Application of ARIMA algorithm and genetic algorithm based on MATLAB language in finance

Sihao Wu

Faculty of Business and Management, Beijing Normal University - Hong Kong Baptist University United International College, Zhuhai, 519000, China

13835509755@163.com

Abstract. The price changes of bitcoin and gold are frequent, as observed in previous studies, and market participants can base their daily investment plans on historical data. Market traders can maximise their investment returns by buying low and selling high on volatile assets based on the day's market conditions, as markets fluctuate every day. Gold and bitcoin are two of the most common volatile investments. In this project, we are trying create an algorithm of high degree of integration which can provide advice to traders before they made their decisions. We used ARIMA algorithm based on the Genetic Algorithm to predict the data and let the algorithm to make based on the predicted data. The complete code could be seen in appendix. The codes were mostly based on MATLAB language, for which it integrated the features of "simple and easy to get started"; "dispatchable sensitivity"; "adaptable and reliable". Which allows the traders to use simple operations to get reliable and visualizable results. Which can let the traders to have the opportunity to get suggestions on when they should trade.

Keywords: ARIMA Model, Profit-Position, Deficit-Position, Sensitivity, MATLAB

1. Introduction

The price changes of Bitcoin and gold are frequent, as observed in earlier studies [1-2], and market participants can base their day-to-day investing plans on historical data. We can assist market traders in maximizing the returns on their investments by utilizing this data.

Market traders can maximize their investment returns by purchasing low and selling high on volatile assets based on the day's market conditions because markets fluctuate every day. Gold and bitcoin are two of the most frequent volatile investments.

With the data given in the title, we already have the two data sets of date and daily price of gold and bitcoin. We need to use these two data to build a model to guide market traders to buy, hold, and sell to maximize their investment over a five-year trading period from November 9, 2016 to October 9, 2021 with a starting amount of $1,000.

And we need to take into account that gold and bitcoin transactions have a cost, gold is 1% and bitcoin is 2%.

Considering the context and constraints identified in the problem statement, we propose the following solution.

2. Assumptions and notations

2.1. Assumptions

To simplify the given problems and modify it more appropriate for simulating real life conditions, we make the following basic hypotheses.

1. Assume that the daily percentage change in multiple market variables follows a normal distribution.

2. The daily change in the value of the trading portfolio also follows a normal distribution.

3. Assuming no severe fluctuations, such as stock meltdowns, occur during the set period.

4. linear assumption, the risk exposure of the portfolio is linearly correlated with the risk factor during the holding period.

2.2. Notations

Table 1. Notations

Symbols	Definition
ACF_P	That is value p, represents the lag number of the time series data itself used in the prediction model [3], also called AR/Auto-Regressive item
PACF_Q	That is value q, the number of lags representing the forecast error used in the forecast model, also known as the MA/Moving Average term
Periodicity	The prediction period in the ARIMA model. The function is to treat the first several times of data as independent variables, and thus to condition the prediction on the data of that time.
n	The number of values to be predicted after each prediction period
ProfitPos	Which stands for Profit-Position, which is the point when profit and loss is judged. This value is compared with DeficitPos when making profit/loss judgments. If the ProfitPos value is greater than the DeficitPos value, then it can be judged as a profit state; if the ProfitPos value is less than the DeficitPos value, then it can be judged as a loss state; if both ProfitPos and DeficitPos are equal to 7, then there is no loss and no profit.
DeficitPos	Which stands for Deficit-Position. When making a profit/loss judgment, the value is compared with ProfitPos. If the DeficitPos value is less than the ProfitPos value, it is judged to be a profit state; if the DefitPos value is greater than the ProfitPos value, it is judged to be a loss state; if both ProfitPos and DeficitPos are equal to 7, there is no loss and no profit.
Rp	which was rate of profit. Rp=p(t)/p (t- 1). For which Rp stands for the rate of profit at time t.
Rd	which was rate of deficit. Rd=p(t)/p (t- 1). For which Rd stands for the rate of profit at time t.
LMData	(LBMAtrade rate of profit of Gold. Seq, k+2)
BCData(i, k+2)	rate of profit of Bitcoin

3. Method

3.1. Have basic forecasting decision ideas

Predictive analysis of data using ARIMA algorithm [4] based on genetic algorithm principles.

When making decisions, it is to look at the forecast within 'n' days based on the 'Periodicity'. If there is a more suitable operation to make money or avoid losses, trade, otherwise, hold.

