trading only, the highest daily transaction price, minimum transaction price, transaction volume and number of transactions in the past are all influencing factors affecting the price of future financial products.In addition, the overall economic situation and political situation of the market are all important influencing factors.If only the historical data is used for reference, the prediction of the financial time series directly shows a trend to follow the input data.Thereafter, we differentiated the data and were not ideal for the prediction of financial time series fluctuations.

2 Assumptions

In order to simplify the solution of the problem, we need to make appropriate assumptions within reasonable limits. To the extent that the model is applicable, these assumptions have been rigorously demonstrated.

- Cash is the only medium of exchange, which means that we believe commission is required for both sides in the transactions between bitcoin and gold.
- Both external environment and our investment will have no impact to the market.

3 Abbreviation and Definitions

We begin by defining a list of nomenclature (symbols) used in this article.

Table 1: Nomenclature

Symbol	Definition
y_{t-s}	price of product s days before today
θ_s	weight for a certain parameter
u	constant
e_t	error

4 Model Explanation and Algorithm Design

4.1 Inchoate Strategy

According to random walk theory, fluctuations in product prices are unpredictable. There are various of sayings about investment that we've all heard, such as, 'The best option is no action.' or 'Making random choices is the only way to deal with the random market.' With little data to support the algorithm, our team chose to put the two ways into practise — no investing or randomly invest.

4.2 Other Statements

To simplify the circumstance, when making a decision to buy or sell, our subject will always be the entire fortune. So the state of property([C, G, B]) in our model will only have three conditions: [w(t), 0, 0], [0, w(t), 0] and [0, 0, w(t)], where w(t) means our total wealth.

According to the common sense as well as the commission difference, we can affirm that price changes of bitcoin are larger in scale than that of gold. To pursue more wealth, bitcoin will always be our first choice to invest. Only when the prediction tells us that bitcoin may suffer a big fall and gold may experience a considerable rise than we will turn to focus on the latter.

4.3 Compound Criterion

To make decisions on buying or selling bitcoin, we first refer to 11 days and 22 days moving averages change. As long as there is an intersection, ARIMA prediction tells us that there won't be a sharp increase/decrease tomorrow, and the number of transactions in the last seven days is less than 3, we choose to sell/buy. If we have cash in hand, and moving average criterion tell us that a sharp increase of gold is likely to take place in the next week, we choose to buy. If the criterion indicates that we'd better buy bitcoin, but we have gold in hand and it is not a trading day, we wait till the next trading day to take action as long as the next intersection has not come yet.

4.3.1 Moving Average Criterion

As one of the key means of quantitative investment, moving average trading strategy is commonly used in the transaction of stocks and many other financial products. Researches show that even simple rules of average-curve strategy can enable investors to obtain excess returns and effectively avoid some risks in the market environment.[5]

The main reference standard of the moving average strategy is usually two or more average curves of different time spans. Generally, the larger the time span, the less the influence of new data on the amplitude change, which means that the curve is relatively smooth. In other words, in the event of sharp or continuous changes of the price, the average curve with a larger time span is more likely to lag behind, resulting in intersections with the average curve with a smaller time span. It is worth mentioning that the change curve of product price itself is the single moving average, but due to its large range of change, it is of limited use value in the strategy of moving average.

For two moving averages with different time spans, the basic strategy for moving averages is to buy when the short-term moving average rises above the long-term moving average and sell when the short-term moving average falls below. But there can be a series of continuous intersections in the choppy data. Considering of the commission in the problem, frequent buying and selling can reduce our capital significantly, hurting returns to a large extent. Therefore, our team chose to limit the times of transactions within a certain time span on the premise of cutting off a large number of continuous trashy points.

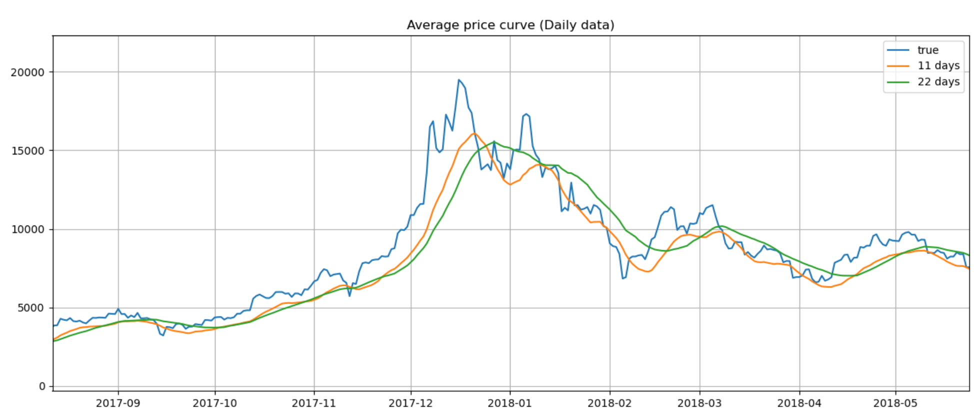


Figure 1: Average Price Curve

4.3.2 ARIMA Prediction

Adopting ARIMA(Autoregressive Integrated Moving Average) model for prediction is one of the predictive analysis methods of time series. The construction and prediction process of ARIMA(p, d, q) model includes stationarity test, parameter estimation and timeseries prediction. Parameter p in AR model, i.e., autoregressive model, is the lag number of the data itself. Parameter d indicates how many differences are needed for time series data to obtain stable data. Parameter q in MA model, i.e., moving average model, is the lag number of prediction error.

