

# Gold or Bitcoins based on ARIMA

Haotian Lv<sup>\*,#</sup>, Yujie Mou<sup>#</sup>, Jiasheng Li<sup>#</sup>

College of Mechanical Engineering, Shanghai Jiaotong University, Shanghai, China

\*Corresponding author: haotianzhijiaozi@sjtu.edu.cn

<sup>#</sup>These authors contributed equally.

**Abstract.** To be or not to be is the question that Hamlet thinks about day and night. Gold or Bitcoins is an inescapable choice for investors. With the ever rising and falling price of gold and bitcoin, making good trading decisions is of paramount importance. In this paper, we systematically investigate how data can be used to quantify the factors that influence trading and make the final decision. We build time series with the prices of gold and bitcoin for the past five years. We obtained forecast curves with excellent fit by seasonality analysis and ARIMA time series model forecasts.

**Keywords:** Quantitative Investments; Time Series; ARIMA.

## 1. Introduction

### 1.1 Problem background

Market traders often trade volatile assets to earn returns. By buying low and selling high, they can make a profit from the spread. Bitcoin and gold are the two primary investments. By forecasting asset price movements and measuring risk, investors can determine whether they should buy, hold or sell assets in their portfolio today. However, due to the complexity of changes in the investment market and the bias of people's perceptions, qualitative analysis based on experience and guesswork can hardly guarantee the optimality of decisions. Therefore, we need to establish suitable mathematical models and formulas quantitatively, write efficient programs using computer technology, study and analyze the future returns and risks of financial products, judge the probability of various market movements, and issue buy and sell orders through the program to realize the trading of investments.

### 1.2 Problem Restatement

- Develop a model that gives the best daily trading strategy only on the basis of price data up to that day. Use your model and strategy to determine what is the value of an initial \$1,000 investment on September 10, 2021.
- Prove that your model provides the best strategy for daily trading.
- Determine the sensitivity of your strategy to transaction costs and discuss how transaction costs influence strategy and results.
- Write a memo of no more than two pages to communicate your strategies, models, and results to traders.

## 2. Model Assumptions

We make some general assumptions to simplify our model. These assumptions are listed below.

1. Gold and Bitcoins both use t+0 trading mode, which means unlimited buying and selling on the same day.
2. Gold and Bitcoins prices do not move during the day, and traders determine whether to buy, hold or sell assets in their portfolios today based on past daily price flows to date and price forecasts for tomorrow.
3. Disregard futures, options, and other trading models, gold and bitcoin instant buy and sell.
4. We do not take inflation and time value into account.
5. Bitcoins can be traded daily, but gold is only traded on open days, namely days other than double holidays and statutory holidays.

### 3. Model Preparations

#### 3.1 Notations

We list the symbols and notations used in this paper in Table 1.

#### 3.2 Data processing

- Develop a model that gives the best daily trading strategy only on the basis of price data up to that day. Use your model and strategy to determine what is the value of an initial \$1,000 investment on September 10, 2021.

- In the official data supplied, daily gold prices are listed for market opening days between November 9, 2016, and October 9, 2021, but some of the data is missing. We derived the predicted values of the missing points by judging the linear trend of the neighboring points and regressing them with the number of periods as  $x$  and the time series values as  $y$ .

- Since gold is only traded on days when the market is open, the amount of gold held by the trader remains the same on non-open days. We filled the amount of gold for a non-open day with the amount of gold from the previous day of that day.

- To measure the movement of gold and bitcoin prices with the same criteria, we normalize the data. We use the min-max normalization method and perform a linear transformation on the original data. The transformation function is as follows.

$$x' = (x - \min A) / (\max A - \min A)$$

Normalization allows variables of different dimensions to be numerically comparable.

### 4. Predictive Model: Time Series Analysis Based on ARIMA Model and Seasonality Test

#### 4.1 Time series

##### 4.2.1 Numerical change pattern

There are four categories of numerical decomposition results for time series, namely, long-term variation trend  $T$ , seasonal variation pattern  $S$ , cyclic variation pattern  $C$ , and irregular variation  $I$ .

##### 4.2.2 Superposition models and multiplication models

Expressing the eventual change in the value of the indicator in terms of  $Y$ , then

1. If the four changes are independent of each other, then the superposition model can be expressed as  $Y = T + S + C + I$

2. If there is a mutual influence relationship between the four variations, then the product model should be used:  $Y = T * S * C * I$

#### 4.2 Forecasting using ARIMA time series analysis model

##### 4.2.3 Stability test

In the time series plot of the seasonality detection, we can determine that the time series of the prices of the two products are not smooth. To be able to forecast using the ARIMA model, we transformed the non-stationary data into stationary data using differencing. The time series plot after first-order differencing is shown in Figure 5.