3.2. Build a basic forecasting model framework

ARIMA prediction model can be expressed as

Predicted value of y=weighted sum of the constants c and/or Y at one or more recent times and/or the prediction error at one or more recent times

Suppose p,d,q is known,AMIMA is expressed in mathematical form as:

Predicted value of $y_t = \mu + \emptyset_1 * y_{t-1} + ... + \emptyset_p * y_{t-p} + \theta_1 * e_{t-1} + ... + \theta_q * e_{t-q}$

Obtain time series data of the observed system;

Plot the data and observe whether it is a stationary time series; for non-stationary time series, d-order difference operation is performed to convert it into a stationary time series;(Stable data has no trend and no seasonality; that is, its mean has a constant amplitude on the time axis, and its variance tends to the same stable value on the time axis. Hypothesis testing can be performed using the Dickey-Fuller Test)

After the second step, a stationary time series has been obtained. To obtain the autocorrelation coefficient ACF and the partial autocorrelation coefficient PACF of the stationary time series, the optimal rank p and order q can be obtained by analyzing the autocorrelation graph and the partial autocorrelation graph.

From the d, q, and p obtained above, an ARIMA model is obtained. Then start model checking on the resulting model.

Specify the structure of the model and Variables' Values.

First, use SPSS to differentiate the data and output the ACF, PACF images.
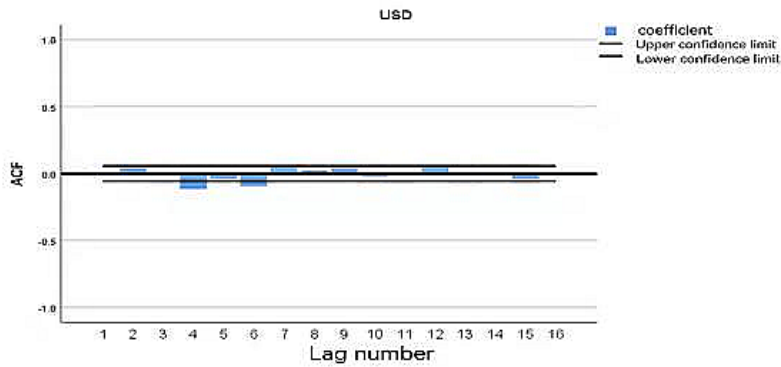
1) The ACF, PACF images of Gold.

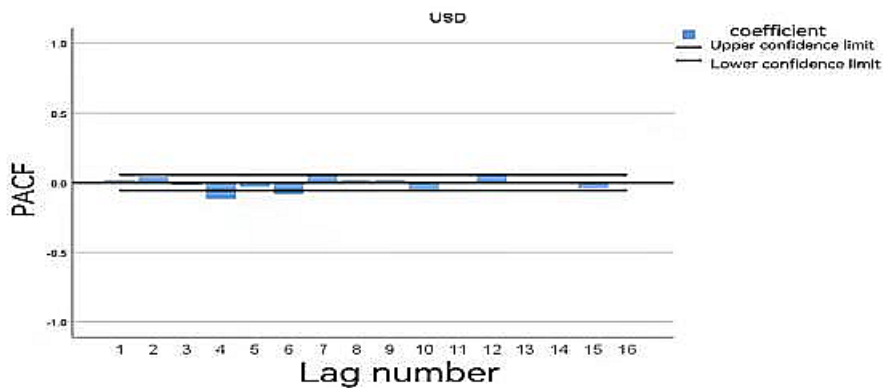**Figure 2.** The ACF images of Gold

Figure 3. The PACF images of Gold

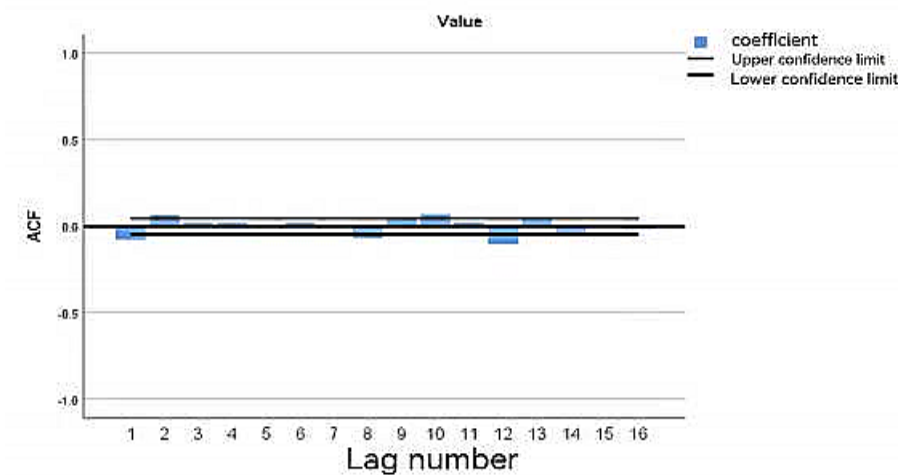2) The ACF, PACF images of Bitcoin.