- **Parameter p and AR model**

  AR model describes the relationship between the current value and the historical value. The AR model with p-order lag can be expressed as follows:

  $$y_t = u + \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + e_t \tag{1}$$

  In the equation, u represent a constant, while e_t represent the error.

- **Parameter q and MA model**

  MA model describes the relationship between the current value and the error accumulation of the autoregressive part. The MA model with lag q order can be expressed as follows:

  $$y_t = u + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{2}$$

  In the equation, u represent a constant, while e_t represent the error.

- **Parameter d and difference**

  First order difference:

  $$y_t = Y_t - Y_{t-1} \tag{3}$$

  Second order difference:

  $$y_t = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) \tag{4}$$

- **ARIMA = AR + MA**

  $$y_t = u + \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{5}$$

- **Stationarity test**

  The time series data processed by ARIMA model shall meet the stationarity test. For the preprocessed time series data, the data shall be transformed by difference method, and then the stationarity test shall be carried out by unit root check.

- **Parameter estimation**

  Based on the information order determination criterion, the order determination score calculation formulas of AIC(Akaike Information Criterion) and other information can be used to determine the order of AR and MA models.

- **Time series prediction**

  Based on the constructed ARIMA model, the time series data are predicted. The specific process is as follows:

  1. ADF test

  Performing ADF test on the time series after preprocessing. When the p-value of the unit detection statistic is less than 0.05, it is statistically significant, and the difference parameter d of ARIMA model can be determined.

  2. Determining the order of AR and MA model parameter

  Set the algorithm parameter configuration information, estimate and rank other parameters of ARIMA model according to AIC criteria and BIC criteria, and fit the prediction model.

  At present, the following criteria are commonly used to select models:

  Akaike Information Criterion:

  $$AIC = -2ln(L) + 2k \tag{6}$$

  Bayesian Information Criterion:

  $$BIC = -2ln(L) + ln(n)k \tag{7}$$

  Hannan-Quinn Criterion:

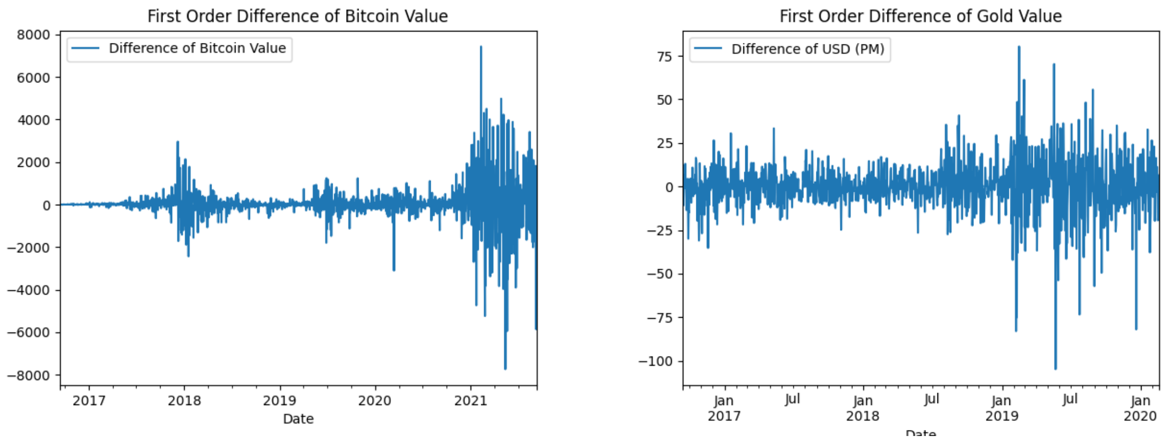
  $$HQC = -2ln(L) + ln(ln(n))k \tag{8}$$

  3. Iterative training and rolling prediction

  In order to avoid insufficient fitting of ARIMA model and insufficient accuracy of prediction model, the time series data are processed in the form of iterative training and rolling prediction, and the prediction step of ARIMA model is set to 1. After each round of iterative training, the algorithm will add the prediction value back to the training data to re fit the model and predict the results at $t + 2$.[6]

In the data processing, first we got the original data which was analyzed from the data of bitcoin and gold value trend given by the problem. Then, we preprocessed the original data, such as de duplication, null value processing, outlier processing, and got the time series data.

In modeling, first we observed the time series data after preprocessing and found that its oscillation is relatively strong. The p-bitcion value is 0.9339 and the p-gold value is 0.9042, which are far greater than 0.05, meaning they are typical non-stationary time series. Therefore, we considered the first-order difference processing. It was found that the p-bitcoin value of the data is 1.011×10^{-13}, and the p-gold value is 9.2697×10^{-13}, which is far less than 0.05. They met the stability condition , so we could adopt ARIMA modeling.

