

# Prediction of the Best Portfolio for Bitcoin and Gold based on the ARIMA Model

Qi Zhou<sup>1</sup>, Zixuan Chen<sup>1</sup>, Zhuoying Cai<sup>2</sup>, Ziwei Xia<sup>2</sup>

<sup>1</sup>School of Accounting, Anhui University of Finance and Economics, Bengbu, Anhui, China

<sup>2</sup>School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, China

**Abstract:** With the prosperity of the financial market, more and more people are involved in securities trading. How to combine investment in bitcoin and gold to achieve the highest profit is one of the issues that market traders think about. To solve this problem, we build a model that predicts future prices in order to better help investors. We constructed an ARIMA model through differential stationarity processing, AEC, white noise test and other methods, and used the data of the current day and the previous day to predict the price of the next day. At the same time, we use the model to predict the average for the next N days. If it is predicted that the price of the asset will decrease in the future, it will be sold on the same day. If the price of the asset will increase in the future, and the income obtained is greater than the transaction cost and the expected income of the investor, the purchase will continue until the end. Last until the last day \$323841.52. Finally, it can be verified that the accuracy of the ARIMA model is the highest by comparing other mainstream machine learning models. In other words, the ARIMA model is the best strategy for this problem.

**Keywords:** ARIMA model, Bitcoin, Gold.

## 1. Introduction

### 1.1. Background

In recent years, due to the relatively stable fundamentals of the securities investment market and its positioning in a professional and market-oriented system and mechanism, the scale and income of financial investment business have shown a trend of rapid growth. More and more residents are willing to invest their income in market transactions in order to make bigger profits. As an important part of market investment, volatile assets are known for their short-term financing flexibility and weak stability which can solve financing difficulties in a relatively short period of time.

In financial investment markets, traders often buy and sell liquid assets to maximize their total returns. Among them, it is not difficult to find that gold and bitcoin are commissions when trading. However, given the instability of the daily prices of Gold and bitcoin, how to optimize the asset trading portfolio in the correct way has become a problem worth thinking about.

## 2. Assumptions and Justifications

- Assume that the past prices of Gold and bitcoin are known.
- Assume that the non-trading time price of Gold is the price of the last day before, and ensure that the data volume of Bitcoin and Gold is the same.
- Assume the trader is risk neutral and will not react too quickly or too slowly.
- Assuming transaction fees is commission.

## 3. Symbols and Notations

Symbols	Definition
p	Number of autoregressive terms
d	Algorithm entry difference coefficient
q	Moving Average Coefficient
r	Expected risk return
Y1	Bitcoin
Y2	Gold
$\rho$	Correlation coefficient
$y_{1-t}$ ( $t=1,2,3,4,5$ )	Bitcoin's price change in five years
$y_{2-t}$ ( $t=1,2,3,4,5$ )	Change in the price of Gold in five years

## 4. The Overall and Specific Thinking of The Problem

### 4.1. Overall Idea of The Problem

To correctly to analyze the optimization of trading strategies, the problem that needs to establish a quantitative indicator system,so we chose the ARIMA algorithm model for prediction.

ARIMA model is a well-known time series forecasting method,which is essentially a linear regression. It performs regression according to its own historical data and is suitable onto data with high and stable correlation.

First, according to the 2016-2021 data of each bitcoin and Gold, put the data from history to today into the ARIMA model to predict future prices. Determine whether traders should buy, hold or sell assets in their portfolios on a daily basis by comparing the prices of two assets, Bitcoin and Gold. At the same time, in order to maximize the accuracy of the forecast, a forecast is carried out every day, the asset comparison is performed using the ARIMA model. Surely, the

model can be optimized in the later stage, adding an elastic amount to reduce the sensitivity to price, such as how much loss is accepted. Finally, it can be verified that the accuracy of the ARIMA model is the highest by comparing other mainstream machine learning models, the ARIMA model is the best strategy for this problem. In order to determine the sensitivity of the strategy to transaction costs, we considered fine-tuning the asset price, and then made the forecast results from the next day. The results show that the impact on

transaction costs is small in the short term and significant in the long term.

Finally, through the ARIMA model, we can smoothly get the decision of each day after the first day, and give the best daily trading strategy based on the trading strategy of the day. And from this, we can know that the initial \$1,000 investment is worth \$94,489.092 five years later.

The overall idea flows chart 1 is as follows:

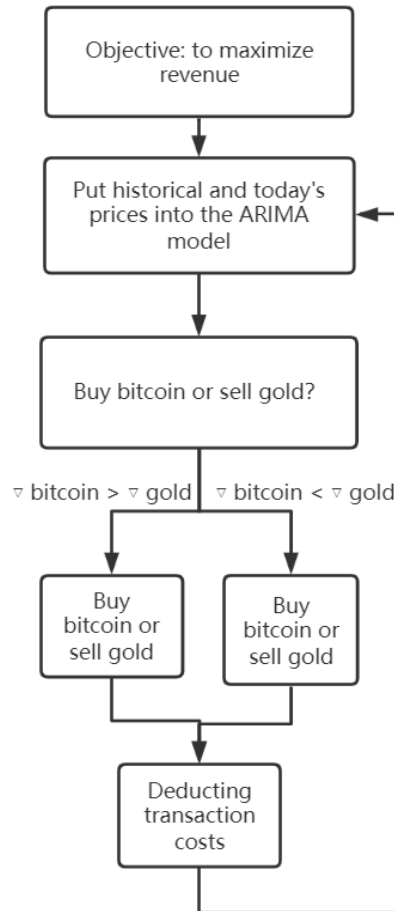


Figure 1. Overall idea flow chart

## 4.2. Problem Specific Ideas

### 4.2.1. Solution to How the Trader Makes Daily Decisions

- Build an ARIMA model

As a typical representative of time series models, ARIMA model has the characteristics of strong short-term forecasting ability, high practicability, and data stability, and is often used to predict value data onto the economic and financial fields. When predicting the value of the asset portfolio, we can first perform the stationary test of the data, then use the autocorrelation coefficient and partial autocorrelation coefficient to determine the order, and perform the significance test and white noise test to confirm that the model is reasonable so that the model can be completed at last. The general form of the model are as follows:

$$Y_t = c + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q}$$

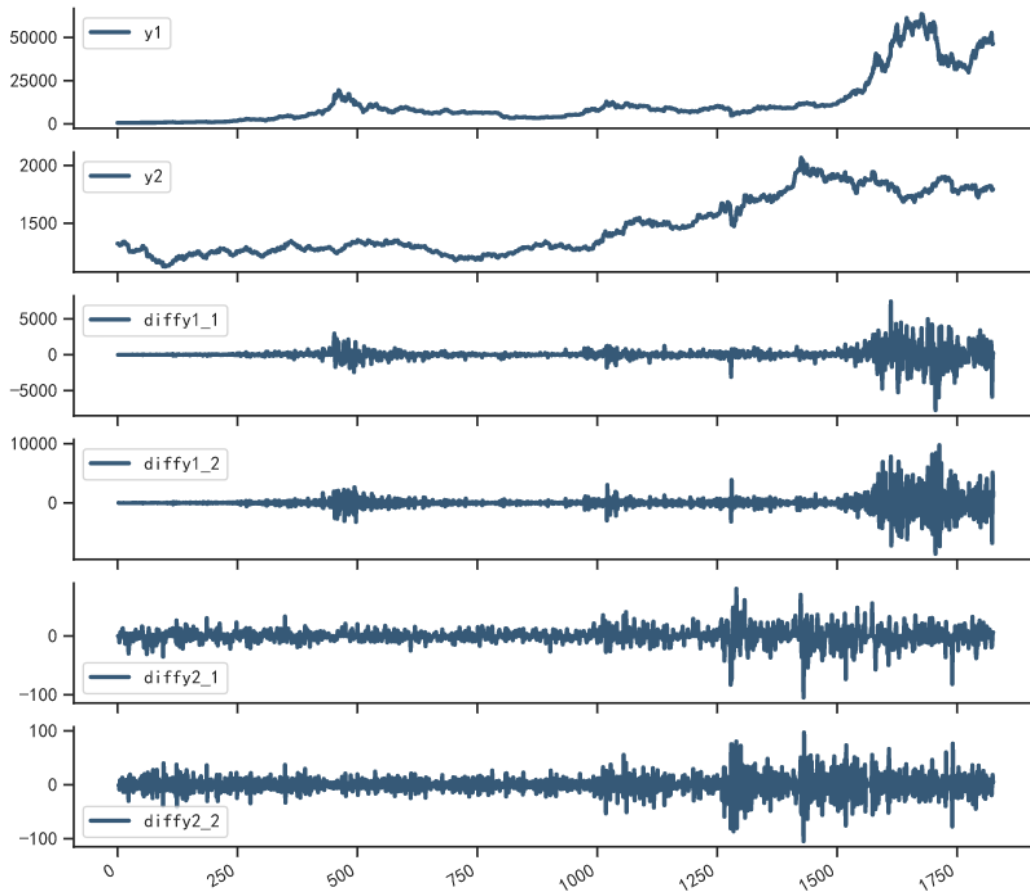
#### 1. Data Preprocessing - Stationarity Test

The preprocessing of ARIMA model includes two aspects: stationary test and white noise test. First, we observed that the data is in a state of non-stationary growth, and difference processing is required. We suppose  $y_1$  is Bitcoin,  $y_2$  is Gold,  $\text{diffy1}_1$  is the first-order difference of  $y_1$ ,  $\text{diffy1}_2$  is the second-order difference of  $y_1$ ,  $\text{diffy2}_1$  is the first-order difference of  $y_2$ , and  $\text{diffy2}_2$  is the second-order difference of  $y_2$ . It is in periodic fluctuations and cannot be stabilized in a certain interval value, so the second-order difference is performed again, and the data tends to be stable. The following are the formulas for first-order difference and second-order difference:

$$\nabla Y_t = (1 - B)Y_t$$

$$\nabla_2 Y_t = (1 - B)^2 Y_t$$

The difference results are shown in Figure 2.



**Figure 2.**  $y_1, y_2$  and their first-order second-order differences

2. Determine  $p, d, q$

● Determine  $d$

The ARIMA model uses three coefficients  $p, d,$  and  $q$  to perform the entry algorithm, where  $p$  represents the number of autoregressive terms,  $q$  represents the number of moving average terms, and  $d$  represents the difference order to ensure the stability of the algorithm entry.

Through the second-order difference, the non-stationary time series is turned into a stationary one. According to Figure 2, we can observe that  $diffy1\_2$  and  $diffy2\_2$  are relatively stable.

