Research on price forecasting and trading strategy based on data insight

Kongrui Hong, Xiaoqi Wang, Liming Xu*

School of Computer Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu, 222005 China

*Corresponding author. Email: xuliming0319@163.com

Abstract. The goal of establishing the model in this paper is to find the historical price patterns of gold and bitcoin according to the data provided. The purpose is to maximize returns under various market constraints and avoid loss risk as much as possible. Traders provide the best trading strategy. In this paper, two models are established: model 1: price prediction model based on ARIMA; Model 2: quantitative trading strategy model based on dynamic programming. For Model 1, a classical time series modeling approach based on stock forecasting was used: the ARIMA price forecasting model. The model's validity was demonstrated by analyzing the intrinsic trend of the data movements and verifying the smoothness. Next, historical data were used to fit the parameters of ARIMA, and the forecasting model was determined to be ARIMA (0, 1, 0) by the exhaustive method. Finally, the ARIMA predicts the up and down trends of gold as well as bitcoin, which provides the basis for the trading decision. For Model 2, to better quantify the relationship between investment risk and investment return, the Sharpe ratio is introduced, the Sharpe ratio's value is used as the main parameter of the trading strategy, and the corresponding planning equation is established. Then, based on the up and downtrend of the data predicted by the ARIMA model, the assets are allocated for investment. The model is optimized by a particle swarm algorithm, which accelerates the convergence of the model. Finally, this paper tests the model's accuracy to verify the correctness of the model. By adjusting the commission rate, it is found that the commission rate is negatively correlated with the number of large transactions of gold and bitcoin.

Keywords: ARIMA; Quantitative Trading Strategy; Dynamic Programming; Particle Swarm Algorithm.

1. Introduction

Market participants always make trading strategies based on the previous gold and bitcoin prices. Due to the influence of subjective factors and investment perspective, the traditional manual selection of portfolio is difficult to achieve the optimal allocation. On the other hand, quantitative investment uses computer programs to configure diversified portfolios and process a large amount of data to eliminate non-systematic risks. The latter overcome human weaknesses and reduces the overall investment risk more effectively, scientifically, and reasonably to achieve the best trade strategy. This is the purpose of this paper. This paper needs to meet the following requirements: formulate the model of the best daily trading strategy based on the current historical price data, analyze the sensitivity of trading, and determine three different types of investment, including positive type, robust type and stop-loss type.

2. Model 1:Price Prediction Model Based on ARIMA

2.1 Problem Analysis

According to the condition restrictions, the daily trading strategies are developed only about the price and are not influenced by complex factors such as trading volume, mean, and variance. For scientific and safety reasons, we need to judge the future price trend before making trading Strategies. Therefore, the whole trading strategy is divided into two parts. Firstly, we need to predict gold and bitcoin prices for the next trading day based on historical data, and then determine the amount of investment based on the prediction. In the first part, we use a price prediction model based on ARIMA to train on past prices and predict future price trends. The second part uses a quantitative trading strategy model based on dynamic programming to develop specific investment strategies.