In the time series plot after first-order differencing,  $B$  value and  $G$  value are symmetrically and uniformly distributed on both sides of the mean. We can see that

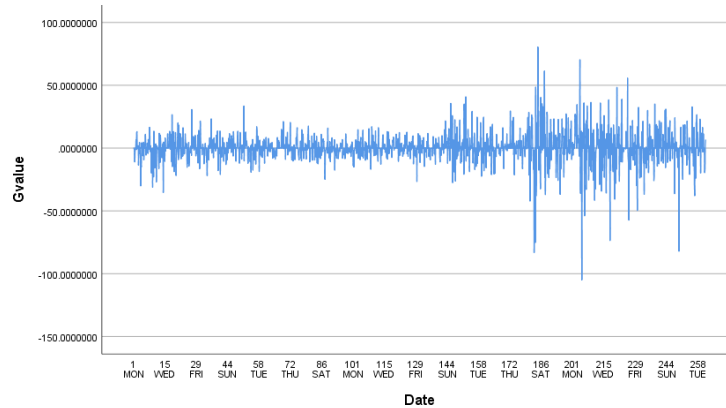


Figure 1: Time series plot after first-order differencing

- Deviation from the mean value is small except for a few outliers
- The variance is relatively small
- There is no trend and seasonality

So we can consider the time series as having stationarity.

#### 4.2.4 Building ARIMA model

First, we exclude the outliers (outliers) from the gold and bitcoin prices.

We use seasonal and time series graph features to make predictions by building appropriate ARIMA models for Bvalue and Gvalue, respectively. The number of differential transformations has all been determined as 1. Using spss software, we obtained the programming results as in Table 3.

Table 1: Model Description

Model ID	Model Type	
Bvalue	model 1	ARIMA(0,1,0)(0,0,0)
Gvalue	model 2	ARIMA(0,1,0)(1,0,1)

The models corresponding to Bvalue and Gvalue are listed in Table 4. The p-value in the AR model part and the q-value in the MA model are zero for both, i.e., they are autoregressive on themselves and no moving average is required. In addition, Gvalue has an additional seasonal component compared to the B value.

Model parameters are shown in Table 2.

#### 4.2.5 Model evaluation

1. Model fit

Table 2: ARIMA Model Parameters

			Estimate	SE	t	Sig.
<b>Bvalue</b>	Natural Logarithm	Constant	0.002	0.001	2.658	0.008
		Difference	1			
		Difference	1			
<b>Gvalue</b>	No Transformation	AR, Seasonal Lag 1	0.950	0.068	13.889	0.000
		MA, Seasonal Lag 1	0.960	0.062	15.485	0.000

- Stationary R-squared: The Stationary R-squared of Bvalue and Gvalue reach 0.243 and 0.357, respectively, which are significantly different from 0. This reflects the clear advantage of the model over the baseline model.

- R-squared: The R-squared value of Bvalue is 0.997 and Gvalue is 0.999, which reflects the excellent fit and accuracy of the model to the time series.

2. White noise test

- Q-test: As is shown in Table 5, the p-values obtained from the Q-test of the residuals by B value and G value are 0.366 and 0.521, respectively, both of which exceed 0.05. Therefore, we cannot assume that the residuals are white noise series. Therefore the two ARIMA models can

identify the sales data in this example very well.

Table 3: Data obtained from Q-test

Model	Number of Predictors	Model Fit statistics			Number of Outliers
		Stationary R-squared	Ljung-Box Q(18) Statistics	DF Sig.	
Bvalue	0	0.243	19.425	18 0.366	18
Gvalue	0	0.357	29.531	16 0.521	26

ACF/PACF test: The plots of ACF and PACF for the residuals are shown in Figure 6. As seen in Figure 6, the autocorrelation and partial autocorrelation coefficients for all lag orders are not significantly different from 0.

4.2.6 Predictions

Combining the model parameters in Table 4, we are able to make reasonable forecasts for the prices of gold and bitcoin using the two ARIMA models obtained.