Figure 4. The ACF images of Bitcoin
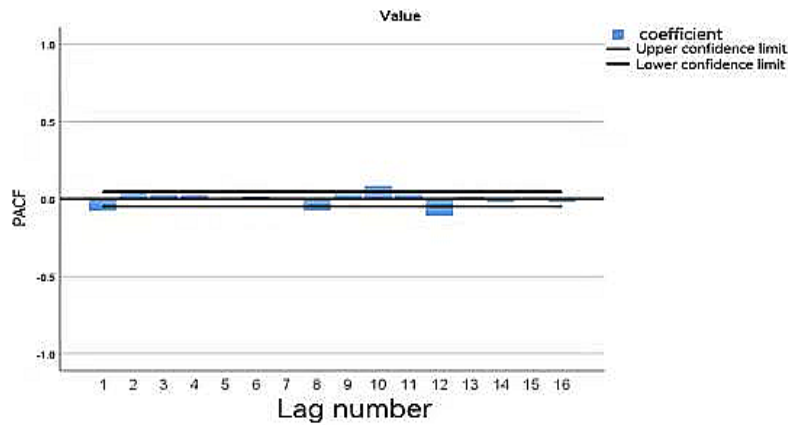

Figure 5. The PACF images of Bitcoin

The code of auto selecting model structure of Gold.

```
TSMODEL
        /MODELSUMMARY  PRINT=[MODELFIT]
        /MOEDLSTATISTICS  DISPLAY=YES MODELFIT=[SRSOUARE RSOUARE]
        /MODELDETAILS  PRINT=[PARAMETERS FORECASTS]
PLOT=[RESIDACF RESIDPACF]
    /SERIESPLOT  OBSERVED  FORECAST  FIT  FORECASTCI  FITCI
        /OUTPUTFILTER  DISPLAY=ALLMODELS
        /SAVE  PREDICTED  LCL(LCL)  UCL(UCL)  NRESIDUAL(NResidual)
        /AUXILIARY  CILEVEL=95  MAXACFLAGS=24
        /MISSING  USERMISSING=EXCLUDE
        /MODEL  DEPENDENT=Value  INDEPENDENT=DATE
               OUTFILE='E:\MCM\spss\BCHAIN\(2).xml'
                PREFIX='model'
          /ARIMA  AR=[1]  DIFF=2  MA=[1]
               TRANSFORM=NONE  CONSTANT=YES
          /AUTOOUTLIER  DETECT=OFF
```

From the output code of SPSS for BITCOIN, it's not hard to find that the best model for the bitcoin is ARIMA [1,2,1].

Same method as bitcoin, we know the best model for the gold is also ARIMA [1,2,1].

4. Trend analysis based on ARIMA model

4.1. Calculate the predictive value of Flows

```
function [result] = ARIMA_algorithm(data, Periodicity, ACF_P, PACF_Q, n)
```

Since we had already Specify the structure of the ARIMA algorithm, the value of ACF_P and PACF_Q could be seen as a constant. Then we only need to define the value of 'periodicity' and 'n'. Which will affect sensitivity the strategy is to transaction costs.

For simplicity, we begin present a special situation. We make some other specific assumptions. 'periodicity=20'; 'n=6'; 'ACF_P=1'; 'PACF_Q=1'

4.2. Calculation based on the algorithm using MATLAB

```
result = data(m1+1:end);
plot(result);
```

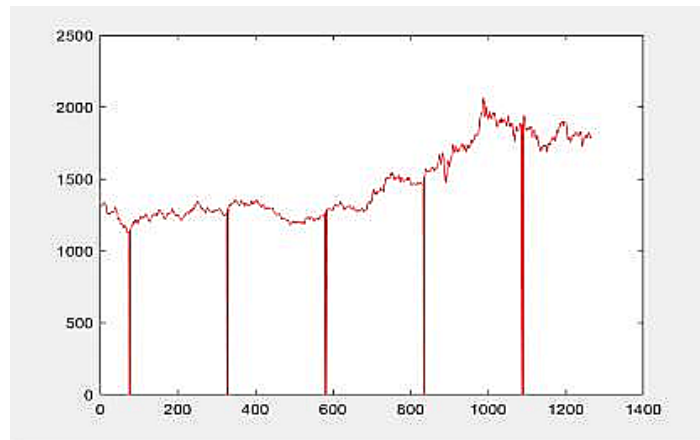Then plot the result calculated by using MATLAB [5], we have the Figure 6.