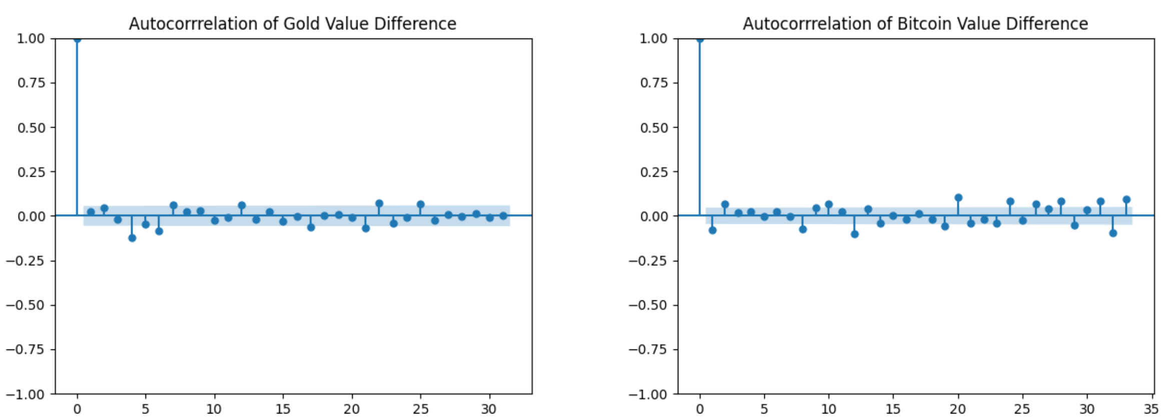


(a) Bitcoin (b) Gold

Figure 2: First order difference

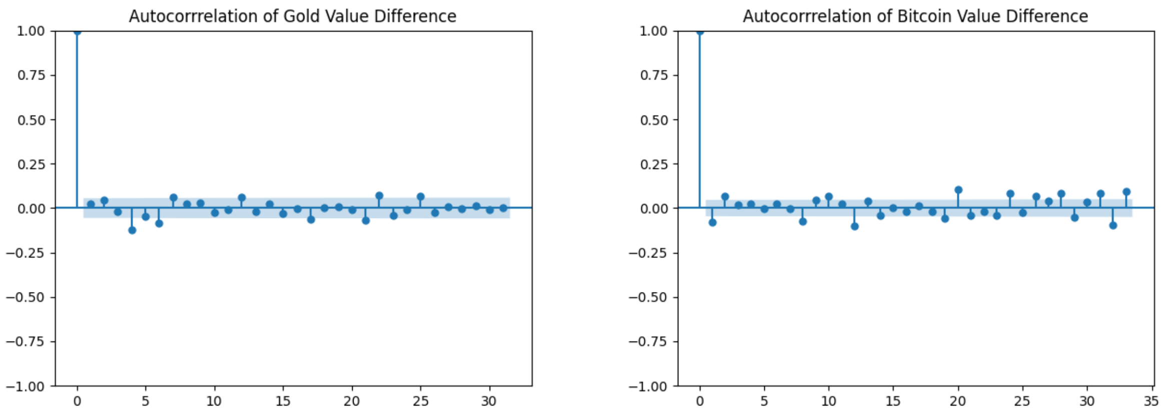
Then, we calculated the autocorrelation coefficient and partial correlation coefficient of the preprocessed time series data, determined the p, d and q parameters of the model by AIC, and fitted the time series model. The parameters of bitcoin (p, d, q) = (6, 1, 4) and gold (p, d, q) = (6, 1, 3) are obtained.

Given that the model requires data support, our team finally decided to adopt it after 30 days, that is to say, after October 11th, 2016.



(a) Bitcoin (b) Gold

Figure 3: Autocorrelation of Value Difference



(a) Bitcoin (b) Gold

Figure 4: Partial Autocorrelation of Value Difference

4.3.3 Apriori Prediction

As requested, the only source of reference for daily trading decisions is up-to-date price data. Due to lacking in other reference coordinates, our team decided to try to explore the statistical laws of the data to find the relevance. This thought is consistent with the idea of apriori algorithm, which is mainly used to analyze frequent item sets in a large amount of data to look for some interesting associations or interconnections between each other.[7]

In order to facilitate the use of apriori algorithm, it is necessary to classify the data first. Considering the commission issues involved in transactions, consider categorizing data into the following four categories:

1. The price of certain product has increased sharply recently;
2. The price of certain product has increased slightly recently;
3. The price of certain product has decreased slightly recently;
4. The price of certain product has decreased sharply recently.

The definitions of terms such as "recently", "sharply" and "slightly" in the above definition are vague. In the specific practice, considering that gold trading involves non-trading days with a span of two days, our team decided to analyze bitcoin on one day and three days respectively, while gold is analyzed on the unit of trading days. To ensure the profits, a "sharp increase" indicates that we should pay a commission to buy the product even if we don't own it. A "slight increase" indicates that if we hold the product, we should continue to hold it, while if we do not hold the product, we may choose not to buy it owing to large amount of commission. A "slight decline" indicates that if we hold the product, it costs less to hold on to it than to pay a commission to sell; A "sharp decline" indicates that if we hold the product, we should not stint on a commission to sell. After some adjustments on the parameters, the corresponding standards of the two products are summarized as follows.