In addition, we performed the Box-Pierce test on the data after the second-order difference. The test results are shown in Table 1 below. It can be found that the P values are all greater than 0.05, indicating that the null hypothesis is accepted, the second-order difference sequence is considered to be a white noise sequence.

Therefore, the second-order difference sequence can be

selected for analysis and processing, so  $d=2$ .

**Table 1.** Box-Pierce test

lag	AC	Q	Prob ( $>Q$ )
6.0	0.020302	464.433412	0.387631
12.0	-0.075962	510.390494	0.136985
18.0	-0.043200	519.793121	0.061569

● Determine  $p, q$

We find that Bitcoin is traded every day, while Gold can only be traded on weekdays, so we first assume that the price of Gold on non-trading days is the price of the last day at the end of the previous transaction, to ensure that the data collected by Bitcoin and Gold are collected and their length are the same.

To determine  $p$  and  $q$ , we observe the ACF and PACF plots and find  $q$  can be determined by the ACF diagram,  $p$  can be determined by the PACF, and the relevant theories are shown in Table 2.

**Table 2.** Theories related to ACF and PACF

Prozess	ACF( $\rho_X(h)$ )	PACF( $\pi_Y(h)$ )
AR( $p$ )	Infinite (dampened exponential or sinusoidal waves)	Finite $\pi_Y(h)=0$ for $h>p$
MA( $q$ )	Finite $\rho_X(h)=0$ for $h>p$	Infinite (dampened exponential or sinusoidal waves)
ARMA( $p, q$ )	as AR( $p$ ) for $h>p$	as MA( $q$ ) for $h>q$

According to Figure 3 and Figure 4, the ACF Figure and PACF Figure of the second-order difference, it is concluded

that the ACF figures becomes very small after the 2 time point, so the end  $q$  value is 2. And the PACF figures to oscillate

between 0, so the p value is 0. Therefore explain the model is MA(2).  
 transformed into an MA model which is ARIMA(0,2,2) or

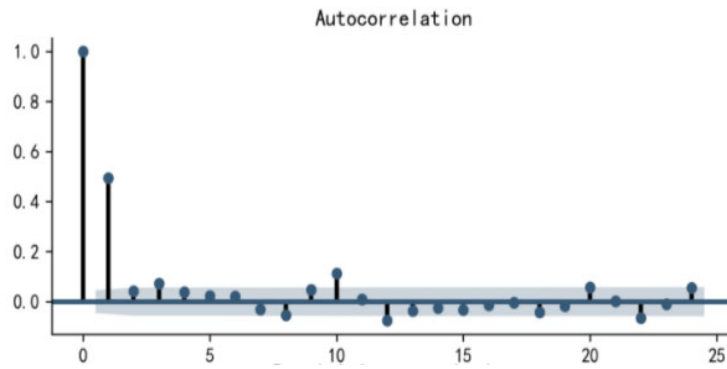


Figure 3. ACF

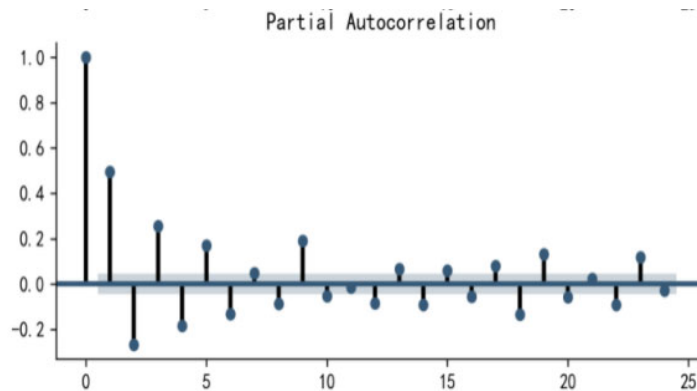


Figure 4. PACF

### 3. Residual PP plot with QQ plot and normality test

We make the residual PP Figure and distribution Figure, as shown in Figures 5 and 6, and find that the residual PP Figure and the QQ Figure basically show a linear relationship, which is relatively stable, and the closer the blue is to the red

reference line, the more in line with the expected distribution; we construct a residual histogram, as shown in the figure, and analyze it in the form of a function. It is found that the density distribution of the residual show a normal distribution, which means that the model has the same error in the high and low directions, and the overall seems more reasonable.

