ISSN: 2688-9323

The Research on Optimal Investment Strategy based on ARIMA and Improved Bolling-line Model

Yuxuan Jia[1], Jiawei Zhang[2], Yuming Feng[2]

[1]The Information Science Technology College, Donghua University, Shanghai, 200051, China

[2]School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai, 201400, China

Abstract

Quantitative investment is an emerging concept of systematic investment approach by using computer programs combined with financial theories to perform investment analysis and trading. This paper focuses on how to design a quantitative trading strategy to optimize the investment strategy, and then design the ARIMA model and an improved Bolling-line model based on optimal iteration with a risk function to facilitate the computation of different scenarios for three different types of trader investments and build a dynamic programming model. And the robustness and sensitivity analysis of the model are tested, and the model performs stably. Finally, the model is less sensitive, has good market adaptability, and has some realistic significance.

Keywords

ARIMA Model; Simple Model; Bolling-line Model; Quantitative Investment; Trading Strategy.

1. Introduction

Volatile assets can earn big returns in smart buying and selling, two of which are gold and bitcoin. Compared with gold, bitcoin has a high yield, high volatility, free of regulation and tax, etc., so it will have a great development space in the financial field. Bitcoin is sometimes called the new gold, replacing gold as a hedge against inflation and as a new safe-haven asset that can complement gold for hedging, which is of great interest to investors of all kinds in the financial sector. In most cases, there is a commission on every sale.

This paper used the data of gold and Bitcoin to design a quantitative trading model to optimize the investment strategy, and perform a sensitivity analysis to evaluate the pros and cons of the model.

2. Model Building and Results

2.1. Gold Prediction Model: ARIMA

2.1.1. Data Description

Table 1. Descriptive statistics

	N	The minimum	The maximum	The mean	The standard deviation
USDPM	1255	1125.70	2067.15	1464.5494	249.29181
Number of active cases (column)	1255				

The maximum unit price of gold is 2067.15, the minimum is 1125.70, the average price is 1464.5494, and the standard deviation is 249.29181. It can be seen that the gold price fluctuates greatly, but still maintains at a certain level.

ISSN: 2688-9323

2.1.2. Model Description

ARIMA (P, D, Q) model (autorepression differential moving average model). This model is a commonly used time series prediction analysis method, which can be considered as the difference combination of AR (P)(autoregression) model and MA(q) (moving average) model [1]. The expression of ARIMA (P, D, Q) model is shown as follows:

$$\begin{cases} \Phi(B)(1-B)^d \chi_t = \Theta(B)\varepsilon_t; \\ \Phi(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p; \\ \Theta(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q; \\ E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s)=0, s\neq t; \\ E(\chi_s \chi_t)=0, \forall s<t. \end{cases} \quad (1)$$

2.1.3. Model Checking and Results

Data seasonality test: Determine whether the time series is stationary according to the autocorrelation coefficient graph of the time series. If the sequence is verified to be non-stationary, it needs to satisfy the stationary condition by different transformation or logarithmic difference transformation. If the sequence after d-order difference satisfies the stationary condition, then d. [2]

Model order determination: According to autocorrelation graph (ACF) and partial autocorrelation graph (PACF), the order p of autocorrelation and the order Q of the moving average are determined, and the model ARIMA (P, D, Q) is determined.

Parameter estimation and diagnosis: the parameter estimation of ARIMA (P, D, Q) model is carried out to determine whether the residual sequence is a white noise sequence. If the residual sequence is white noise sequence, the test is passed [1]. If the fitted model not is able to pass the test, the parameters P and Q are re-determined, and the model is re-selected for fitting.

Finally, ARIMA (P, D, Q) model is used for modeling.

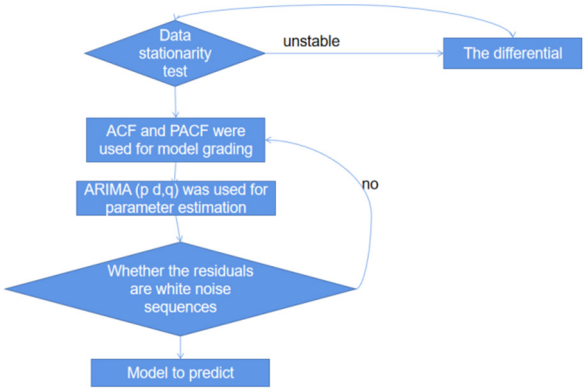

Figure 1. The Prediction Flow Chart of ARIMA (P, D, Q) model

The ARIMA(3,1,3)(1,0,0) model was selected to fit the gold data after verification estimation and diagnosis. The goodness of fit was 0.427, and there were some heteroscedasticity in high order after fitting, but the overall fitting effect was satisfactory.