Table 2: Principle of classification in apriori prediction

Product	Unit time span	Feature	Range
gold	1 trading day	sharp increase	>+1%
gold	1 trading day	slight increase	0%~+1%
gold	1 trading day	slight decrease	-1%~0%
gold	1 trading day	sharp decrease	<-1%
gold	7 days	sharp increase	>+3%
gold	7 days	slight increase	0%~+3%
gold	7 days	slight decrease	-2%~0%
gold	7 days	sharp decrease	<-2%
bitcoin	1 day	sharp increase	>+2%
bitcoin	1 day	slight increase	0%~+2%
bitcoin	1 day	slight decrease	-2%~0%
bitcoin	1 day	sharp decrease	<-2%
bitcoin	3 days	sharp increase	>+6%
bitcoin	3 days	slight increase	0%~+6%
bitcoin	3 days	slight decrease	-5%~0%
bitcoin	3 days	sharp decrease	<-5%

The boundaries between ranges are set according to the scale of commission.

First, we count the frequency with which three consecutive time units had different ups and downs in the available data. Then each day we move forward, we record the day's ups or downs and update the set. In this way, we are able to get a up-to-date frequency sheet. When functioning along, we treat it as the only judgement. While in the compound criterion, this algorithm can offer us the probability of tomorrow's up-or-down feature.

It should be noted that to explore such statistical law, we need quite a long period in preparation. Our team finally decided to adopt this judgement after 365 days, that is to say, after September 11th, 2017.

5 The Model Results

In the first 30 days, if we invest gold and bitcoin in a completely random ratio every day, the maximum value is 531 and a minimum value is 450 after 30 times. If the frequency is reduced to once every four days, the maximum value become 857 and the minimum value become 772. This indicates that reducing trading frequency to avoid high commission can reduce the possibility of loss to some extent.

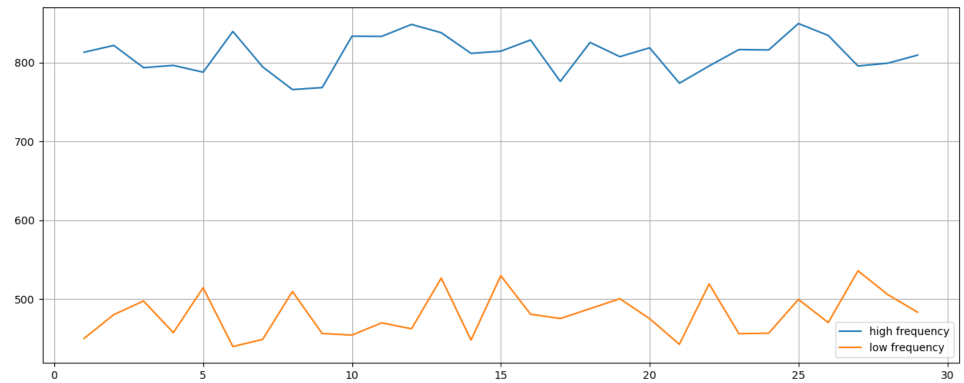


Figure 5: Results of Random Investment in the Early Stage

The following results are on the basis of no trading in the early period.