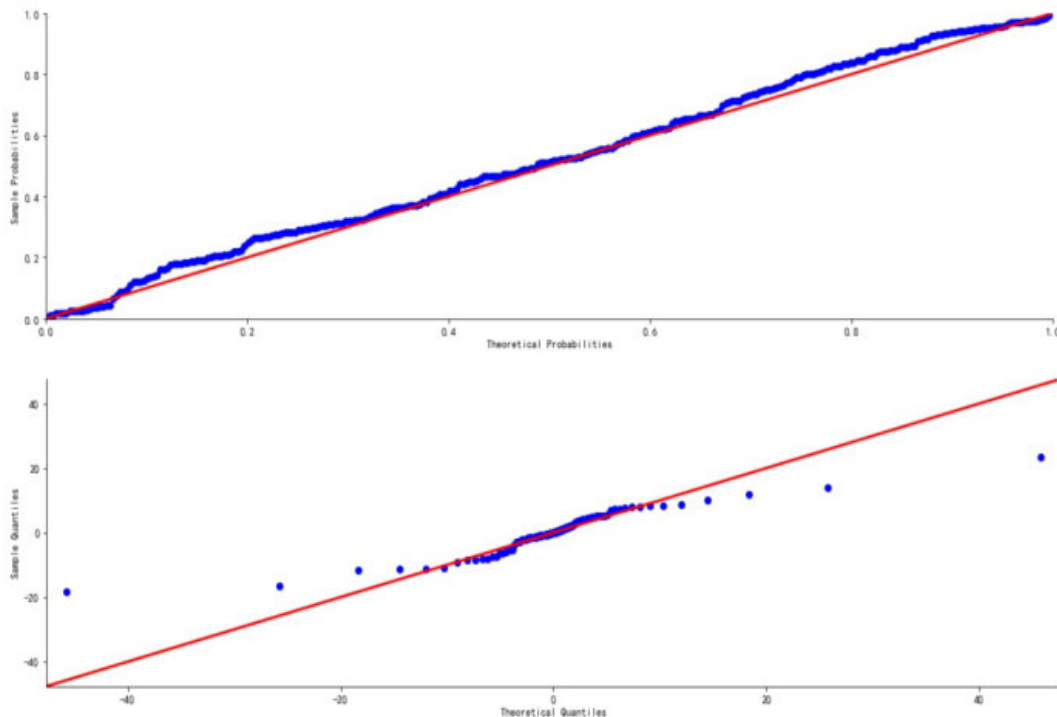


Figure 5. PP figure and QQ figure

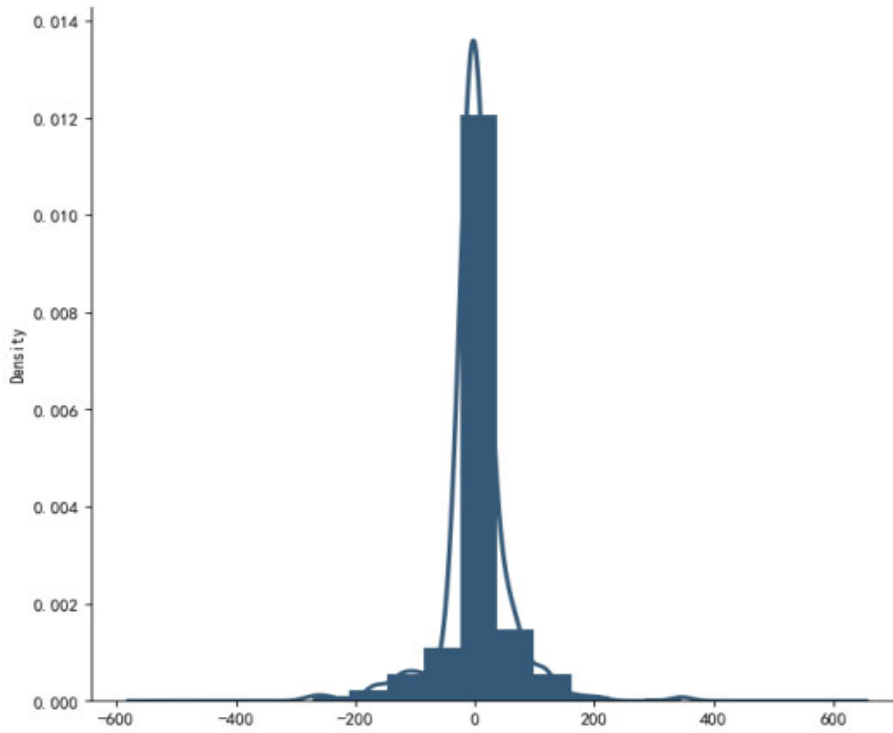


Figure 6. Residual Normality Test

- **Comparison of Bitcoin and Gold 100-day and five-year forecast data**

We use the ARIMA (0, 2, 2) model to compare the predicted price of Bitcoin and Gold with the actual value, as shown in Figure 7 and 8, to demonstrate the predictive ability and accuracy of the model. According to the predicted value of Bitcoin and Gold, we find that the data is more

concentrated than the actual value, indicating the stability of resistance to risk, indicating that the model has strong predictive ability and risk prevention.

The core Python code is as follows:

```

model=sm.tsa.ARMA(df['y1'].diff(2).iloc[1:].dropna(),(0, 2)).fit(method='css')
model=sm.tsa.ARMA(df['y2'].diff(2).iloc[1:].dropna(),(0, 2)).fit(method='css')

