Financial Engineering and Risk Management (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2022.050312
ISSN 2523-2576 Vol. 5 Num. 3

# *Research on gold and bitcoin portfolio strategy based on ARIMA time series prediction model*

**Zhixuan Hong[1], Ruiyu Yang[2], Dongsheng Liu[3]**

[1] *School of Automation (School of Artificial Intelligence), Hangzhou Dianzi University, Hangzhou, 310020, China*
[2] *School of Electronics & Information (School of Micro-electronics), Hangzhou Dianzi University, Hangzhou, 310020, China*
[3] *Hangzhou Dianzi University-ITMO University Joint Institute, Hangzhou Dianzi University, Hangzhou, 310020, China*

***Keywords:*** The best investment strategy, Gold and bitcoin, Investment strategy model, ARIMA time series prediction model, The best transaction model

***Abstract:*** In order to formulate the best investment strategy, this paper takes gold and bitcoin as examples to establish an investment strategy model. The model is based on ARIMA time series prediction model. Finally, the best transaction model is determined on this basis.

## 1. Introduction

In recent years, with the rapid development of information technology and the change of financial engineering technology, quantitative investment relying on data modeling and programmed trading has gradually emerged, and as a new investment method, it has gradually become the focus of attention.

We will formulate gold and bitcoin portfolio strategies based on ARIMA time series prediction model. Finally, formulate the investment strategy model. The research results provide guidance for stock market investment.

## 2. ARIMA Model-based Stock Price Forecasting

### 2.1 Fundamentals of the ARIMA model

The ARIMA model is a combined model, containing an autoregressive model, the moving average model and the difference model, resulting in a differential autoregressive moving average model *ARIMA* (*p,d,q*) [1], where d is the number of orders that need to be differenced for the data.

The model can be expressed as follows:

$$y_t' = a_0 + \sum_{i=1}^{p} a_i y_t' + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-1} \quad (1)$$

$$y_t' = \Delta^d y_t = (1-L)^d y_t \quad (2)$$

$$(1-\sum_{i=1}^{p} a_i L^i)(1-L)^d y_t = \alpha_0 + (1+\sum_{i=1}^{q} \beta_i L^i)\varepsilon_t \quad (3)$$

↑*AR(p)* ↑*differential* ↑*represents MA(q)*

## 2.2 ARIMA Model Building

There are generally four stages in model building, which are data preprocessing, smoothness test, model identification, order fixing and model testing. Since the actual use of the model for calculation needs to be based historical data before each day to predict the price on next day, so the number of predictions is huge. Meanwhile the model obtained from each solution is also different, so in this paper we choose some of the pre-diction results as a case study for analysis. Assuming that the daily bitcoin price is $BIT_i$, the gold price is $Gold_i$, the original bitcoin price sequence is $\{BIT_1, BIT_2, BIT_3, ..., BIT_i\}$, the original gold price sequence is $\{GOLD_1, GOLD_2, GOLD_3, ..., GOLD_i\}$ and the prediction value of the $i$th day is obtained from the historical data before the first $BIT_{40}$ day $BIT_i$ and so on, each time the prediction value is calculated from the data of all the previous days. In this example, we selected $BIT_1$ to $BIT_{40}$ in order to predict $BIT_{41}$.

The data in the original data are discontinuous therefore interpolation is needed to achieve better prediction effect. We reckon interpolation can reasonably compensate for the missing in the data. This model uses a linear interpolation method. The values are estimated by taking the average of the two data points adjacent to the left and right of the point in the data series that needs to be interpolated. according to the linear interpolation model as follows:

$$\begin{cases} BIT_i = \dfrac{BIT_{i-1} + BIT_{i+1}}{2} \\ GOLD_i = \dfrac{GOLD_{i-1} + GOLD_{i+1}}{2} \end{cases} \quad (4)$$

Smoothness is an important characteristic of time series, so we first need to test the smoothness of the series. The ADF unit root test is used here.

The timing diagram and the autocorrelation function diagram are first made using Matlab [2,3].