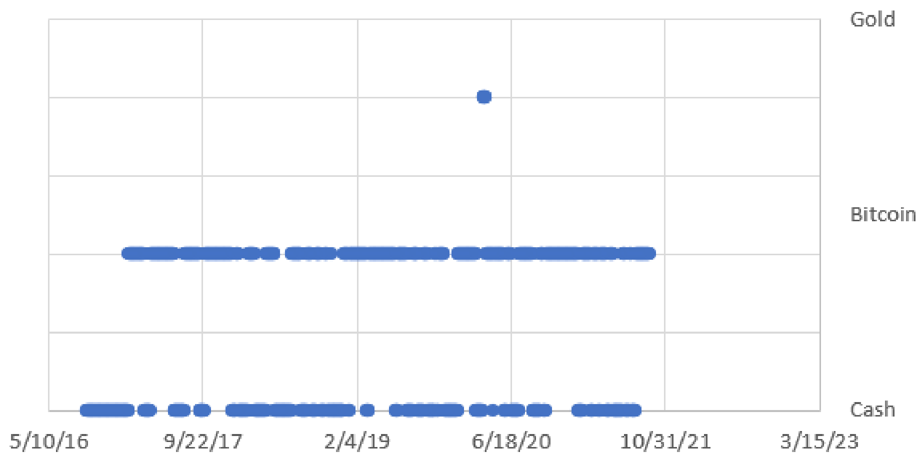


Figure 6: Decision-making Situations of Compound Criterion

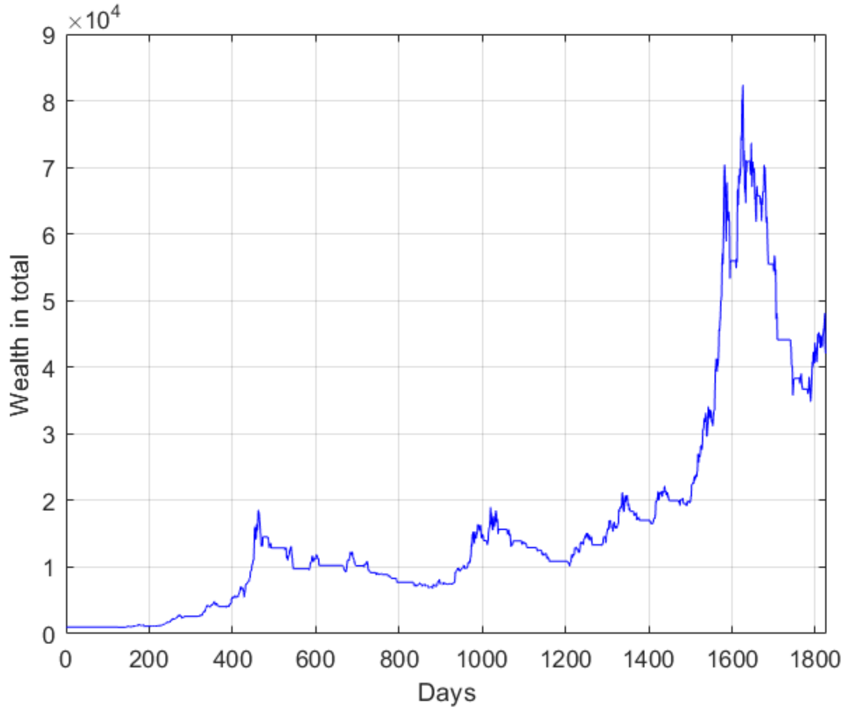


Figure 7: Result of Compound Criterion

When we use the composite criterion for investment selection, we get the investment results as shown in the figure below. We can see that the wealth can reach as much as 80000 at most, but unfortunately it will eventually drop to about 42353.

The reason may be that the price trend of bitcoin and gold is actually affected by many factors, so it is difficult to make accurate judgment.

6 Validating the Model

- **Moving Average Criterion Only**

  Using moving average criterion independently, the result is as follow:

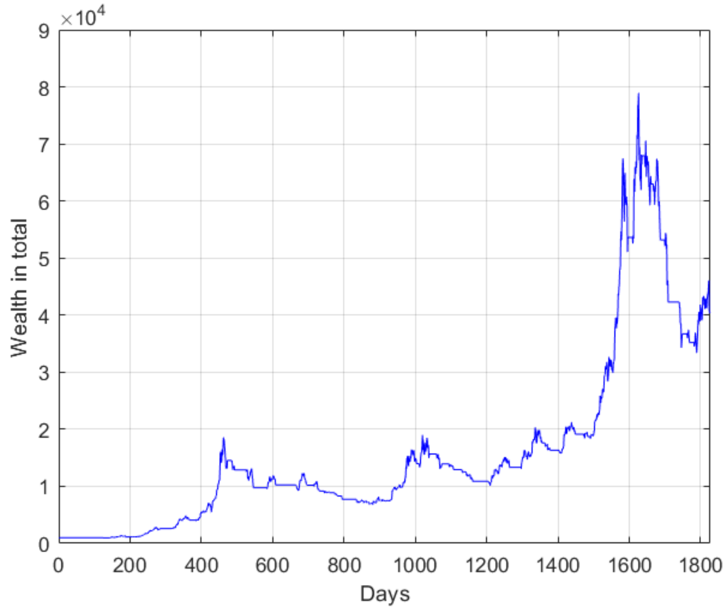


Figure 8: Result of Moving Average Criterion

We can see that it can get a fairly good result. The final wealth is about 40576 dollars.

- **ARIMA Prediction Only**

  Using ARIMA prediction independently, the result is as follow:

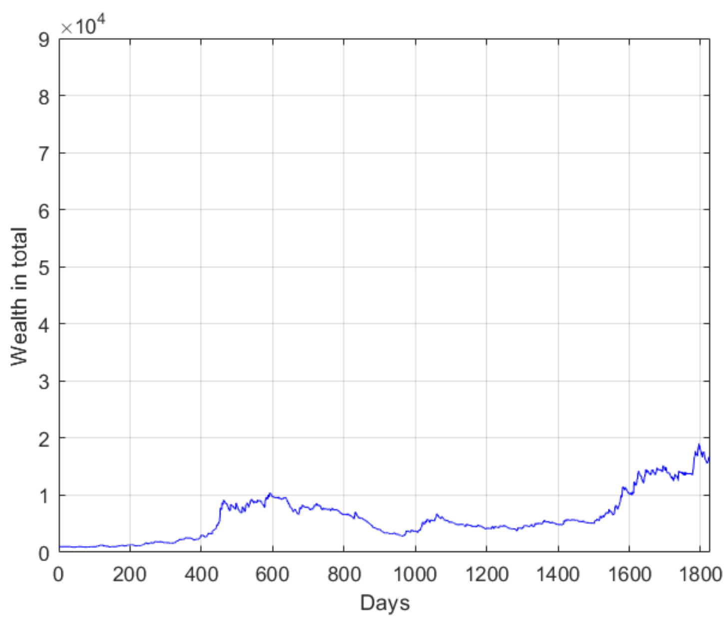


Figure 9: Result of ARIMA Prediction

We can see that it works badly. That's because with only one dimension of coordinate, it hasn't a considerable accuracy. The final wealth is about 16401 dollars.