The SARIMA model has the following form:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{i=1}^p \phi_i L^{mi})(1 - L)d(1-Lm)Dy_t = \alpha_0 + (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{i=1}^Q \theta L^{mi})\epsilon_t \tag{1}$$

By substituting the parameter estimates in equation (1), we obtain respectively:

$$ARIMA(0, 1, 0)(0, 0, 0) : y_t = \alpha_0 + \epsilon_t + y_{t-1} \tag{2}$$

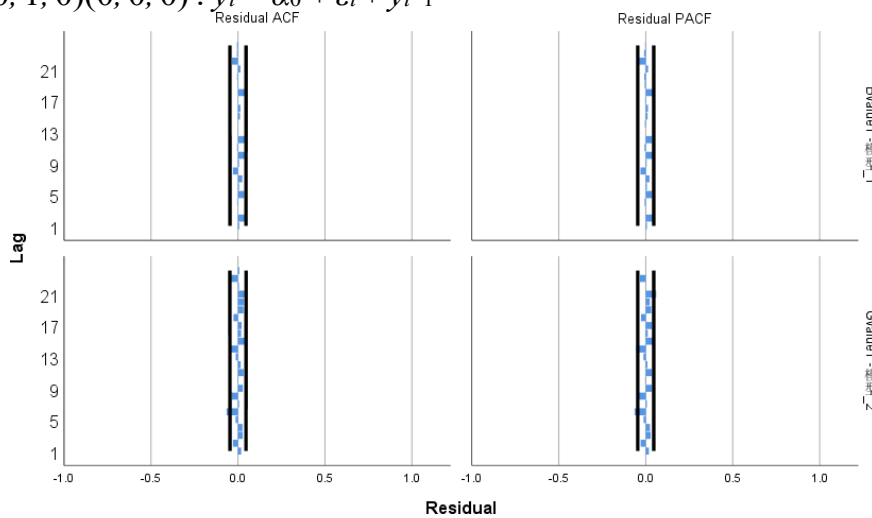


Figure 2: The plots of ACF and PACF test

$$ARIMA(0, 1, 0)(1, 0, 1) : y_t = \alpha_0 + \epsilon_t + \phi_1 y_{t-7} - \phi_1 y_{t-8} + y_{t-1} + \Theta_1 \epsilon_{t-7} \tag{3}$$

The predicted prices of gold and bitcoin can be obtained from the data and equation (2)(3), as shown in Table 4.

Table 4: The predicted prices of gold and bitcoin

Model		261 Sat	262 Sun	262 Mon	262 Tue	262 Wed	262 Thu	262 Fri
Bvalue	Forecast	46505.88	46643.48	46781.49	46919.91	47058.73	47197.97	47337.61
	UCL	49936.20	51562.07	52873.52	54024.37	55073.99	56052.74	56978.59
	LCL	43253.14	42081.03	41225.11	40531.61	39941.15	39423.37	38960.30
Gvalue	Forecast	1794.63	1794.64	1795.15	1794.81	1794.90	1794.65	1794.55
	UCL	1812.91	1820.48	1826.81	1831.36	1835.76	1839.41	1842.90
	LCL	1776.35	1768.79	1763.50	1758.26	1754.03	1749.88	1746.19

## 5. Decision Model Based on Linear Program

### 5.1 Model building and solving

The amount of gold(bitcoins) owned on day  $i+1$  is equal to the sum of the amount of gold(bitcoins) owned on day  $i$  and the amount of gold(bitcoins) traded on day  $i$ .

$$G_{i+1} = G_i + g_i \tag{4}$$

$$B_{i+1} = B_i + b_i \tag{5}$$

The number of dollars owned on day  $i+1$  is equal to the number of dollars owned on day  $i$  minus the number of dollars spent on gold or bitcoin on day  $i$  and commissions.

$$C_{i+1} = C_i - gValue(i) * |g| - g * gValue - bValue(i) * \alpha_b * |b| - b * bValue(i) \tag{6}$$

On day  $i$ , we take day  $i+1$  asset maximization as our goal. This is because on day  $i$ , we do not know the prices of gold and bitcoin on day  $i+1$ . Therefore, our objective function can only be based on the price of gold and bitcoin predicted on day  $i+1$ .

Our objective function is as follows.

$$\text{maxgoal1} = C_{i+1} + G_{i+1} * gValue\_p(i + 1) + B_{i+1} * bValue\_p(i + 1) \tag{7}$$

According to the actual situation, the constraint conditions are that the number of dollars, gold and bitcoins after the transaction are not less than zero.

$$\begin{cases} G_{i+1} \geq 0 \\ B_{i+1} \geq 0 \\ C_{i+1} \geq 0 \end{cases} \tag{8}$$

We use matlab to solve the linear programming problem to get the number of gold and bitcoins that should be traded daily, as shown in the Figure 7(a).