Figure 6. The prediction curve of Gold

From Figure 6, we notice that there are some data missing. We define that the day while missing data are unable to trade.
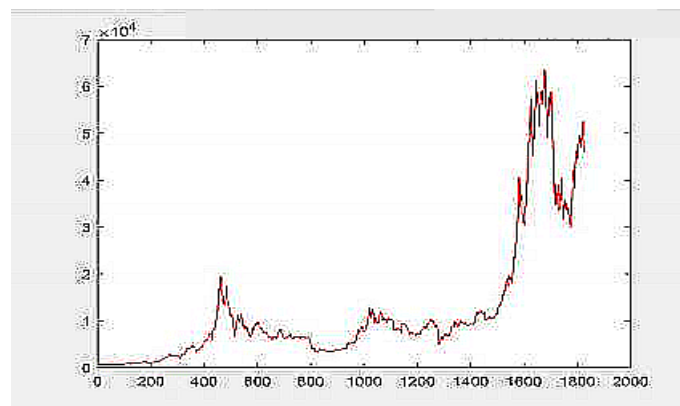

Figure 7. The prediction curve of Bitcoin
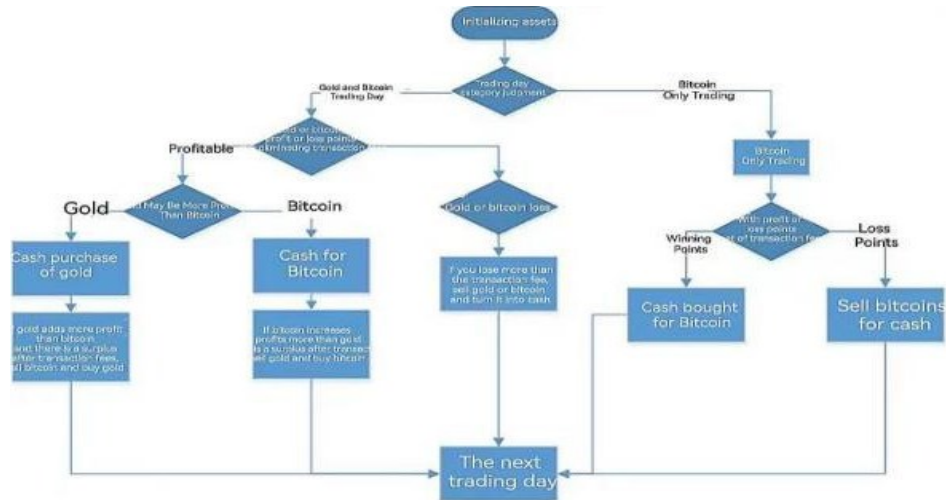
5. Decision stage

Figure 8. Visio diagram

Through the basic conditions given by the material, we know initialized cash, the holding status of gold and Bitcoin are [1000, 0, 0] respectively.

Therefore, we can get the holding status before the daily transaction, and then get the asset holding status after the daily transaction after the transaction.

On the days when gold is allowed to trade, gold transactions need to be processed, and both gold and bitcoin can be traded at this time.

For the decision of dollar cash not gold or bitcoin, the decision is made by interpreting whether gold is more profitable than bitcoin. In the decision-making process, for the confirmation of profit and loss points, if there is neither a profit point nora loss point, then Profit-Position and Deficit-Position are both 7.

If there is a profit point, and the profit point is before the loss point, a buy operation is performed. If gold is more profitable than bitcoin, buy gold and turn all your cash into gold. When judging whether it is necessary to convert Bitcoin to gold, if there
is neither a profit point nor a loss point, then Profit-Position and Deficit-Position are also both 7.

If bitcoin is more profitable than gold, buy bitcoin and convert all cash to bitcoin. There is a loss point, and the loss point is before the profit point, sell.

When gold loses, sell all gold held; similarly, when bitcoin loses, sell all bitcoin.

After the above operations are completed, the next gold allowable trading day is processed. Handles the relationship between Bitcoin and Cash when only Bitcoin is tradable. Similarly, first confirm the profit and loss point.

