Research on the application of time series ARIMA model in trade strategy

Shuohong Ye[1,*,#], Mingyu Zhang[2,#], Jiabeizi Yu[2,#]

[1] School of mathematics and statistics, Chongqingjiaotong University, Chongqing, China

[2] School of economics and management, Chongqingjiaotong University, Chongqing, China

*Corresponding author: 2505510206@qq.com

[#] These authors contributed equally.

Abstract. Market trading has always been a popular choice for bold investors, but market trading doesn't rely on pure luck. More need for daily experience and data. In this article, we will model the random forest algorithm for a specific situation and find the best strategy. Then it is proved that the model can provide the best trading strategy, which can be understood as proving the fit of the model established above. Using the real value of bit coin and gold prices in 5 years, establish a time series ARIMA model, analyze the smoothness and pure randomness of the series.Then get the ARIMA model fitting of bit coin and gold, indicating that the prediction model is very suitable. For the dynamic programming model, this paper adopts a genetic algorithm for the judgment of the optimal buy-sell node, which accelerates the convergence speed. To understand how the trading price affects the strategy and the results, sensitivity analysis is also done in this paper. Considering giving a certain perturbation to the results and giving a perturbation margin of 5% to the buying and selling fees respectively, the algorithm is iterated to calculate the sensitivity of the investment strategy to the trading cost and analyze how the change in the trading cost would affect the buying and selling strategy.

Keywords: ARIMA; Random Forest; Dynamic Planning Model; Genetic Algorithm.

1. Introduction

Market traders often choose to buy and sell shares to maximize their earnings. For each purchase and sale, there is also a return commission [1-3]. Suppose that as of September 11, 2016, the trader has an account in which the initial states of cash, gold, and bit coin are [1000,0,0]. During the five-year trade period, the cost of each trade (purchase or sale) is a% of the trade amount, where a% for gold and bit coin is 1% and 2%, respectively, and the stock held costs nothing. Considering that the topic requires a given daily optimal trade strategy, we choose a time series model with a random forest algorithm to predict the next day’s market situation [4-6]. A planning model is then built and the model is iterated using a genetic algorithm to finally output the best node for trading. To prove the feasibility of the strategy we will test it and perform sensitivity analysis. The mind map of the full text is shown in Figure 1.

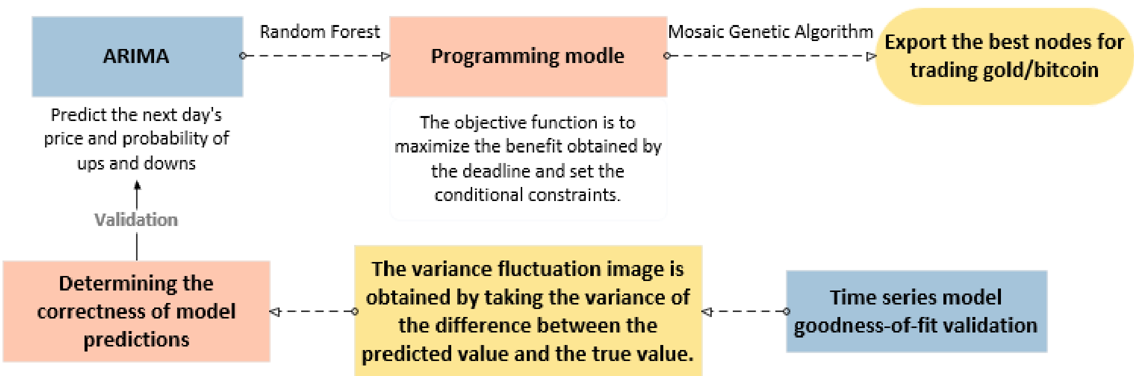

Figure 1. Mind Map

2. Assumption and Justification

1. It is assumed that no additional investment is made during the trading period.
2. Assume that the trade is made at the close of the stock market on the nth day, i.e., the price does not change on that day.
3. Assume that transaction costs are settled at the same day's price and generate revenue from the next day onwards.
4. The uncertainty of the return is measured by the variance of the return.
5. Assume that at the start of the starting time, the entire principal is chosen to be purchased in gold or bit coin.

3. Symbols and assumptions are shown in Table 1

Table 1. Symbols and assumptions

Symbol	Description
N	Sample size
t_i	Time of the i-th transaction
Q	Number of transactions at t_i
P	Sold Price
Z	Holdings
C	Cash volume
n_f	Sample size for update time

*The subscript t_i is the parameter for bit coin and the subscript t_k is the parameter for gold.