```

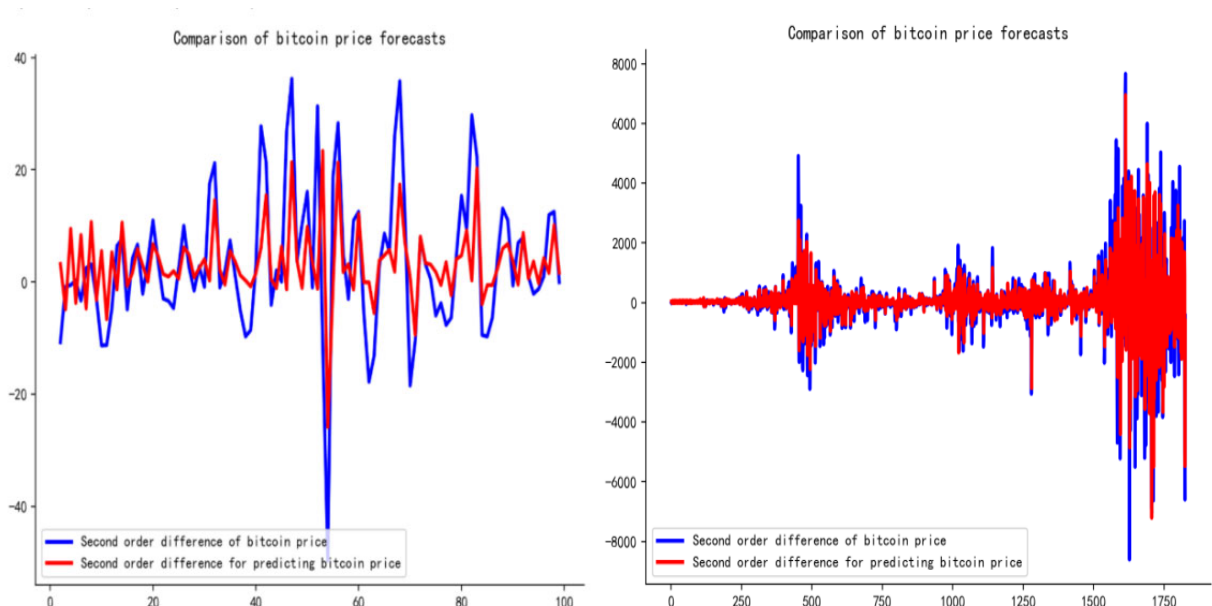


Figure 7. Bitcoin's first 100-day and five-year forecast and actual value

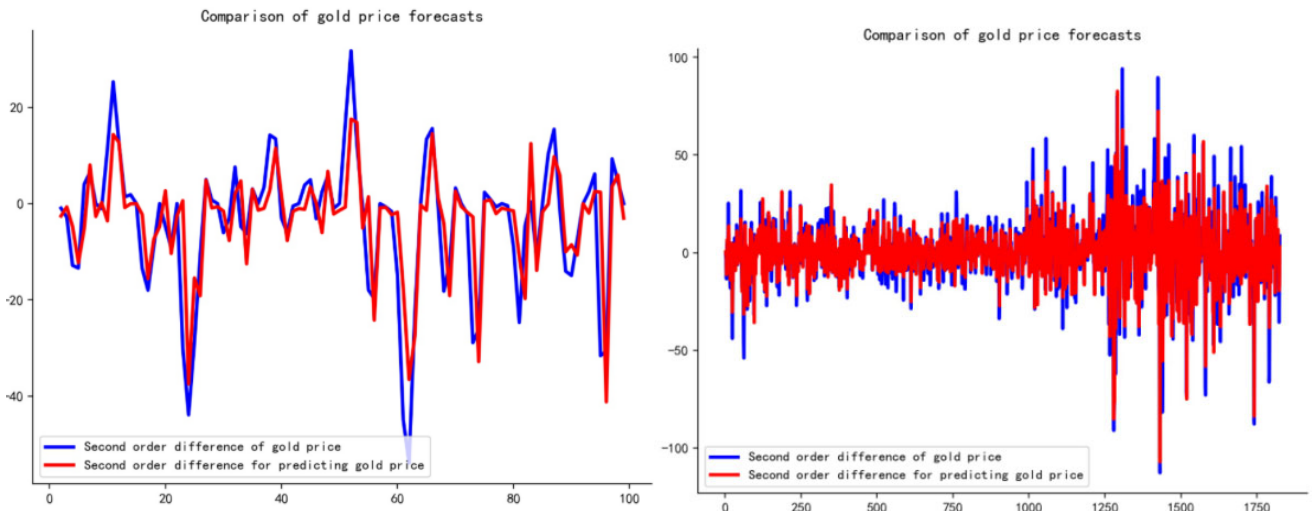


Figure 8. Gold's first 100-day and five-year forecast and actual value

The error between the five-year predicted value of Gold price and the actual value is shown in Figure 9 below. It can

be seen that the overall situation is relatively stable, so the ARIMA model can be used for prediction.