If there is neither a profit point nor a loss point, then Profit-Position and Deficit-Position are both 7. If Profit-Position < Deficit-Position, then cash is exchanged for Bitcoin; if Profit-Position > Deficit-Position, then Bitcoin should be exchanged for cash

6. Decision making result
For simplicity, we begin by presenting a special situation. We make some other specific assumptions.

```
If LMData (LBMAtradeSeq, k+2) >2.0 ||BCData (i, k+2) > 2.0,
There is a profit point, and the profit point is before the loss
point; buy operation
There is a deficit point, and the deficit point is before the profit
point, sell.
If LMData(LBMAtradeSeq,k+2)<0.2, sell gold
```

```
BCData(i, k+2)<0.2, sell bitcoin
ValueEnd =
CashHold(i)+BCHold(i)*BCDataPredict(i,2) *0.98+LMData(end,2)
*LBMAHol d(end)*0.99;
```

For this situation, we have the final result 3342.14033250550 USD.

Among the factors influencing the sensitivity, we can analyze them in two different parts, the prediction part and the decision part. For the prediction part, we can refer to two values, Periodicity and n. When the ratio of Periodicity to n is larger, the higher the volatility of the stock (security), resulting in lower sensitivity. Conversely, when the ratio of Periodicity to n is smaller, the volatility of the stock (security) is smaller, resulting in higher sensitivity. And for the decision-making part, we can refer to Profit rate and Deficit rate. the larger the Profit rate, the lower the sensitivity, and conversely, the smaller the Profit rate, the higher the sensitivity. For Deficit rate, the larger the Deficit rate, the higher the sensitivity, and vice versa, the smaller the Deficit rate, the lower the sensitivity.

Given the refinement and rationality of our model, our model has unparalleled advantages, starting with our forecasting model, as follows.

1.Most of our models are in MATLAB language, so they are more integrated and easier to use for our customers.

2.Our model prediction principle has a strong basis, so the predicted data and situation are more scientifically reliable.

3. The sensitivity of our models is adjustable, so we can adjust the sensitivity of our models to meet the needs of our clients for different types of investments.

4.4. Our model is very simple and requires only endogenous variables without resorting to other exogenous variables. It is easy for clients to understand and operate.

Although we have tried our best. Time is not quite enough, and some data are missed. As a result, the missing data can still bring the errors in evaluation. Then there are the following deficiencies for our prediction model.

1. The time-series data in our model are required to be stable or stable after differencing

2. Our model can only capture linear relationships by nature, but not nonlinear relationships.

3. If the data used in our model is unstable, then it cannot capture patterns, e.g., stock data is unstable due to the influence of policies and news, and therefore cannot be predicted.

7. Conclusion

This study summarizes several MEMORANDUMS for the trader.

1. Simple and easy to get started. Because the vast majority of our models use the MATLAB language. The highest level of model integration is guaranteed with more concise steps. For most traders, our model is cheaper to learn, but more effective.

2. Our model is commonly practiced. Because most securities firms are using and promoting this type of model, it proves that our model is scientific and reliable, and has wide practice and application in the financial market.

3. The sensitivity is dispatchable, so it is possible to meet the different needs of different clients for different types of investments with just one model, achieving simplicity and efficiency, and maintaining maximum benefits

4. Our model is adaptable and reliable methods have been found (e.g. Var/cVar). So even in volatile financial markets, we can still use this model to predict more reliable data to guide our decisions in the face of uncertainties such as policies, news and other factors.

As you can see from the analysis of our model above, our model is very reasonable and reliable in analyzing some relatively smooth data, but the actual financial and market data are often affected by various factors and are not always smooth.

For some unstable data, we can use some Var and cVar to ensure that the results we obtain can better guide our decisions.

Meanwhile, due to the time limit, we do not find the optimal sensitivity ratios. But we can add an automatic meritocratic algorithm to the existing algorithm to find the optimal sensitivity ratios to maximize the benefits.

References
[1]   Bhuiyan, R. A., Husain, A., & Zhang, C. (2023). Diversification evidence of bitcoin and gold from wavelet analysis. Financial Innovation, 9(1), 100.
[2]   Jareño, F., de la O González, M., Tolentino, M., & Sierra, K. (2020). Bitcoin and gold price returns: A quantile regression and NARDL analysis. Resources Policy, 67, 101666.
[3]   Vantuch, T., & Zelinka, I. (2015). Evolutionary based ARIMA models for stock price forecasting. In ISCS 2014: Interdisciplinary Symposium on Complex Systems (pp. 239-247). Springer International Publishing.
[4]   Hřebík, R., & Sekničková, J. ARIMA model selection in Matlab.
[5]   Pedregal, D. J., Contreras, J., & de la Nieta, A. A. S. (2012). ECOTOOL: A general MATLAB forecasting toolbox with applications to electricity markets. Handbook of networks in power systems I, 151-171.