4. Optimal Trade Strategy

Market trading is often incredibly dangerous; it may both bring you a fortune and cost you everything you own [7]. To avoid this, we must forecast the market to maximize our profits. In this post, we utilize the example of combining two markets, gold, and bit coin, to determine the best trade strategy for investors.

4.1 Model Preparation

To decide the optimal trade strategy, we first introduce the time series model and the random forest algorithm and introduce them [8].

Random forest is a decision tree-based integrated method that uses numerous decision trees to avoid over fitting [9]. It also includes properties like being simple to understand, handling class features, being simple to extend to multi-classification situations, and requiring no feature scaling. The random forest trains a series of decision trees independently, resulting in a parallel training process. Because the algorithm includes a random process, each decision tree is slightly different. To reduce the number of prediction scenarios, combine the prediction results of each tree to improve performance on the test set.

The principle of time series predictive analysis is to use the characteristics of the time of an event in the past period to predict the characteristics of the event in the future period, that is, to predict the unknown event from existing data [10]. Unlike regression analysis models, this is a relatively complex class of predictive modeling problems. The time series model relies on the sequence of events, and the results generated by the same values will change when the sequence is changed. The first point that needs to be clarified is that time series can be divided into smooth series, i.e., those in which there is some period, seasonality, and trend in the variance; and non-smooth series, i.e., those in which the mean does not change over time.

In this paper, we first use a time series model to forecast the data of the first two days to forecast the data of the next day to decide whether to buy or sell on that day.

The model is shown in the following equation:

$$(1-\sum_{i=1}^{p}\emptyset_1 L^i)(1-l)^d X_t = (1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t \tag{1}$$

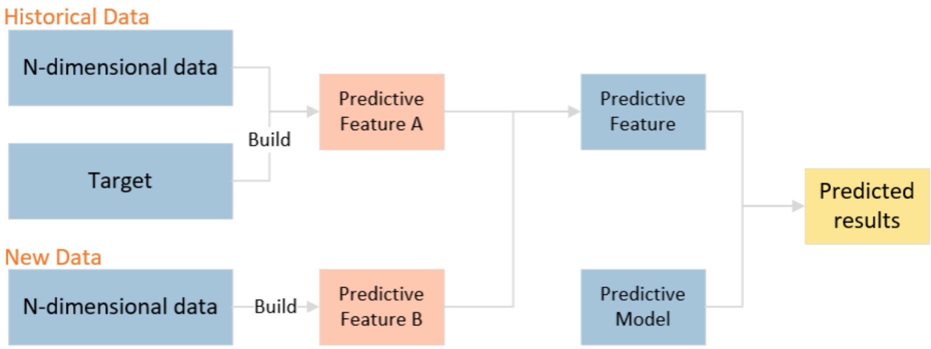

Figure 2. Machine learning algorithm schematic

4.2 Model Building

First, the objective function is given to buy gold or bit coin from November 9, 2016, until October 9, 2021, based on the price data up to that date, giving the best strategy for each day's trade to make the highest total return, according to the conditions of the question. Let the total assets of the day i be Wi (total assets = amount of cash + bitcoin holdings + amount of gold holdings × today's price - buying and selling fees), then the relationship can be listed according to the above equation:

$$W_i = C_k + Z_i \cdot P_{t_i} + Z_{t_k} \cdot P_{t_i} - |Q_{t_i} \times 1\%| - |Q_{t_k} \times 2\%| \tag{2}$$

Assuming that day i is traded, the price trend of day i+1 is first predicted by the data of day i-1, i. With the price increase or decrease of day i+1 as the constraint, firstly, the profit from the transaction needs to be greater than the commission; the trading volume must not be greater than the holding volume; secondly, the prescribed function is satisfied.

Specify the function F(X):

$$\text{F(X)} = \begin{cases} 1 & \text{When the trade is a buy} \\ 0 & \text{When not buying or selling} \\ -1 & \text{When selling} \end{cases} \tag{3}$$

The constraints are obtained by collating the above requirements and stock market trading laws:

$$C_k \geq F(X)Q_{t_i} \cdot P_{t_i} + F(X)Q_{t_k} \cdot P_{t_k} \tag{4}$$

The final dynamic planning model can be obtained as *max*Wi:

$$\begin{cases} W_i = C_k + Z_i \cdot P_{t_i} + Z_{t_k} \cdot P_{t_i} - |Q_{t_i} \times 1\%| - |Q_{t_k} \times 2\%| \\ \frac{P_{t_{i+1}}}{P_{t_i}} - 1 > 0.02 \\ Z_i \geq Q_{t_i} \\ \frac{P_{t_k}}{P_{t_{k+1}}} - 1 > 0.01 \\ Z_{t_k} \geq Q_{t_k} \\ C_k \geq F(X)Q_{t_i} \end{cases} \tag{5}$$

Genetic algorithms are essentially a series of operations performed on chromosomal patterns, i.e., inheritance of good patterns from the current population to the next generation population by selection operators, pattern recombination by crossover operators, and pattern mutation by variation operators. Through these genetic operations, the pattern gradually evolves in a better direction and finally, the optimal solution of the problem is obtained.