We found that the overall price trend is upwards, so it is possible that it is not a smooth series (Fig. 1). From the autocorrelation function graph can be seen that the image is a sinusoidal fluctuation pattern, which does not meet the characteristics of a smooth series, and the ADF test using Matlab results in zero, therefore we need to transform the series into a smooth series.

The difference operation on the original series is performed to transform the original series into a smooth series, and the correlation function plot method and the ADF test are applied to continue the test on the difference series, and the results are shown as follows (Fig. 2,3).

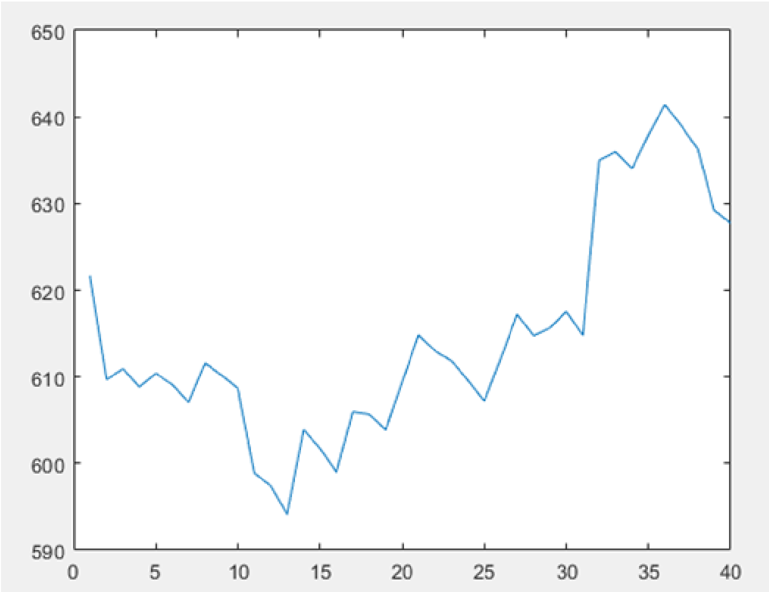


Fig. 1 Time series of prices

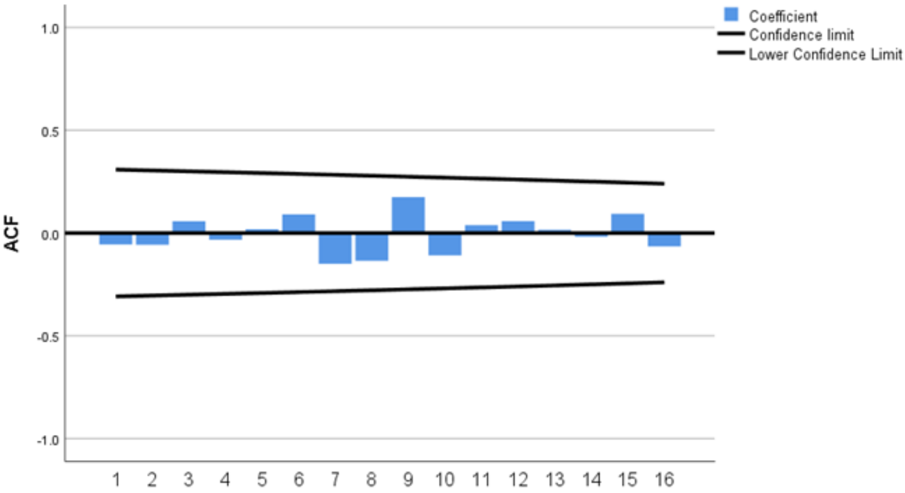


Fig. 2 Autocorrelation function diagram

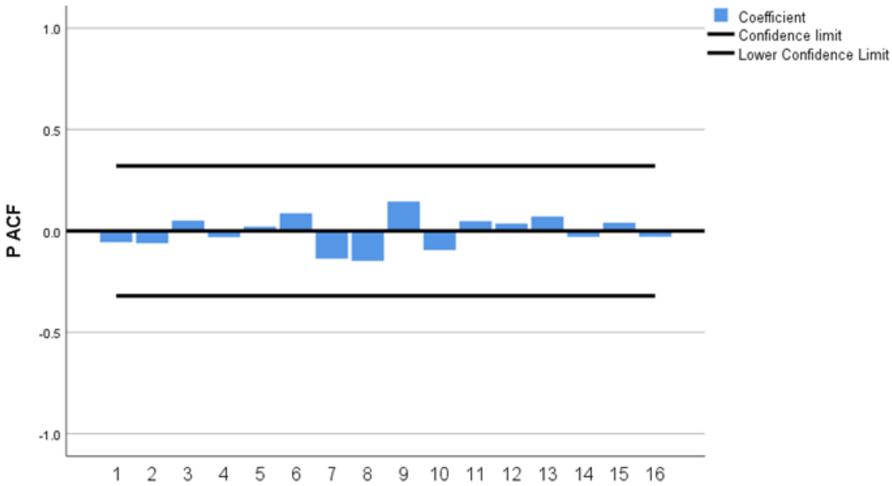


Fig. 3 Plot of partial autocorrelation function

As can be seen from Fig. 2, Fig. 3, both the autocorrelation coefficient and the partial autocorrelation coefficient decrease rapidly to zero or turn into zero after a certain order of calculation, which suggested it's a smooth series; the ADF test also shows that the difference series has become a smooth series and the next step can be carried out.

The problem of determining the order equals to determine the *p, d, q* three parameters, and since this series is transformed into a smooth series after one difference, the order of that *ARIMA* (*p, d, q*) can be determined, i.e $d$ = 1. The main issue here is the determination of the $p$ sum $q$. We use the autocorrelation function $ACF$ and the partial autocorrelation function $PACF$ to determine.

The correlation between all observations $X_t$ , $X_{t-1}$, $X_{t-2}$, $\cdots$, $X_{t-p}$ of the time series $X_t$ is called autocorrelation, and this relationship between observations can be measured by defining an autocorrelation function. If a series $X_t$ is a smooth time series, then its autocorrelation coefficient is defined in the form of the following equation.