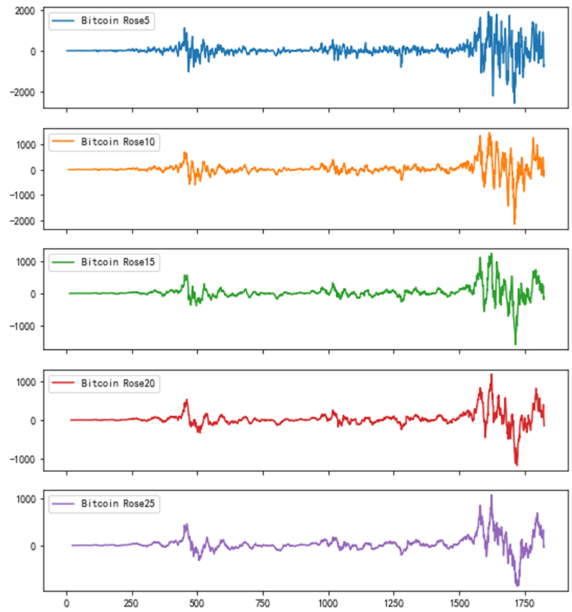

Figure 1. Different Time Mean Increase of Bit

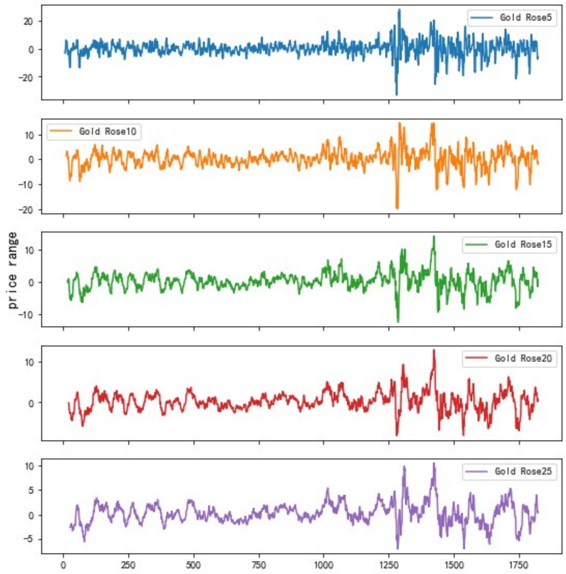

Figure 2. Different Time Mean Increase of Gold coin

2.2 Model Construction

Consider the ARIMA model as a "filter" that seeks to separate the signal from the noise and then extrapolates the signal into the future to obtain a prediction.AR is the autoregressive model, I is the difference model, and MA is the average sliding model in the overall ARIMA model. The AR model establishes a relationship between the eigenvalues of current and historical data, and the following equation defines the p-order autoregressive process:

$$y_i = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \tag{1}$$

where y_i represents the current value; μ represents constant term; p represents order; γ_i represents auto-correlation coefficient, ε_t represents measurement error.

The moving average model focuses on the accumulation of error terms in the autoregressive model, which can effectively eliminate the random fluctuations in the forecast. The formula for the q-order moving average process is defined as follows:

$$y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} \quad (2)$$

ARIMA is a model constructed by converting a non-stationary time series to a stationary one and then regressing the dependent variable on its lagged values alone, as well as on the present and lagged values of the random error term.

Since the ARIMA model requires a smooth time series of the processed data, a smoothness test is considered for the historical gold price series. In order to perform the smoothness test, we consider drawing a gold trend chart based on the gold historical price series and perform observation analysis and statistical test for the gold trend chart.

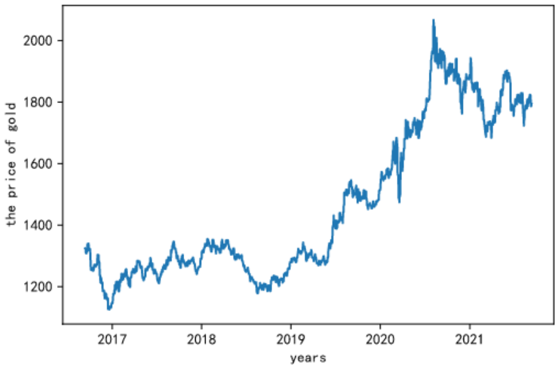

Figure 3. Gold trend chart

For the observation and analysis of the gold price trend chart, the price trend exhibits small fluctuations on the local and overall shows a rising, falling, and rising trend without seasonality and singularity. By further conducting the ADF test on the price series, the p-value of the gold price series was solved to be greater than the significant level 0.05, so the gold price series was judged to be a non-stationary series. Since the gold price series does not meet the smoothness requirement, multiple difference operations are considered for the gold price series until a smooth price series that meets the requirement of the ARIMA model is obtained.

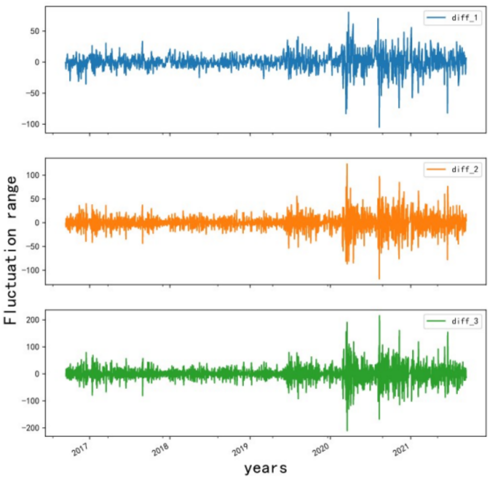

Figure 4. multi-order difference chart

By examining the multi-order difference plot of the gold price series shown above, we can see that the images are uniformly distributed on both sides of the zero value. The trend effect is negligible, and the gold price series meet the smoothness requirement after a single difference. In order to further quantitative analysis of data smoothness, the ADF test is applied to the multi-order difference series of gold price series to quantify the data smoothness of the gold price series.