Genetic algorithm to achieve global convergence, easy to parallelize the processing, the use of heuristic search and the adapted function is not subject to the constraints of continuous, differentiable and other conditions, a wide range of applications.

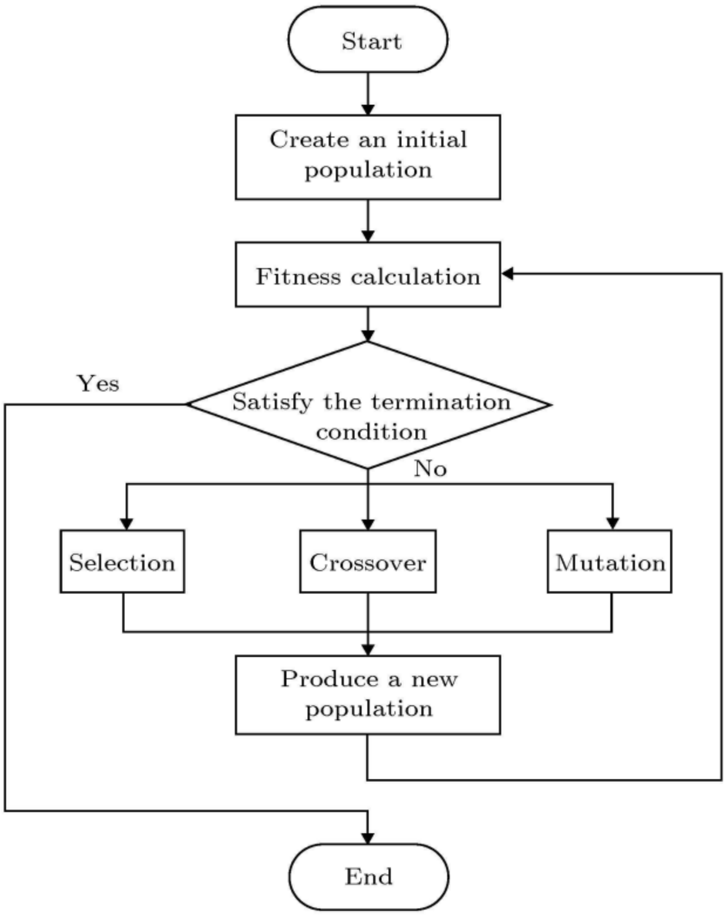

Figure 3. Genetic algorithm flow chart

4.3 Model solving

After the model is built, the existing data is integrated and the algorithm is used to find the daily returns of gold and bit coin respectively.

As the chart shows, yellow represents the daily return of gold and pink represents the daily return of bit coin, which visually reflects that the daily return of bit coin is significantly larger than the daily return of gold. It also reflects the fact that gold’s price is more stable while bit coin’s price is more volatile.

A random forest algorithm is used to predict the next day’s market price and the probability of increase or decrease, and the predicted values of gold and bit coin are compared with the real values respectively.

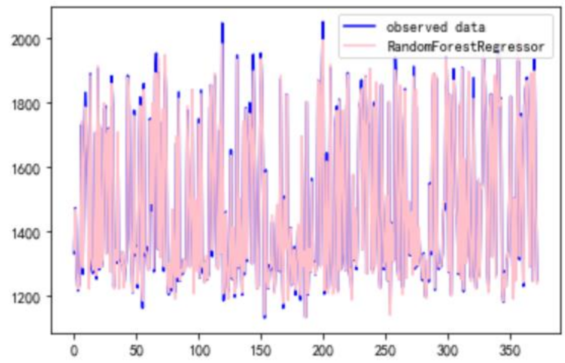

Figure 4. Comparison of predicted and real values of gold.