From the fitting effect, the overall trend is roughly the same. After some extreme values are removed, only high order autocorrelation graphs and partial autocorrelation graphs have some heteroscedasticity. We believe that the model has a good fitting effect.

ISSN: 2688-9323

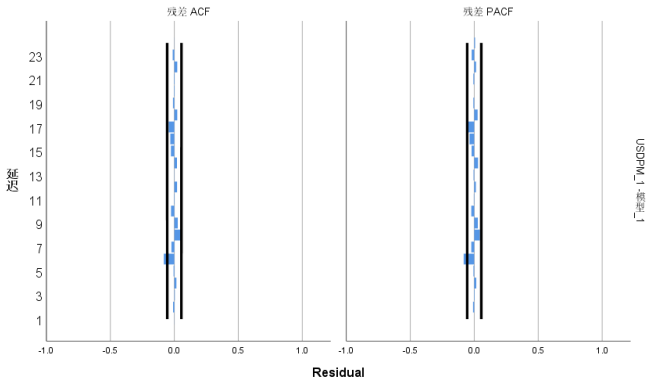

Figure 2. ACF & PACF of GOLD.

Figure 3. Fitting Effect Diagram of Gold.

2.2. Forecast Model for Bitcoin: Simple

2.2.1. Data Description

Table 2. Descriptive statistics

	N	The minimum	The maximum	The mean	The standard deviation
Value	1826	594.08	63554.44	12206.07	14043.89
Number of active cases (column)	1826				

The minimum unit price of Bitcoin is 594.08, the maximum is 63554.44, the mean value is 12206.068, and the standard deviation is 14043.9. It can be seen that the unit price of Bitcoin is extremely unstable, with large fluctuation range.

2.2.2. Model Description

$$\begin{cases} S_t = \hat{\chi}_{t+1}, \\ \hat{\chi}_{t+1} = \alpha\chi_t + (1-\alpha)\hat{\chi}_t. \end{cases} \tag{2}$$

2.2.3. Model Checking and Results

Obviously: each smoothed data is obtained by weighted sum of past data. The closer the data is to the current period, the greater its weight is, indicating that the data closer to the current period has a greater influence on the current period. Conversely, the earlier the data, the less impact on the current period.

The selection principle of smoothing coefficient:

ISSN: 2688-9323

1. If the time series has irregular fluctuation changes, but the long-term trend is close to a stable constant, the value is generally small (0.05-0.02).
2. If the time series has a tendency to change rapidly and obviously, a larger value should be set (0.3-0.5).
3. If the time series changes slowly, a smaller value should also be selected (generally between 0.1 and 0.4).

After checking and estimating the order, we chose the simple exponential smoothing model as the bitcoin fitting model with a goodness of fit of 0.243. It can be seen from the observation of ACF and PACF plots that there is no heteroscedasticity.