From the ADF test conducted for the first-order difference series in the above table, the p-value of the first-order difference series is less than the significant level 0.05, and the t-value is less than the statistical value at 1%, 5%, and 10%, which can reject the original hypothesis (the original gold price series is a smooth series) in a highly significant way. It demonstrates that the first-order difference series meets the smoothness requirement and thus is chosen as the treatment series for the ARIMA model.

After the first-order difference of the gold price series, the sequence smoothness requirement of the ARIMA model has been satisfied. Thus, it is necessary further to determine the p and q values. Then the ACF and PACF plots of the first-order difference series must be examined to provide parameters for building the ARIMA model. The ACF and PACF distributions are calculated and plotted by programming, and then it can be concluded that the values of ACF and PACF are both 0, and thus the model is determined as ARIMA(0,1,0).

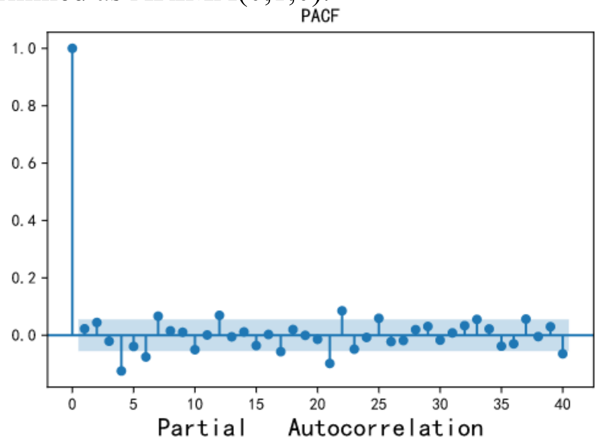

Figure 5. PACF

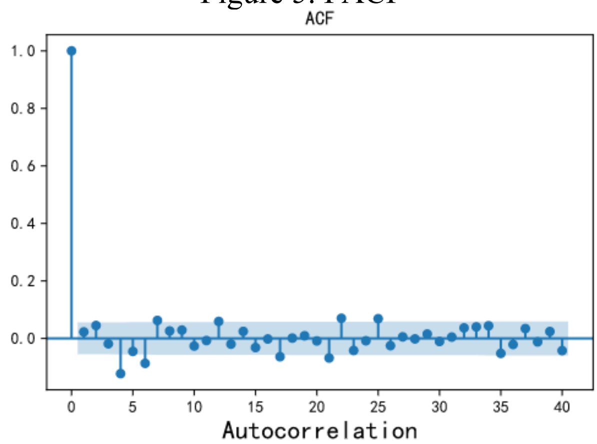

Figure 6.ACF

2.3 Model Solution And Test

By applying the model based on the historical price data, we can forecast the price data for the next three days, adjust the portfolio based on the forecasted price data, and then solve for the optimal investment plan. The price data is predicted according to the prediction model, and the predicted price of gold and bitcoin is plotted against the actual price as follows:

The goodness-of-fit test is analyzed for the price prediction graphs, and the goodness-of-fit of the gold and bitcoin price prediction curves are 0.99 and 0.92, respectively. This indicates that the gold and bitcoin price regression curves are well fitted to the observed values, indicating that the model is valid. Thus, before developing an investment strategy, the model mentioned above can be used to forecast the coming trading day's price trend.