- **Apriori Prediction Only**

  Using apriori prediction independently, the result is as follow:

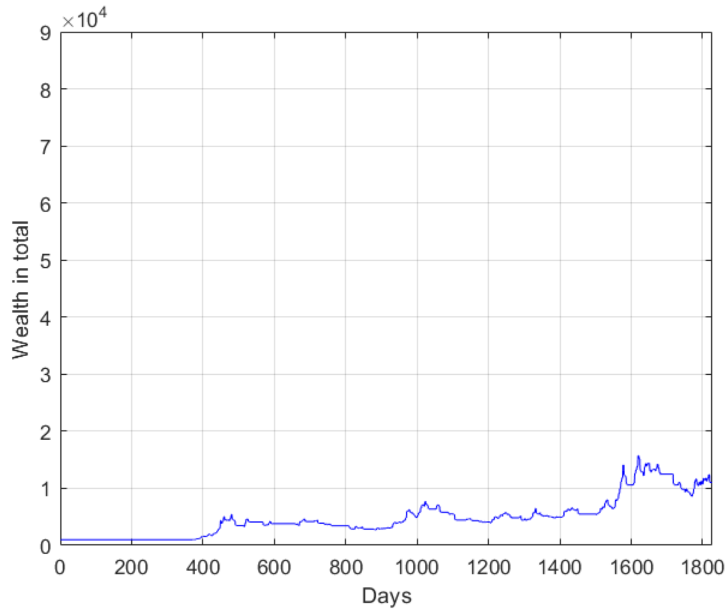


Figure 10: Result of Apriori Prediction

We can see that it also works badly. That's because of both its low accuracy and late start. The final wealth is about 11030 dollars.

- **Additional Compare Group**

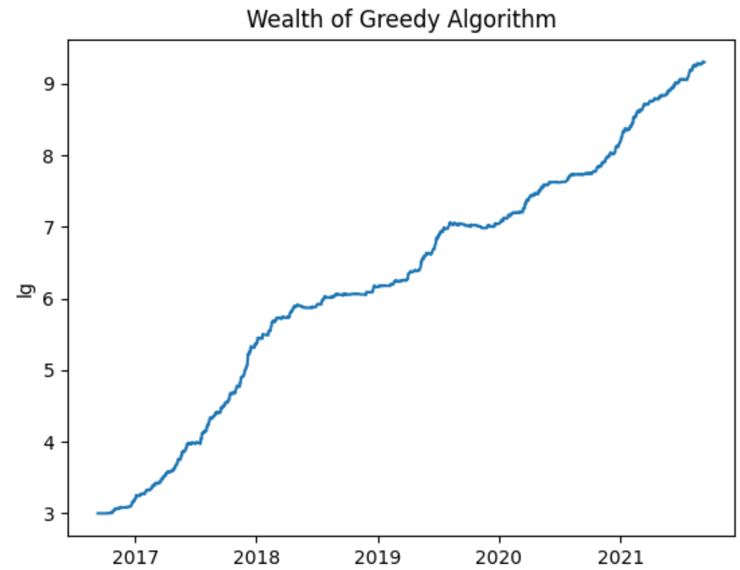


Figure 11: Result of Greedy Algorithm

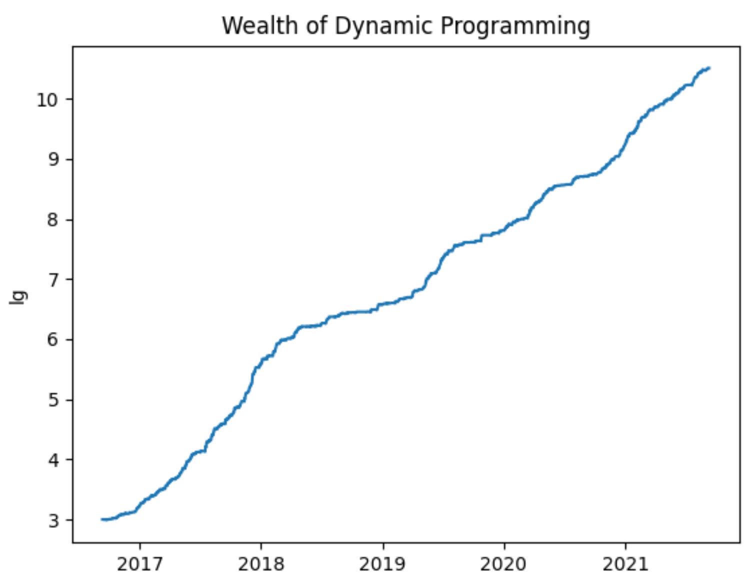


Figure 12: Result of Dynamic Programming Algorithm

When we accurately know the value of bitcoin and gold the next day and adopt the greedy algorithm, we find that we can finally get 1.99×10^9; When we accurately know the daily value of bitcoin and gold in these five years and adopt dynamic planning, we can even earn more, up to 3.21×10^{10}.