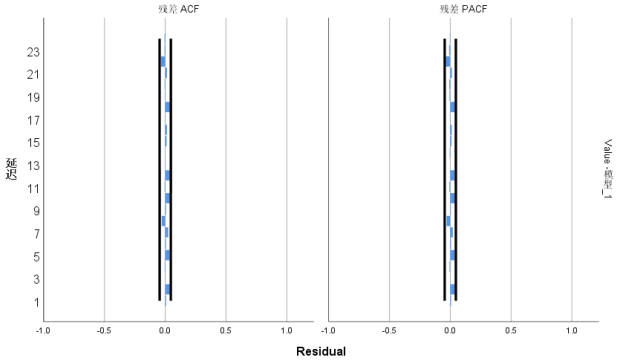

Figure 4. ACF & PACF of Bitcoin

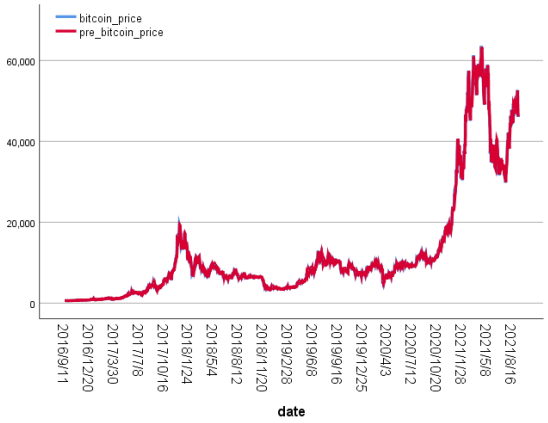

Figure 5. Fitting Effect Diagram of Bitcoin

2.3. Decision Model: Improved Bollinger Line Model based on Optimal Iteration

2.3.1. Traditional BOOL Mode

Bollinger bands are computed using a combination of mean and standard deviation in statistics and are divided into mean, upper rail and lower rail lines.According to the Bollingline mean regression strategy, the stock price floats within the range of the upper and lower tracks, and even if it breaks through the upper and lower tracks in the short term, it will still return to the Bollingband in the long run.Accordingly, once break down track, from buying and sell signal namely.When the stock price exits the upper bound, it is a sell signal, when the stock price breaks the lower bound, it is a buy signal.

BOLL line calculation formula:

Middle rail =N daily moving average

ISSN: 2688-9323

Up rail = middle rail + K standard deviation
Lower rail = middle rail -K standard deviation

The most effective of all the indicators of the Brin line is bandwidth, which shows how much the price fluctuates and is calculated using the standard deviation method, usually 2 standard deviations of the average price.In statistics, standard deviation is used to describe the degree of deviation from the mean.According to the statistics, the 20-day moving average is adopted in the rail line, and the bandwidth designed with two standard deviations can ensure that 95% of the historical price data falls within the bandwidth, so the probability of the price falling outside the range is less than 5%.From a statistical point of view, when the price is in the area outside the Bollinger line, it can be considered as high or low, and there is a small probability time. This judgment is reasonable.

In the traditional Bolling line model, N=20, K=2.But,in the actual financial environment, the probability distribution of the signal line is random, not normal. According to the logic of bolling line breaking through the upper and lower rails with a small probability, we can improve the Bolling line and estimate the actual probability distribution of the signal line. In other words,The optimal N and K for a given financial system at a given time is not necessarily 20 and 2 collimations

2.3.2. The Advanced Mode

When one day price is greater than the rail line, the price and the actual value is greater than the predicted value, and if sold at this point, the value is greater than the check surplus rate of p (p = 1.3) buying value, then sell.

when the price of a one day less than under the rail line and the actual value is less than price forecast, then buy.

According to the data in the past 15 days, we make the variance reciprocal scatter plot, used to measure the level of risk,

In order to reduce risks and improve the stability of the model, we divided the funds invested in bitcoin and gold into two parts, not buying or selling at the same time, and then compared the results of undivided funds and split funds. Since we need to find the optimal N and K of each day, we traverse N and K according to the past and current data and use the improved Bolling line model to calculate the optimal value of N and K of each day, based on which N and K decide whether to buy or sell. Based on the amount of buying and selling determined by the risk level at that time. If we invest gold and bitcoins at 500,500, the total income will be 5099 and the annual yield will be 38%.

2.3.3. Risk Assessment

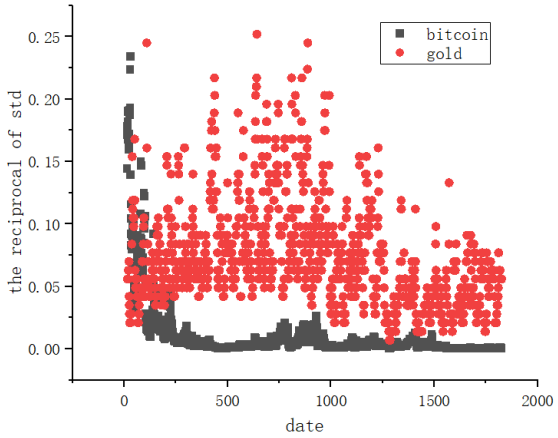

Figure 6. The Reciprocal of STD

ISSN: 2688-9323

The inverse of the standard deviation of the price in the past 15 days, the smaller the value, the greater the risk. We can see that bitcoin has less risk in the early stage and higher risk in the later stage, while gold has less risk in general but higher risk volatility. when we buy or sell, we determine the volume of trading based on the risk level of bitcoin or gold at the time, combined with a number of other indicators.

2.4. Trading Strategy

Different colors represent the divided two parts of the capital, the "+" sign represents buy, the "*" represents sell, the horizontal axis represents the daily price of the stock and the value of the invested capital changes.