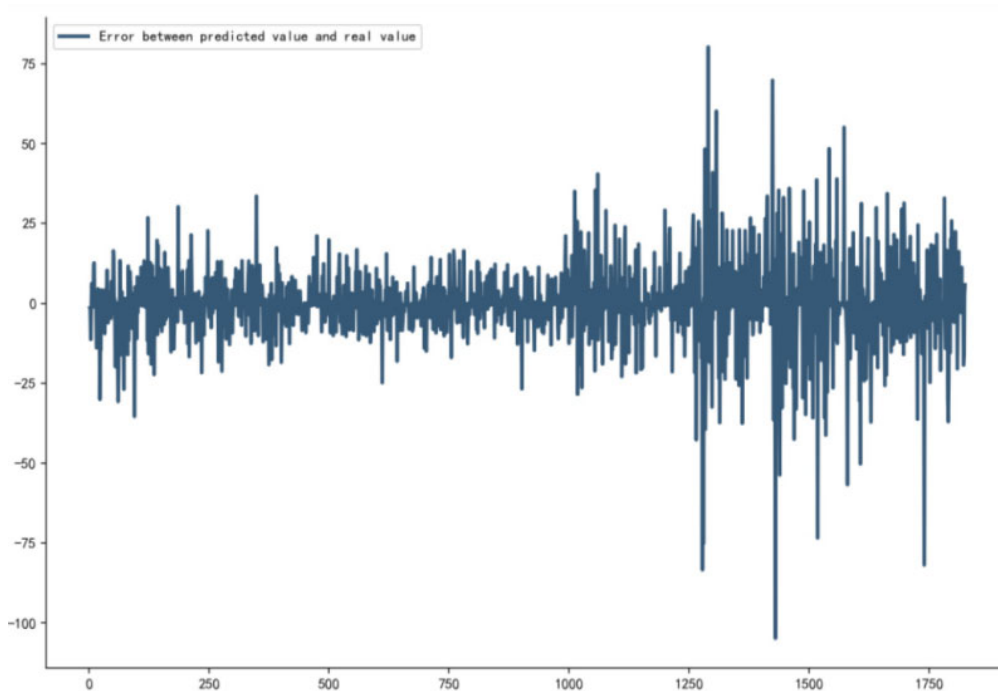


Figure 9. The difference between Bitcoin's predicted value and its true value

#### 4.2.2. Specific Analysis of Giving the Best Daily Trading Strategy

It requires us to give the best daily trading strategy which is based on the trading strategy of the day and provides relevant evidence. In this regard, we recommend using the model to predict the average value of  $N$  days in the future.  $N$  is the risk factor which is determined by the investor's own risk awareness and other factors. The larger the  $N$ , the more conservative the investor, and the smaller the acceptable risk fluctuation. If it is predicted that the price of the asset will decrease in the future, it will be sold on the spot. On the contrary, if the price of the asset will increase in the future, and the income obtained is greater than the transaction cost and the expected income of the investor, it will be purchased. Until we get the best investment return amount the last day.

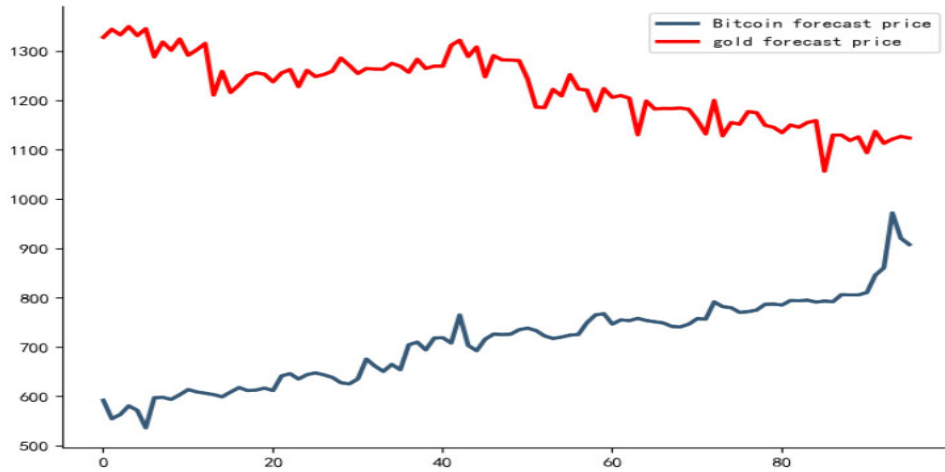
#### 4.2.3. Solution

##### ● Examples of trading strategies

Assuming that we hold bitcoin and gold today, if the price of bitcoin today is \$100 and the price of gold is \$100, the average price of bitcoin in the next  $N$  days is predicted to be \$150, and the price of gold is \$90. Therefore, it is expected that the price of bitcoin will rise while the price of gold will fall. At this time, you ought to buy bitcoin and sell gold.

##### ● Determination of the mean predicted value

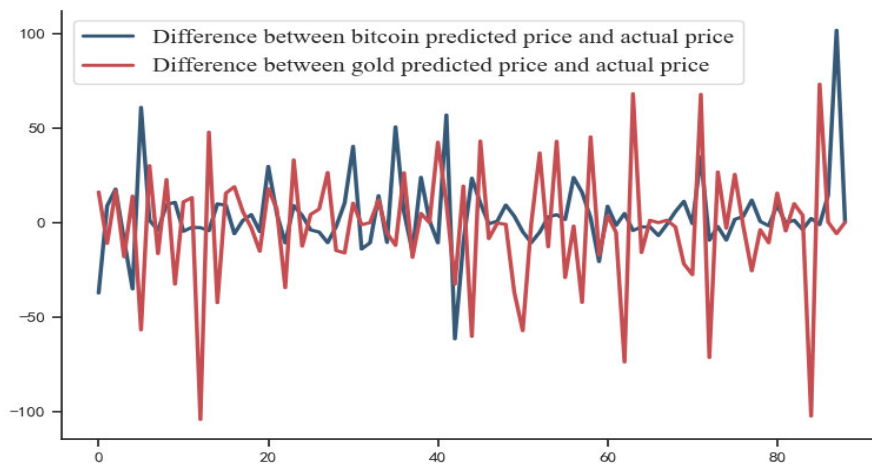
According to the investor's own needs, we use the ARIMA model to predict the asset prices in the next  $N$  days. For example, we calculate the predicted values of Bitcoin and gold prices for 10 days and average them. At the same time, in order to calculate the best asset price on the last day, we repeated this prediction every day and the process continues until the last day, as shown in Figure 10.