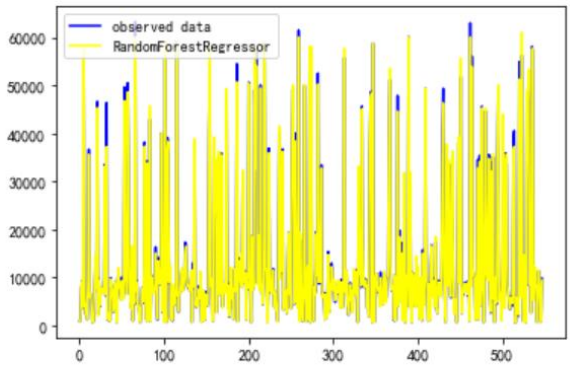

Figure 5. Bit coin predicted value compared to the true value

Using MATLAB to perform calculations, the optimal strategy for each day's trade, the daily trade volume, gold holdings, bit coin holdings, and the amount of gold and bit coin traded can be derived. Due to the large size of the data, the output is detailed in the appendix. Then based on the obtained result data, the trend of cash holdings share, total assets, bit coin and gold holdings share are plotted. model fits well, so you can refer to our model with confidence.

5. Test of Optimal Trade Strategy

In the above article, we used a time series model based on machine learning to predict the next day's market price. To validate the reasonableness of this model, we tested the prediction results for gold and bit coin respectively using all available data, and the steps to validate the time series prediction model are as follows.

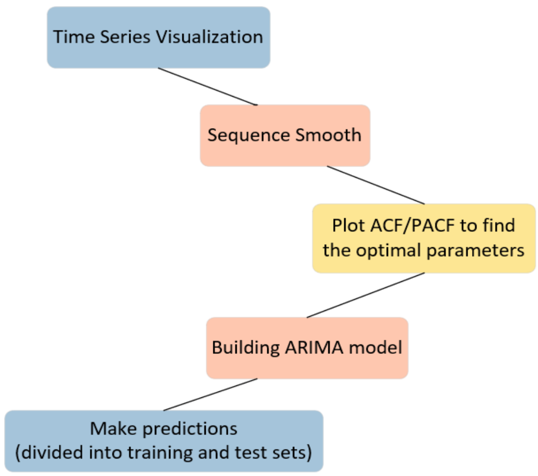

Figure 6. Time series prediction proof ideas

According to the time series plot, it can be seen that there is no obvious trend or shows periodic characteristics. Firstly, the bit coin series plot is made differential, and after the ADF test, the pvalue is less than 0.05, which means that the series is smooth; after that, the pure randomness test is done, and according to the results, the series is a non-white noise series, thus judging that this time series is a smooth non-white noise series. After that, autocorrelation and partial autocorrelation plots were drawn.

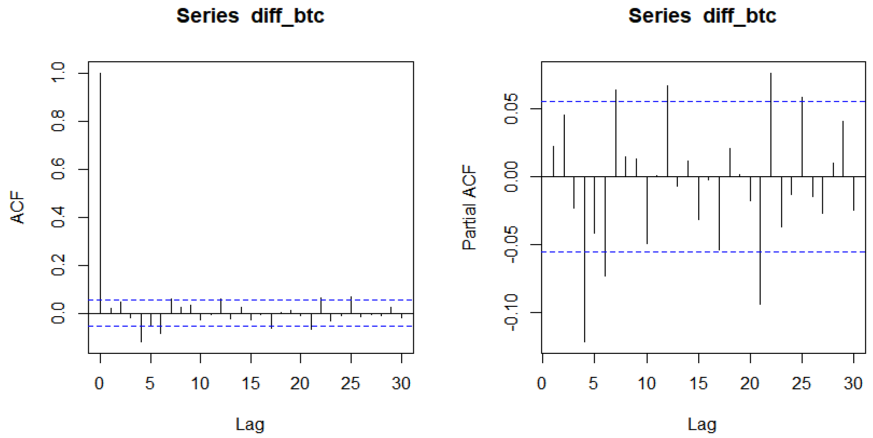

Figure 7. Bit coin's autocorrelation and partial autocorrelation graphs

From the figure 7, the autocorrelation (ACF) plot is in the second-order truncated tail and the partial auto correlation (PACF) plot is in the trailing tail, based on the AIC information criterion to find the optimal parameters, the model results for the ARIMA model (0,1,2) test table and based on 1 difference data, the model equation is as follows:

$$y(t) = 25.067 - 0.076\varepsilon_{t-1} = 0.068\varepsilon_{t-2} \tag{6}$$

Similarly, the series of gold can be differenced using the above method, after which a pure randomness test for smoothness is performed to obtain a smooth non-white noise series, after which the model is obtained as ARIMA model (0,1,0) by fixing the order with the above method. And the model's goodness-of-fit $R^2 = 0.997$ can be derived, indicating that the model is a good fit. In summary, our time series model fits well and validates its usability, so we can use the model to forecast the market.