If not sold by the deadline, the current price is used to measure the value.

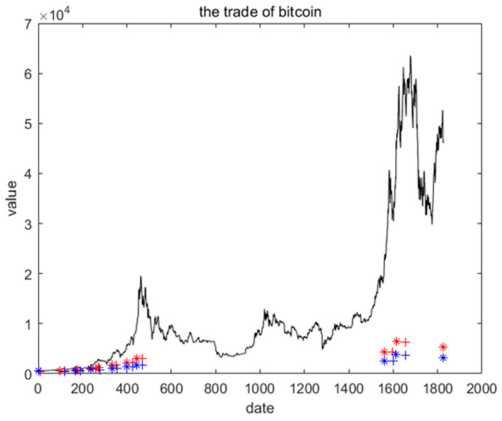

Figure 7. The Trade of Bitcoin

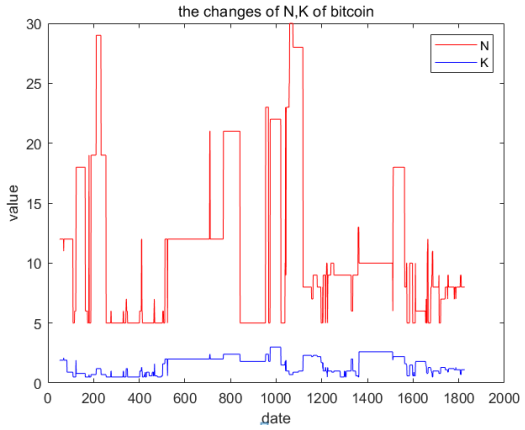

Figure 8. The changes of N, K of Bitcoin

In the case of Bitcoin, there were 36 buys and sells for a total gain of $8,625 based on $1,000.

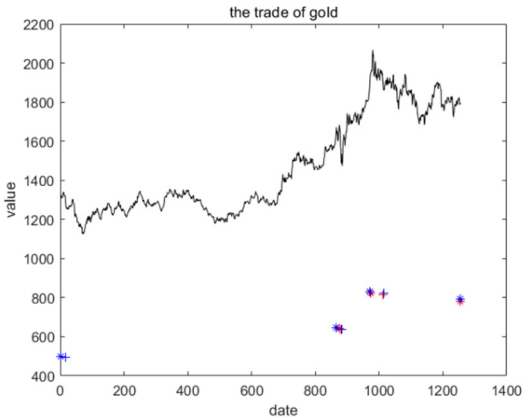

Figure 9. The Trade of Gold

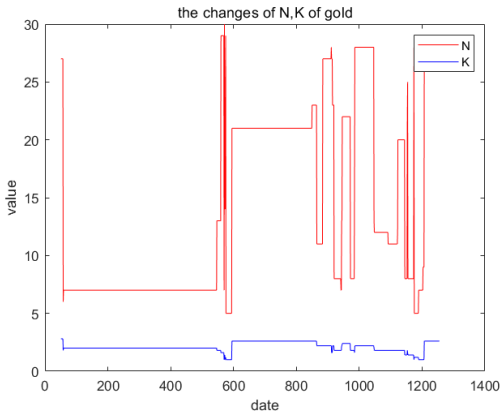

Figure 10. The changes of N, K of Gold

For gold bought and sold for a total of 10 times, based on $1000, the profit was $1573.