Obviously, this result is far from ours, but we are neither Buffett nor God. Investment is like gambling. But it's more than it. In fact, there are many factors affecting the prices of bitcoin and gold. The market is changing rapidly. No one knows what will happen tomorrow.

7 Sensitivity Analysis

As mentioned above, the compound criterion used in our model is a composion of three independent criteria. Among them, moving average criterion is used to look for trading opportunities, while ARIMA model only provides forecast data for existing values. So they have nothing to do with the change of commission, or to say, transaction costs. The classification method of apriori prediction is based on the percentage of commission, which is used to refer to whether a decision is worthwhile or not. Therefore, the following content mainly analyzes the sensitivity of apriori prediction to the change of commission when functioning alone or in combination.

For apriori prediction, the most important thing is to determine the classification standard. And the classification standard partly depend on the commission. When commissions are high, we need to look for numbers that rise or fall more sharply, and these numbers can be too few to make the pattern significant. When commissions are low, too much data may fall into the 'sharp' range, causing the rule to lose reference value. However, the standards are also influenced by real changes, so high sensitivity is not necessarily a bad thing. We have to admit that our team have not summed up an exact equation. Only through parameter adjustment could we find a suitable boundary value.

In summary, this algorithm is sensitive to commission. However, considering that it is only used for gold prediction in the composite criterion, which is not our focus, the whole model still has strong stability.

8 Strengths and Weaknesses

8.1 Strengths

- **Updated over time**

  Our model can remember the data of the value trend of bitcoin and gold in the past and update them in real time. It can pursue advantages and avoid disadvantages in time, and improve returns while reducing risks.

- **Each component gives full play to its strong point**

  Above, we enumerate the results of the three components of the composite criterion when they work separately. One of them even could not reach the level of 'buy in the first day and sell in the last day'. But when they are combined together, we picked out each component's strengths together, which improved the performance to a certain extent.

8.2 Weaknesses

- **Simplication for the decisions**

  For our own convenience, all of our decisions consist of buying and selling only, which means we always have just one kind of property in hand. Indeed, we proved that theoretically optimal decisions are made in a familiar way. However, considering the feature and risk of the two products are not the same, modern investment theory as well as actual investment decisions often choose to hold two kinds of property or more. With three different judgements, the compound may upgrade by detailing each decision's object as well as a scale.

- **Weight allocation of the compound criterion have yet to be improved**

  Due to the limited time, when the three criteria were combined into one, we just adjusted the primary and secondary relations of the three criteria according to their accuracy and effect in their respective functions. We believe that after further analysis, more reasonable weight of the three can be assigned, or other ways to compound can be found to better serve the requirements of the problem.

References

[1] https://baike.baidu.com/item/%E6%8A%95%E8%B5%84/211753?fr=aladdin

[2] Ye Wuyi, Sun Liping, Miao Boqi. Research on dynamic cointegration of gold and bitcoin – Based on semi parametric Midas quantile regression model [J] Systems Science and mathematics, 2020,40 (07): 1270-1285.

[3] Capital University of economics and trade.Sun Ruiqi: Research on price trend prediction model of US stock index based on LSTM neural network Algorithm Introduction.2015.5.

[4] Zhou Mingshan, Feng Xinli, Lin Liang, Fang Xuyun, Zhou Guiguo. Research on the effectiveness of a-share market moving average strategy and random characteristics of return rate [J]. Securities Market Review,2013(01):58-64.

[5] Shanghai Normal University. Guo Sihan:Research on bitcoin price prediction and trading strategy based on improved recurrent neural network.2021.6.

[6] https://blog.csdn.net/weixin_42382211/article/details/81133807

[7] Zhao Hongying, Cai Lecai, Li Xianjie. A review of Apriori algorithm for mining association rules [J]. Journal of sichuan university of science and technology (natural science edition),2011,24(01):66-70.

Appendices

Appendix A Core Codes of Python

- **DP**

```
for i in range (10,length-1):
    for j in range(0,10):
        ads_eleven[i]+=float(ads_a.Value[i-j])
    ads_eleven[i]/=11
np.savetxt('out1.txt',ads_eleven)

for i in range (21,length-1):
    for j in range(0,21):
        ads_22[i]+=float(ads_a.Value[i-j])
    ads_22[i]/=22
np.savetxt('out2.txt',ads_22)

plt.figure(figsize=(18, 6))
plt.plot(ads_a.Date,ads_a.Value,label="true")
plt.plot(ads_a.Date,ads_eleven,label="11 days")
plt.plot(ads_a.Date,ads_22,label="21 days")
plt.title('Average price curve (Daily data)')
plt.grid(True)
plt.show()
```