**Figure 10.** Top 100 Days Forecast for Bitcoin and Gold

Simultaneously, we make a prediction of the difference between tomorrow's price and today's price as shown in Figure 11. Then we calculate the asset price. Finally, we sell the asset at the highest point of the portfolio and buy it at the

lowest point. As the chart shows, we can sell around the 40th day to get the best value and buy around the 10th day in order to earn higher profits at lower costs.

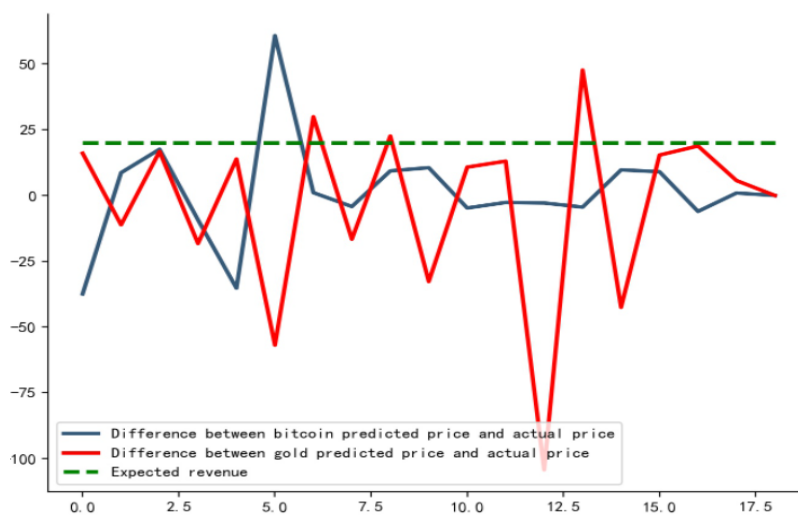


**Figure 11.** The difference between the first 100 days and today's price

● **Expected returns and risk avoidance strategies**

1. Fixed income of \$20

Regarding asset forecasts for bitcoin and gold, we need to consider earnings of a fixed \$20 as shown in Figure 12.



**Figure 12.** Fixed income of \$20

2. Risk avoidance

By comparing a fixed profit which considering risk factors or not, we can discover that the expected return increases when we consider the risk situation. Meanwhile, with the increase of risk, the expected earnings are also basically

increased. At this time, the assets can be bought or sold above the fixed income line as shown in Figure 13. The expected risk and return formula is as follows:

$$r = \text{Log} (\nabla y_1 + \nabla y_2 + 1) * 5$$

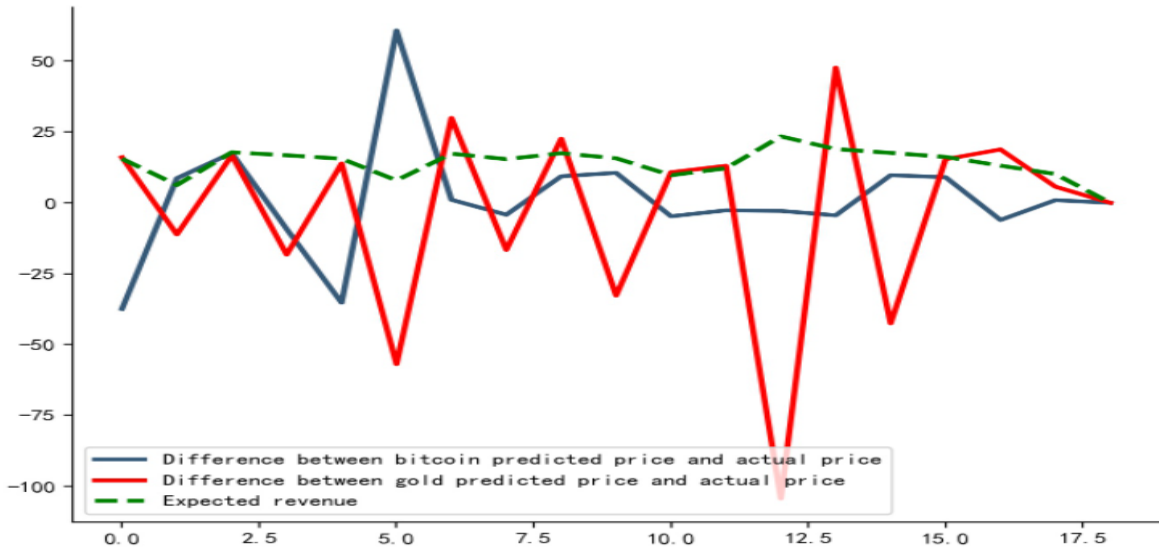


Figure 13. Risk-expected returns

● Transactions and Earnings

Through the above analysis, we invested from 2016 to 2021, initially holding funds of \$1000. In order to maximize the final income, we suppose that every investment is spent. Then we can get the daily investment strategy as shown in Annex

1. Through analysis, it can be concluded that a total of 88 transactions have been carried out, most of which are mainly Bitcoin transactions, because it is highly volatile. Buying close when the price is low and selling it when the price is high can earn more money. The cumulative return after each trade is shown in Figure 14. The we stop investing after reaching a maximum gain of \$323,841.52 on October 1, 2021.