$$\gamma_k = \frac{cov(X_t, X_{t-k})}{\sqrt{DX_t \cdot DX_{t-k}}} \quad (5)$$

$k$ is thought to be a variable of the equation. Both the $AR(p)$ model and $\gamma_k$ in the ARIMA (*p, d,q*) model are decreasing in some negative exponential form, this feature is called $\gamma_k$ the $p$ step tail, that is, its value does not quickly converge to zero, but in a slower form to 0. For the MA(q) model, $\gamma_k$ it will converge to zero immediately at this point in time $k > p$ later, this situation is called $\gamma_k$ the q step truncation tail.

According to the previous correlation function graph, we determine the model as ARIMA (0, 1, 2), so that the parameters of this prediction model are selected, the prediction model is initially established, and the model can be tested in the next step (Table 1).

Table 1 ARIMA model parameters (Bitcoin price prediction model)

| | Estimate | Standard error | t | Significance |
|---|---|---|---|---|
| Constant | 0.156 | 0.862 | 0.18 | 0.858 |
| Difference | 1 | - | - | - |

$$(1 - \sum_{i=1}^{p} a_i L^i)(1 - L)^d y_t = \alpha_0 + (1 + \sum_{i=1}^{q} \beta_i L^i)\varepsilon_t \quad (6)$$

## 2.3 Model Testing

After the estimation step of the model parameters, the established model must also be tested to check the suitability of the currently selected model, remodel is needed if failed. The residual white noise test and parametricity test are performed to determine whether the model built is desirable or not.

From the ACF and PACF graphs of the residuals (Fig. 4), it can be seen that the auto-correlation coefficients and partial auto-correlation coefficients of all lag orders are not significantly different from zero. In addition, the p-value obtained from the Q-test of the residuals is 0.973, which means that we cannot reject the original hypothesis that the residuals are white noise series. At the same time, the model $R^2$ is 0.819, which has a high degree of fitness. According to the model, our predicted the value of $BIT_{41}$ is 627.88. In this way, we can get the predicted value of each day's price.