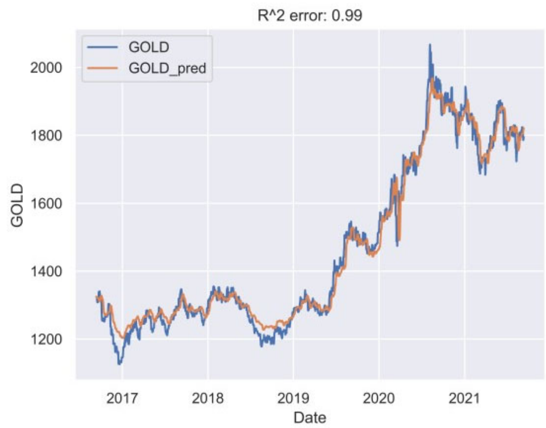

Figure 7.Gold forecast price versus actual price

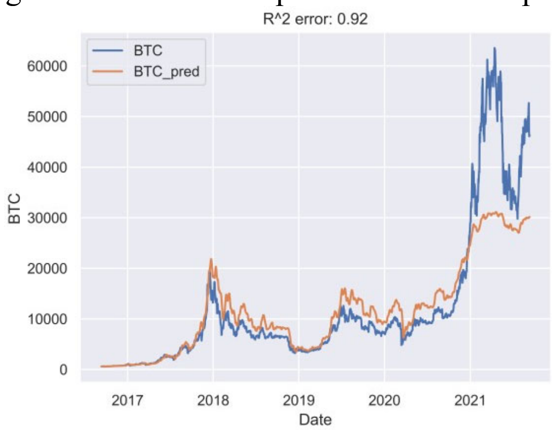

Figure 8. Bitcoin forecast price versus actual price

3. Model 2:Quantitative Trading Strategy Model Based on Dynamic Programming

3.1 Problem Analysis

Based on the analysis, on the current day, we have a complete set of data up to date and require a trading strategy to guarantee profitability after that. Today's data is added to the historical data on the next trading day, and the forecast is made backward again. This dynamic process is a typical dynamic programming process, where we take the results of the previous forecast, specify the adjustment strategy for today, and then recursively until the last day.

3.2 Model Construction

Based on the dynamic programming model, at day k, we set the objective function $f(k)$ to the value of the Sharpe ratio on the third day when the current portfolio is held constant for the next three days, then the maximum value of $f(k)$ is obtained when the risk and return are balanced. Taking day k as an example, from the above model assumptions, the shares of cash, gold and bitcoin in total assets are $[c_k, g_k, b_k]$. Assuming that the increase in cash, gold, and bitcoin on that day are $[0.0662, \bar{G}_k, \bar{B}_k]$, then the percentage will change to $[(1+0.0662\%)c_k, (1+\bar{G}_k)g_k, (1+\bar{B}_k)b_k]$. Due to the total holding amount change, the total assets no longer add up to 100 percent. Therefore, it is required to re-normalize, repeat the procedure for the following trading day, alter the investment strategy, and calculate the transaction cost.

If and only if a transaction for gold purchases is conducted, the amount of change in gold share at this time is $\Delta G_k > 0$. Considering that the gold transaction cost is the total transaction

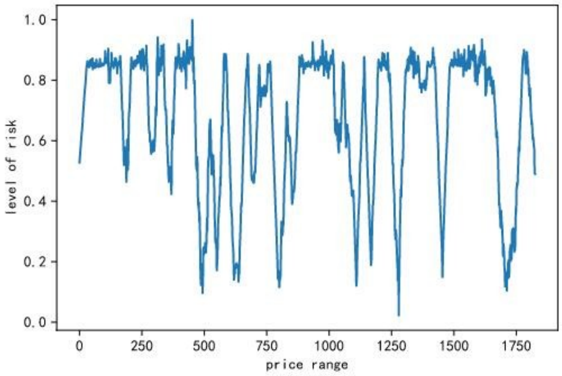

Figure 9. Risk graph of partial bitcoin price

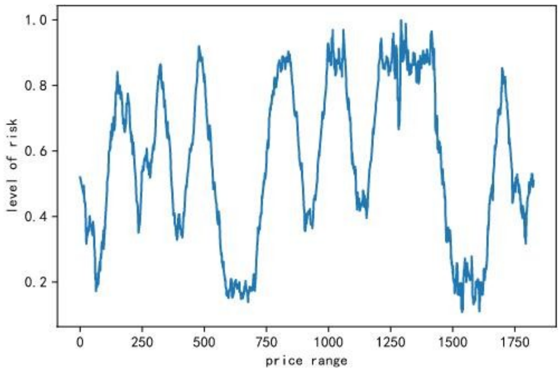

Figure 10. Risk graph of partial gold price range

amount of 1%, the amount of change in cash share at this time is 1.01*ΔGk. Similarly, while purchasing bitcoins, the amount of change in cash share at this time is 1.02*ΔBk

if and only if a transaction for gold sells is conducted, the amount of change in gold share at this time is ΔGk < 0, the amount of change in cash share at this time is 0.99*ΔGk. Similarly, while selling bitcoins, the amount of change in cash share at this time is 0.98*ΔBk.