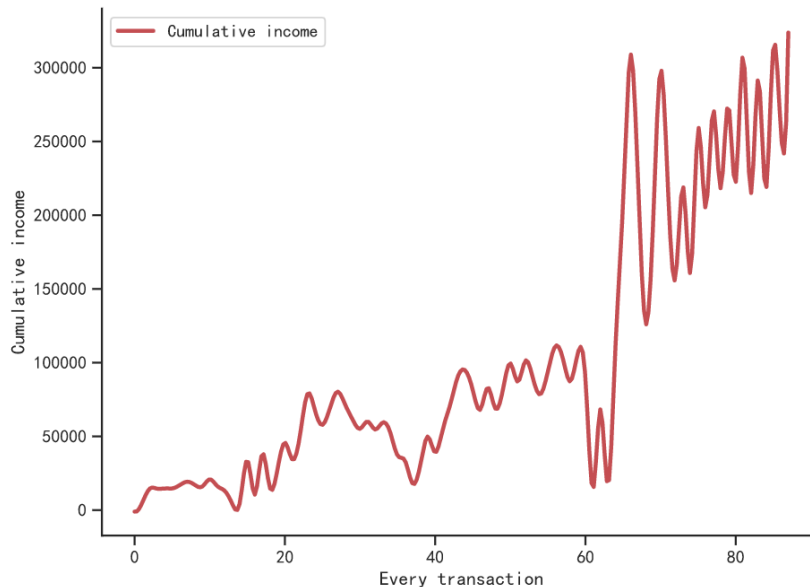


Figure 14. Cumulative gain after each transaction

5. Best Evidence for the Reasonableness of the ARIMA Model

We provide evidence for the ARIMA model as the best

strategy from two aspects: the rationality of the ARIMA model itself and the comparison of the accuracy of mainstream machine learning algorithms.

5.1. Rationality of the ARIMA Model

The key to the rationality of the ARIMA model lies in the

composition of the stationary time series, that is, the data of the steps carried out needs to have certain rules to form a linear relationship. Through observation, we found that the data provided is weakly stationary, so it is necessary to first differentiate into a stationary time series.

## 5.2. Comparison with Mainstream Algorithms

We compare the ARIMA model with other mainstream algorithm models, as shown in Table 3.

**Table 3.** Algorithmic comparison

Algorithm name	Scope of application	Features
ARIMA	Linear data	Simple implementation, good accuracy, and an inability to fit the nonlinear data
Ridge	Linear data	Sacrifice unbiased to improve data stability
KNeighborsClassifier	Numerical type, nominal type	High accuracy and insensitive to outliers
RandomForestClassifier	unrestricted	Processing of high-dimensional data, accuracy is affected by the noise
SVM	Second classification field	Support for multidimensional processing, inefficient and sensitive to missing data

For the same data, these algorithms are predicted. As shown in Table 4, the ARIMA model is correct at 0.75 which is significantly higher than other models with stable data. From the analysis process, the ARIMA model considers the data of each past period, while other machine learning algorithms only consider the data results of this period.

**Table 4.** Comparison of the algorithm accuracies

Algorithm name	Accuracy
ARIMA	0.749
Ridge	0.453
KNeighborsClassifier	0.484
RandomForestClassifier	0.491
SVM	0.503

## 6. Evaluation of the Model

### 6.1. Advantage

- The ARIMA model has a high degree of fit through the smoothing and white noise tests.
- The model is relatively simple, only endogenous variables are needed and no other exogenous variables are needed, which reduces the interference of exogenous variables.
- It can repeatedly perform stationary fitting of time series data, and process weakly stationary data by difference to obtain a new and more stationary time series.

### 6.2. Disadvantage

- The model data has higher requirements on stability, only pretreatment can form better results.
- Essentially a linear regression, it can only capture linear relationships, not nonlinear relationships.

## References

- [1] Huiping Huang, Bloomberg. Market Valuation and Value Investment Strategy — Based on China Securities Market [J]. Accounting Research, 2010, (10): 40-46 + 96.
- [2] Xin Yang, Zhanhui Chen. Empirical research on inertia and reverse investment strategy in Chinese stock market [J]. Journal of Tsinghua University (Natural Science Edition), 2004, (06): 758-761.
- [3] Peng Zhang. No comparative mean-variance and mean-VaR portfolio studies without short selling [J]. China Management Science, 2008, (04): 30-35.
- [4] Lei Yuan. Study on Stock Index Prediction based on the ARIMA-LSSVM mixed model [D]. Harbin Institute of Technology, 2015.
- [5] Tianlun Yu. Research on Stock Investment Strategy Based on Piotroski Method and ARIMA-SVR Model [D]. South China University of Technology, 2016.
- [6] Liping Xu, Mingzhi Luo. Short-term analysis and prediction of gold prices based on the ARIMA model [J]. Financial Science, 2011 (01): 26-34.
- [7] Yan Zhang. Prediction of the RMB exchange rate based on the ARIMA and GRNN combination model [D]. Hubei University of Technology, 2017.
- [8] Jingwen Fei. Analysis and prediction of Gold Futures Price in China based on ARIMA model [J]. Contemporary Economy, 2017 (09): 148-150.
- [9] Libin Zhu. Application of the ARIMA model in stock market prediction [J]. Jiangsu statistics, 1999 (01): 27-28.
- [10] Shuang Ji. Predictive and empirical study of the RMB exchange rate based on the ARIMA-SVM combination model [D]. Jiangxi University of Finance and Economics, 2019.
- [11] Lei Ding, Wanshan Guo. Gold price prediction study based on the ARIMA-GARCH family mixed model [J]. Journal of Xuchang College, 2019, 38 (06): 124-129.