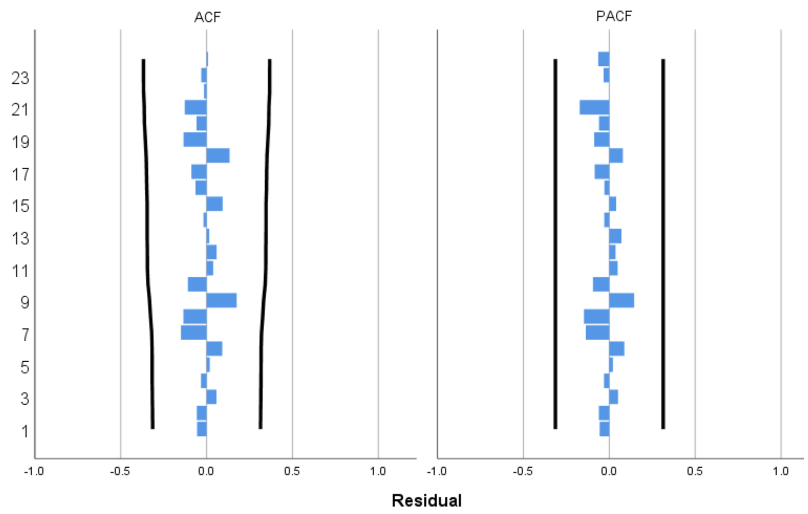


Fig. 4 White noise residual test

## 3. Investment Strategy Model

### 3.1 Quantitative Timing Strategy Model - Double Average Strategy

For the calculation of the transaction fee at the time of purchase and sale the formula is as follows.

$$P_{tax} = P_t^{(i)} \times \alpha^{(i)} \quad (7)$$

Where $P_{tax}$ denotes the commission for the transaction, $P_t^{(i)}$ denotes the price of gold or bitcoin at time $t$, $\alpha^{(i)}$ denotes the respective commission ratios for gold and bitcoin, 0 represents gold and 1 represents bitcoin.

The formula for calculating cumulative earnings is as follows:

$$W = \sum (P_{t2}^{(1)} - P_{t1}^{(i)} - P_{buytax} - P_{selltax}) \quad (8)$$

### 3.2 Construction of a Single-strategy Multi-variety Portfolio - the ”1/N” Strategy [4]

The ”1/N” strategy is one of the simplest and most intuitive ways to construct a portfolio, which in this paper means investing equal amounts of money in different commodities at the beginning of the investment [5]. Since strategies in each species are independent trading system, there is no interaction between different trading strategies, any strategy in any species open and close positions or profit and loss does not affect the opening and closing of other trading strategies.

### 3.3 Moving Average Model (MA)

According to Wikipedia’s definition, The moving average is a statistical calculation that analyzes data points by calculating a series of averages over various subsets of the entire data set. With time series data, moving averages are commonly employed to smooth out short-term variations and emphasize long-term trends or cycles. Its calculation formula is as follows:

$$MA_t(n) = \frac{1}{n} \sum_{i=t-n+1}^{t} P_i \quad (9)$$

Where $M_{at}$ denotes the simple moving average value at moment $t$, $n$ denotes the period being averaged, and $p_i$ denotes the price of the financial product during the period. There are many types of moving averages, generally divided into short-term averages and long-term averages according to the length of time. Moving averages W time units as the basis, commonly used are 5 days (weekly average), 10 days (semi-monthly average), 20 days (monthly average), 60 days (quarterly average), 120 days (semi-annual average) and 240 days (annual average). The averages of different periods reflect the stock price trend over different lengths of time, represent the average cost over time, and provide the basis for different types of trading strategies (short term, medium term, and long term) [6].