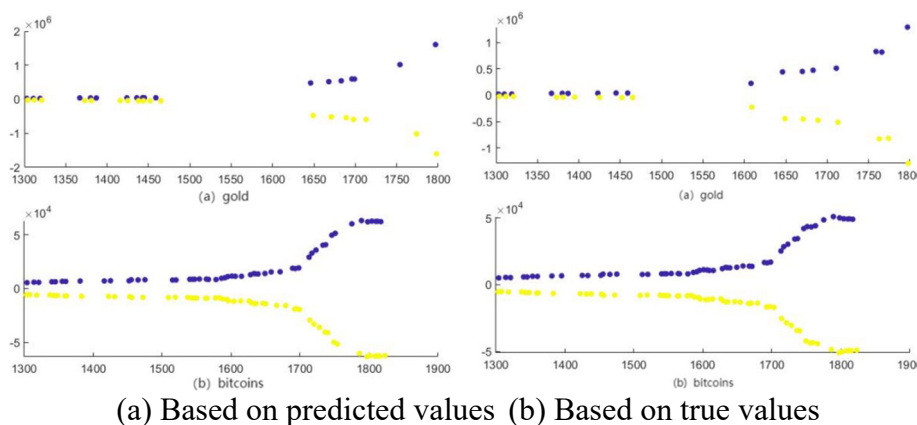


Figure 3: The number of gold and bitcoins that should be traded daily

We calculated the total value of assets owned on day  $i+1$  based on the number of gold and bitcoins bought and sold on day  $i$  obtained by solving.

$$\text{Value}(i + 1) = C_{i+1} + G_{i+1} * gValue(i + 1) + B_{i+1} * bValue_p(i + 1) \tag{9}$$

According to the results derived from the linear programming model, on October 9, 2021, the initial investment value of the starting \$1,000 is worth  $10^{9.403} \approx 2.529 * 10^9$  dollars.

### 5.2 Model evaluation

To test the accuracy of the results we obtained, we stand in God’s perspective and substitute the actual price of gold and bitcoin on day  $i+1$  into the objective function.

$$\text{maxgoal2} = C_{i+1} + G_{i+1} * gValue(i + 1) + B_{i+1} * bValue_p(i + 1) \tag{10}$$

We solve the linear programming with the same constraints (Equation 8) to obtain the number of gold and bitcoins that should be traded daily, as shown in the figure6(b).

Likewise, we make a plot with the number of days as the x-axis and the logarithm of the total value of assets owned on day  $i$  as the y-axis, as in Figure 8.

As can be seen in Figure 8, the difference between the optimal solution derived from linear programming with predicted and true values is small. This also reflects the high accuracy of our

predictions for gold and bitcoin prices.

## References

- [1] Yao, H. C., Lai, J. W., Xia, S. H., Chen, S. M.. Fuzzy trading decision based on Apriori algorithm and neural network[J]. System Science and Mathematics,2021,41(10):2868-2891.
- [2] Wang Huiying,Hao Yongtao. Stock price trend prediction algorithm based on technical indicators and random forest[J]. Modern Computer,2021,27(27):43-47+52.
- [3] LiuLJ,LiangGP,WangHT.Researchonquantitativetradingstrategysystemmodel[J]
- [4] .ShanghaiBusiness,2021(09):158-159
- [5] Wang M T. Research on futures price trend prediction and quantitative trading strategy based on machine learning [D]. Nanjing University of Finance and Economics,2021.
- [6] Gu Wentao, Wang Ru, Zheng Suhao, Yang Yongwei. Research on financial market return direction Prediction Model based on text big data Method [J]. Statistical research, 2020 ((11) : 68-79. The DOI: 10.19343 / j.carol carroll nki. 11-1302 / c. 2020.11.006.
- [7] Zhu Junlin, Fu Yingzi, Chen Yi. Computer technology and development,2013,23(12):168-170+174.
- [8] Sun Yi-xiao, Zhang Li. Research on portfolio model and its algorithm [J]. Industrial Technology Economics,1994(01):45-48.
- [9] Zhang Haipeng, Zhu Jiaming. Journal of henan university of science and technology (natural science edition),2020,48(04):62-70.
- [10] Wang Tong. Research on quantitative trading strategy based on EMV index in China's A-share market [J]. Tax, 2020, 14 (01) : 199.
- [11] Huang Wenlan. Portfolio optimization based on multi-objective evolutionary algorithm research [D]. Harbin business school, 2020. The DOI: 10.27787 / , dc nki. GHRBS. 2020.000514.