ISSN: 2688-9323

2.5. Sensitivity Analysis

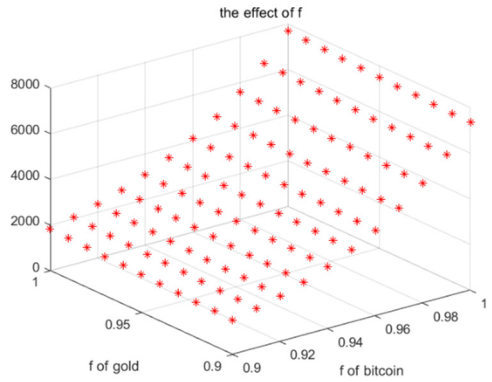

Figure 11. The Effect of F

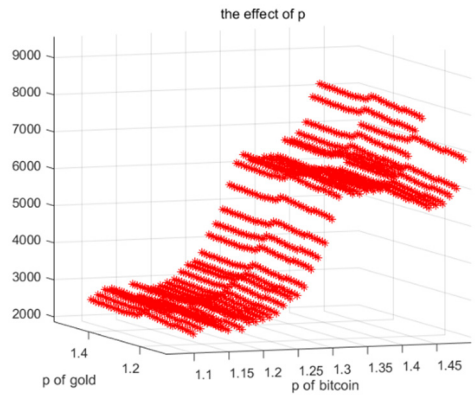

Figure 12. The Effect of P

Based on the sensitivity analysis of the transaction cost of gold and Bitcoin, the transaction conversion rate of gold and Bitcoin is traversed from 0.9 to 1 at an interval of 0.01. It is found that the change of transaction cost of Bitcoin has a great influence on the result, while the transaction cost of gold has relatively little influence on the result. The overall income trend is that the higher the transaction cost, the smaller the income, which is consistent with our perception. The rise of transaction cost will increase the investment risk, and the probability of the model to make investment decisions will be reduced accordingly, and then the income will be reduced.

Based on the sensitivity analysis of the stop-profit rate P, it is found that the larger the stop-profit rate is, the greater the profit is, but there are some aggregation peaks in the local area. Total earnings are greatly affected by the Bitcoin stop-profit ratio, which reaches its first peak when the Bitcoin stop-profit ratio is 1.3, while for gold, it also reaches its peak when the gold stop-profit ratio is = 1.3.

2.6. Results

model has good stability, strong ability to resist risk, total revenue change with basic smooth transaction cost value, no mutations prove model can automatically according to the level of risk frequency when buying and selling is regulated by the check surplus rate of P Models, when P = 1.3 gold and currency earnings are peak, while the risk is relatively small 3 model decision was based on the optimal N, K, the decision is the best solution at the time.

The diversification strategy of gold and bitcoin funds in the investment process can better reduce risks and increase returns. It is shown that the total return of concentrated investment is $4,870, which is less than the $5,099 of diversified investment.

3. Summary

Thorough consideration and analysis of the realities of different types of traders, this paper design the ARIMA model and an improved Bolling-line model based on optimal iteration with a risk function to facilitate the computation of different scenarios for three different types of trader investments and build a dynamic programming model.

Finally, we had a final return of $5,099, an annual return of 38%. And, we conducted a sensitivity analysis on transaction cost and stop-profit ratio, then verified the optimality of the model.

However, when time series models are used to analyze gold price streams, there are still high-order non-white noise sequences. The time series forecasting model belongs to short-term forecasting and can only forecast up to 14 periods. And we will try to solve them in the future.

ISSN: 2688-9323

References

[1] Fang Ni, JU Guopeng, Hui Jiaojiao, Lei Ji. Prediction of municipal solid waste production based on Lasso regression and ARIMA model: A case study of Baoji City [J]. Journal of Henan Science, 202, 40 (01): 98-103.

[2] Wang Z. Forecast and analysis of GDP of Guangdong Province based on ARIMA model. Modern business, 2021 (36): 69-71. The DOI: 10.14097 / j. carol carroll nki. 5392/2021.36.021.

[3] Gu X. Analysis of return volatility in China's stock market -- empirical research based on GARCH model [J]. Investment and entrepreneurship,2020,31(23):5-7.

[4] Wu Guo, Ran Hongliu, Zhou Qing. Fault dislocation probability risk analysis based on monte carlo simulation [J/OL]. Earth science: 1-12 [2022-02-19]. HTTP: // http: // kns. cnki.net/ kcms/ detail/ 42. 1874.P.20220217.1605.021.html.

[5] Li Guowen, LIU Zhiwei. The comprehensive unit price set price based on monte carlo simulation technology risk analysis [J]. Journal of building economy, 2013 (12): 48-51. DOI: 10.14181 / j. carol carroll nki. 1002-851 - x. 2013.12.023.

[6] SONG Jusheng, HUANG Yixuan, ZHENG Wei. Effectiveness of brin line study [J]. Journal of statistics and decision, 2014 (17): 168-171. The DOI: 10.13546 / j. carol carroll nki tjyjc. 2014.17.102.

[7] XIA X P. Quantitative model construction and backtest test based on Bolling line theory [J]. Modern Economic Information,2015(18):283-284.

[8] The Characteristics of bull Market in stock market [J]. Pearl River Economics, 1997(01): 44.

[9] Based on evidence network C2C commodity purchase risk dynamic evaluation model [J]. Control and decision, 2018 (03): 521-528. The DOI: 10.13195/j. k. zyjc. 2017.0020.