6. Sensitivity Analysis

To prove that the scheme is the best solution we provide to investors, we consider applying a range of disturbances by giving results, through the disturbance and handling of investment programs, further determine data to predict the uncertainty of disturbance for the specific influence the output, and then we can analyze the disturbance after processing data, To improve the reliability and credibility of the investment scheme we output, the actual operation process is to repeat the above investment judgment and trading process within the given disturbance range, add 5% disturbance range to the trading scheme of bit coin and gold, and output the following results:

To determine the sensitivity of our specified trading strategy to transaction costs, we explored the extent to which the parameters in the genetic algorithm related to transaction costs: 'the percentage of transaction fees for gold' and 'the percentage of fees for bit coin transactions' affect the final trading strategy as well as the outcomes, and we calculated the sensitivity of this investment strategy to transaction costs by running iterations of the algorithm varying the parameters to obtain the following results. where "a gold", "a bit" represent the proportion of gold and bit coin transaction fees respectively, returns represent the final revenue of the overall scheme under different fee ratios; "g tans times" represents the number of times gold changes in the whole strategy; "bit trans times" represents the number of times bit coin changes in the whole strategy. According to the results, with the change of transaction fees, the overall return changes more obviously, with the increase of transaction fees, the number of large transactions of gold and bit coin decreases significantly, with the decrease of transaction fees, the number of large transactions of gold and bit coin has a significant increase.

7. Strengths and Weaknesses

Strengths: Separate models for gold (non-equal interval day trading) and bit coin (price time-continuous trading) are built to synchronize the non-synchronous trading and make the dynamic planning model more concise and rigorous.

Weaknesses: Since only today's prices of gold and bit coin are available, not including data such as opening price, closing price, and trading volume, the risk model can only use simple variance fluctuations to determine the uncertainty of returns. In practice, the risk model can be improved with more characteristic data, such as the value-at-risk model VaR.

8. Model Improvement

The model produced in the previous part is a planning model that divides the price persistence model and the non-equal interval trading model by time, respectively, and it is recommended below to improve the model to make non-synchronized trading synchronized.

While the relationship between various assets in both time series and cross-section is crucial, the research of cross-sectional correlations can be hampered by multiple time series of nonsynchronous transactions. To transmit price events to new prices in a non-equally spaced, nonsynchronous form, the first step is to synchronize the non-synchronous time series.

All of the above models only consider the basic constraints at the time of trading but do not take into account the risk issue. Here we introduce a risk model with $\frac{Predictedvalue-Actualvalue}{Actualvalue}$as the risk assessment criterion and calculate its variance, which is used to judge the uncertainty of the return. Given the sequence of returns B will be substituted into the model as a constraint, making the model more realistic and enhancing the model credibility.

References

[1] Barndorff N.O.E., Hansen P.R., Lunde A., et al. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise andnon-synchronous trading [J]. Journal of Econometrics, 2010,162(2): 149-169.

[2] Lu, Wanbo, Chen, gallop, Wang, Jianye. A study on the intrinsic risk value of asset portfolios with non-equal interval days [J]. Mathematical Statistics and Management, 2019, 38(06): 1104-1118.

[3] Ye W.Y., Sun L.P., Miao B.Q.. A dynamic cointegration study of gold and bitcoin - based on a semiparametric MIDAS quantile point regression model [J]. Systems Science and Mathematics, 2020, 40(07): 1270-1285.

[4] Zhang, H.P., Zhu, J.M.. Optimal design of multi-stock portfolio investment strategy based on risk diversification [J]. Journal of Henan University of Science and Technology (Natural Science Edition), 2020, 48(04): 62-70.

[5] Ye Wuyi, Sun Liping, Miao Boqi. A dynamic cointegration study of gold and bitcoin - based on a semi-parametric MIDAS quantile point regression model [J]. Systems Science and Mathematics, 2020, 40(07): 1270-1285.

[6] Wu Hanzhang Research on asset pricing of investors' disappointment aversion in China' stock market [D]. Shanxi University of Finance and economics, 2021.

[7] Jing Pengfei Comparative study on bitcoin price prediction based on multi model [D] Shanxi University, 2021.

[8] Suresh K K, Priya S R K. Forecasting Sugarcane Yield of Tamilnadu Using ARIMA Models [J]. Sugar Tech: An international journal of sugar crops and related industries, 2011.

[9] Kuhe D A, Obed T A. An ARMA Model for Short-term Prediction of Hepatitis B Virus Seropositivity among Blood Donors in Lafia-Nigeria [J]. Journal of Scientific Research and Reports, 2019:1-11.

[10] Xu Liping, Luo Mingzhi. Short-term analysis and forecast of gold price based on ARIMA model [J]. Finance and Economics, 2011(1): 9.