- **Apriori**

```
def Apriori(dataset, min_support = 2):
    c1 = item (dataset)
    f1, sup_1 = get_frequent_item(dataset, c1, min_support)

    F = [f1]
    sup_data = sup_1

    K = 2

    while (len(F[K-2]) > 1):
        ck = get_candidate(F[K-2], K)
        fk, sup_k = get_frequent_item(dataset, ck, min_support)

        F.append(fk)    #
        sup_data.update(sup_k)
        K+=1
    return F, sup_data
```

- **ARIMA**

```
def optimizeSARIMA(parameters_list, d, D, s):
    results = []
    best_aic = float("inf")
    for param in tqdm_notebook(parameters_list):
        try:
            model = sm.tsa.statespace.SARIMAX(ads.Value, order=(param[0], d, param[1]),
                                              seasonal_order=(param[2], D, param[3], s)).fit(disp=-1)
        except:
            continue
        aic = model.aic
        if aic < best_aic:
            best_model = model
            best_aic = aic
            best_param = param
        results.append([param, model.aic])
    result_table = pd.DataFrame(results)
    result_table.columns = ['parameters', 'aic']
    result_table = result_table.sort_values(by='aic', ascending=True).reset_index(drop=True)
    return result_table
def plotSARIMA(series, model, n_steps):
    data = series.copy()
    data.columns = ['actual']
    data['sarima_model'] = model.fittedvalues
    data['sarima_model'][:s + d] = np.NaN
    forecast = model.predict(start=data.shape[0], end=data.shape[0] + n_steps)
    forecast = data.sarima_model.append(forecast)
```

- **DP**

```
for i in range(1,length):
    dp_money[i] = max(dp_money[i-1],0.98*bitcoin_value[i]/bitcoin_value[i-1]*dp_bitcoin_value[i-1],0.99*gold_USD[i]/gold_USD[i-1]*dp_gold_USD[i-1]*t[i])
    dp_bitcoin_value[i] = max(0.98*dp_money[i-1],bitcoin_value[i]/bitcoin_value[i-1]*dp_bitcoin_value[i-1],0.99*0.98*gold_USD[i]/gold_USD[i-1]*dp_gold_USD[i-1]*t[i])
    dp_gold_USD[i] = max(0.99*dp_money[i-1]*t[i],0.98*0.99*bitcoin_value[i]/bitcoin_value[i-1]*dp_bitcoin_value[i-1]*t[i],gold_USD[i]/gold_USD[i-1]*dp_gold_USD[i-1])
```

Appendix B Memorandum

Decision-making Situations

From: Team 2221608

To: Market trader

Date: 21st February, 2022

In the past research, the prediction of dynamic prices and fluctuations of financial products has been one of the great challenges facing market traders, and to judge the direction of market prices wisely based on existing information is the necessary ability of every successful market trader.This paper takes Bitcoin and gold as investment objects, and constructs a portfolio investment strategy based on price prediction and multiple judgment basis to seek excess returns.

The strategies that we used are as follows:

1. Data-based trading strategy: Wait for the data-based decision model to start run.That means no transactions was made for the first thirty days;

2. The High-risk Trading Strategy: Invest bitcoin in the long term. Only when there is a risk of a sharp drop in bitcoin price, and only after a projected surge of gold price at least 3%, in a seven-day cycle (including at least not trading days), we choose to put our money into the gold market to maintain our wealth;

3. The Bitcoin Average Value Trading Strategy: Buy when the 11-day average price curve goes up through the 22-day average price curve and sell when it passes down;

4. Price Forecast Investment Strategy: Select the forecast prices with significant fluctuations as real future prices;

5. Price Characteristic Investment Strategy: Predict the future combinations

of price fluctuations based on the frequent combination of price fluctuations in historical data;

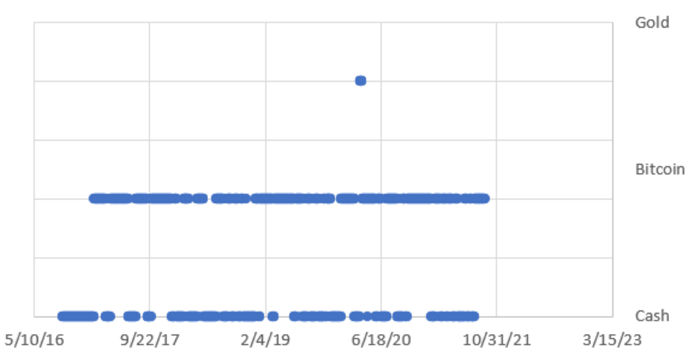
6. Low-frequency Trading Strategy:Hold bitcoin to avoid high fees unless the market shows a significant downward trend.

The investment model that we used is as follows:

1. Build an ARIMA algorithm model that is automatically updated daily to predict the price changes of bitcoins and gold (Initially train the model with data for 30 trading days before adding the latest daily data update);

2. Build the Apriori algorithm model which studies the past bitcoins and gold time series features to extract the combinations of price fluctuations with high confidence (Initially train the model with data for 400 trading days before adding the latest daily data update);

3. Draw the average trading curve for 11 and 22 days to forecast the overall market trends in the future Our final results are presented as follows:



Hope that our model will be helpful in guiding market transactions. We can continue to refine our model later.