### 3.4 Traditional Dual Average Strategy

Double SMA trading system is a classic trading strategy. It is well known that financial markets have obvious trend characteristics, but because markets are always in the process of fluctuation, under the micro-trading behavior, asset prices change rapidly in the short term and are in a process of repeated up and down oscillations for a long time. With the moving average model (Eq. (1)), we can well filter those invalid noise trading and mathematically offset unneeded price changes to reflect long-term trends in asset prices [5].

The operation principle of the Double SMA system is relatively simple. The short-period SMAs have a short calculation window, representing the short-term price trend, which is less affected by the long-term price trend. While the long-period SMAs have a long calculation window, reflecting to a certain extent the long-term trend of the securities price. Using the long and short averages, investors can visually see the changes in the price trend of the averages in the short and long term therefore buy and sell transactions accordingly. We set the two averages are Slow and Fast, and their periods are Ts, $T_l$. Where $T_l > T_s$, double average trading strategy trading principles are as follows: when the short-term moving average from the bottom up through the long-term moving average, called the ”golden cross”[5], indicating that the price of securities has a strong upward trend in the short term while the upward momentum is strong, which is a buying opportunity; when the long-term moving average is broken by the short-term moving average from the top, a ”dead fork” appears, it means that the short-term price has strong downward momentum, and the market outlook is bearish, which is a typical sell signal [7].

And if we just rely on the signal of the golden fork dead fork, it is likely that there will be a loss just buy, so the last question to establish a time series model is very important, we set t and all the time before the price flow through the formula below to get t + 1 day financial product prices, if the price falls the next day, then we still will not buy, if the next day is up, then we do not do the sell operation. That is, we use the prediction of the time series as the second judgment signal, whose mathematical expression is:

$$buysignal_2 : predictP_{t+1} > P_t \quad (10)$$

$$sellsignal_2 : predictP_{t+1} > P_t \quad (11)$$

Later, according to the ”1/N” strategy, we use this modified traditional double mean strategy to invest in a combination of gold and bitcoin. We review the paper and find that the best results are generally obtained when $T_l = 30$ and $T_s = 5$[4].

### 3.5 Dual Averaging Strategy Based on Bull-Bear Market

In order to improve the winning rate of the SMA strategy and reduce the losses caused by frequent stop-losses, we have reviewed the paper [8] and referred to the improved ideas in which we use the arrangement of the three periodic SMAs as the current market trend judgment. Let the three long-period averages be $F_1$, $F_2$, and $F_3$, it is advisable to set the calculation period of the three averages; when the averages show a long alignment, i.e., the average $F_3$ is above the average $F_2$, while $F_2$ is above $F_1$, the current market is judged as a long market, and the mathematical expression is as follows:

$$bullmarket: MA_t(F_3) > MA_t(F_2) > MA_t(F_1) \qquad (12)$$

$$bearmarket: MA_t(F_3) < MA_t(F_2) < MA_t(F_1) \qquad (13)$$

We take $T_l = 30$ and $T_s = 5$. Randomly taking $F_1 = 60$, $F_2 = 50$ and $F_3 = 40$, we apply python and get the following return volatility graph. The final return in 3051.87 USD with a maximum retracement of 13. 34%.\ The fluctuation graph of the return is obtained by python as follows (Fig. 5)