After the transaction fees are settled, the final share of cash, gold, and bitcoin in the total assets of that day is

$$\begin{bmatrix} (1+0.0662\%)c_k+(1+0.01)\Delta G_k+(1\pm 0.02)\Delta B_k, \\ (1+\bar{G}_k)g_k-\Delta G_k,(1+\bar{B}_k)b_k-\Delta B_k \end{bmatrix} \tag{3}$$

After normalization, the sum of the three percentages is multiplied by the previous day's total holdings to get the total holdings and returns for that day and the percentages for that day. According to the assumptions, the current portfolio is held constant for the next three days. Thus,ΔGk+1 and ΔBk+1 are equal to 0 in the latter process. Up to the third day, the Sharpe ratio of the third day is calculated.

In summary, the model is as follows:

$$\begin{cases} (1+0.0662\%)c_k+(1+0.01)\Delta G_k+(1\pm 0.02)\Delta B_k \geq 0, \\ (1+\bar{G}_k)g_k-\Delta G_k \geq 0, \\ (1+\bar{B}_k)b_k-\Delta B_k \geq 0, \end{cases} \tag{4}$$

3.3 Model Optimization and Solution

the model above is based on the condition that the current investment portfolio would remain fixed for the following three days. Actually, it is the same portfolio that causes problems in daily decisions. In the case of frequent fluctuations, it is likely to sell at the low point and then buy at the high point. As a result of our forecast, we can change our investment strategy for the next three days in order to maximize our profit. Set middle value $[\Delta G_{k+1},\Delta B_{k+1},\Delta G_{k+2},\Delta B_{k+2}]$, the constraints are maintained, and the above model is reused for the solution.

In processing the results, many decision variables make it difficult to determine the range at programming time. The convergence process is lengthy, so the intelligent algorithm is adopted to help speed up convergence speed. We chose to use the particle swarm optimization algorithm, a population-based and self-developing optimization algorithm with the advantages of simplicity and adjustable parameters.

4. Conclusion

The main advantage is the scalability of this article model, where we put all risk factors, price increases, etc., in a single, robust framework.

Firstly, this article has done an excellent job of visualizing both the model-building and solution processes. The thought diagram used during the model-building process, the risk diagram for gold and bitcoin, the price prediction vs. actual price for gold, and bitcoin, and finally, the asset possession diagram all prove this. While the dull data may accurately reflect the outcomes, the images are more intuitive.

Secondly, Using particle swarm optimization algorithm to solve the problem of gradient disappearance and gradient explosion in the training process of long sequences. Maintaining stability based on effective risk avoidance is possible, keeping returns smooth and with a higher safety margin.

Finally, the effectiveness of the model under different commission parameters can be demonstrated by sensitivity analysis. Furthermore, this article optimization model sets up roughly three types of investments for the investment population with different risk tolerance. Thus the model can be used in different investment scenarios.

References

[1] Wijesinghe, G. W. R. I., & Rathnayaka, R. M. K. T. (2020, December). Stock Market Price Forecasting using ARIMA vs ANN; A Case study from CSE. In 2020 2nd International Conference on Advancements in Computing (ICAC) (Vol. 1, pp. 269-274). IEEE.

[2] Worasucheep, C., Nuannimnoi, S., Khamvichit, R., & Attagonwantana, P. (2017, June). An automatic stock trading system using particle swarm optimization. In 2017 14th interna- tional conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON) (pp. 497-500). IEEE.

[3] Crone, N., Brophy, E., & Ward, T. (2021). Exploration of Algorithmic Trading Strategies for the Bitcoin Market. arXiv preprint arXiv:2110.14936.

[4] Ye, A., Chinthalapati, V. R., Serguieva, A., & Tsang, E. (2017, December). Developing sustainable trading strategies using directional changes with high frequency data. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4265-4271). IEEE.

[5] Nenortaite, J., & Simutis, R. (2004, June). Stocks' trading system based on the particle swarm optimization algorithm. In International Conference on Computational Science (pp. 843-850). Springer, Berlin, Heidelberg.

[6] Liu, M., Liu, X., Jia, W., & Du, Y. (2020, July). The trading strategy of inflection point futures analysis based on AFS theory. In 2020 39th Chinese Control Conference (CCC) (pp. 2170-2175). IEEE.

[7] Bali, T. G., & Peng, L. (2006). Is there a risk–return trade-off? Evidence from high-frequency data. Journal of Applied Econometrics, 21(8), 1169-1198