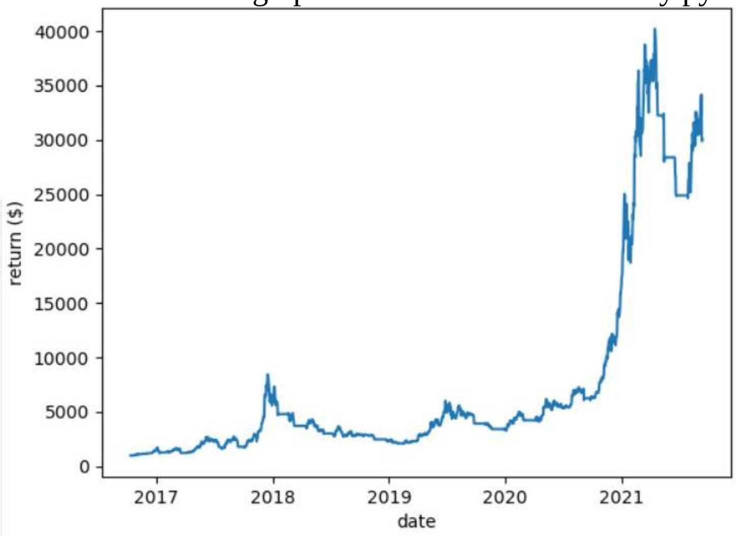


Fig. 5 Returns - Time

### 3.6 Dual Averaging Strategy Based on the Target Profit and Loss Limit

In order to facilitate the balance of return and risk, and to maximize the use of capital, we replace the sell point of the double average trade from the dead cross to the target profit and loss limit parameters.

The take-profit parameter is the mathematical expression that we will choose to sell the take-profit when the current price over the bid price reaches the take-profit parameter, which is:

$$stop\_profit\_signal: \frac{P_{buy}}{P_{now}} > \sigma_1 \qquad (14)$$

The loss limit parameter is that we will sell all the stock when the current price over the bid price reaches the loss limit parameter.

$$stop\_loss\_signal: \frac{P_{buy}}{P_{now}} < \sigma_2 \qquad (15)$$

And the rest of the principles are consistent with the previous traditional double averaging strategy. We take the target profit as 1.3, the loss limit as 0.8, and still take $T_l = 30$ and $T_s = 5$. The fluctuation graph of the return is obtained by python as follows (Fig. 6). The final return was 30072.28, with a maximum retracement rate of 17. 98%.

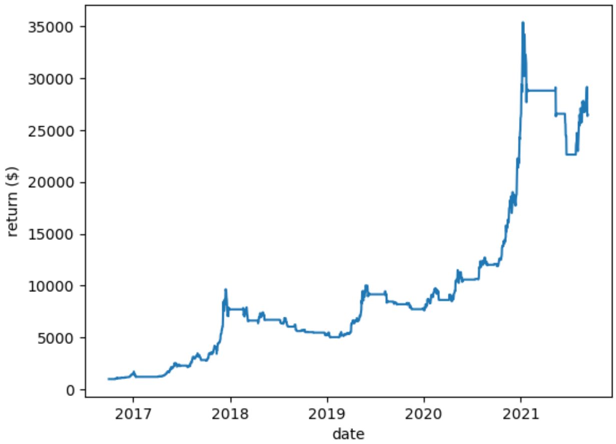


Fig. 6 Returns – Time

## 4. Strengths and Weaknesses

Our decision model for setting stop-loss and take-profit points is simple and easy to implement for the average person. It is also very easy to replicate because it takes into account each person's risk preferences.

However, ARIMA model requires time series data to be stable or stable after differential differentiation. In essence, it can only capture linear relationships, not nonlinear relationships.

## References

*[1] T.H. Qi. Stock price forecasting based on gray model and ARIMA model[J]. Computer Age, 2021(10):83-85+89.*

*[2] H.Y. Tong, L.D. Ma, Z.J. Meng. Motion characteristic analysis of er50a baling robot based on MATLAB [J]. Journal of Taiyuan University of science and technology, 2022,43 (02): 154-159.*

*[3] L. Gu. Application analysis of MATLAB in basic mathematics teaching in Colleges and universities [J]. Heilongjiang science, 2022,13 (05): 150-151.*

*[4] J.F. Fei. Research on multi-variety multi-strategy quantitative investment port-folio in Chinese commodity futures market [D]. Zhejiang University,2018.*

*[5] W.H. Ye. Research on the effectiveness of double-average trading strategy based on CSI 300 full-return index [D]. Guangdong University of Foreign Studies,2016.*

*[6] H.G. Lai. Bitcoin - a new type of "gold" in the digital era [J]. China Business Journal,2018(02):33-34.*

*[7] T.L. Yu. Research on stock investment strategy based on Piotroski method and ARIMA-SVR model[D]. South China University of Technology,2016.*

*[8] Z.C. Zhou. Research on the strategy of individual investment in A-shares [D]. Guangxi University, 2013.*