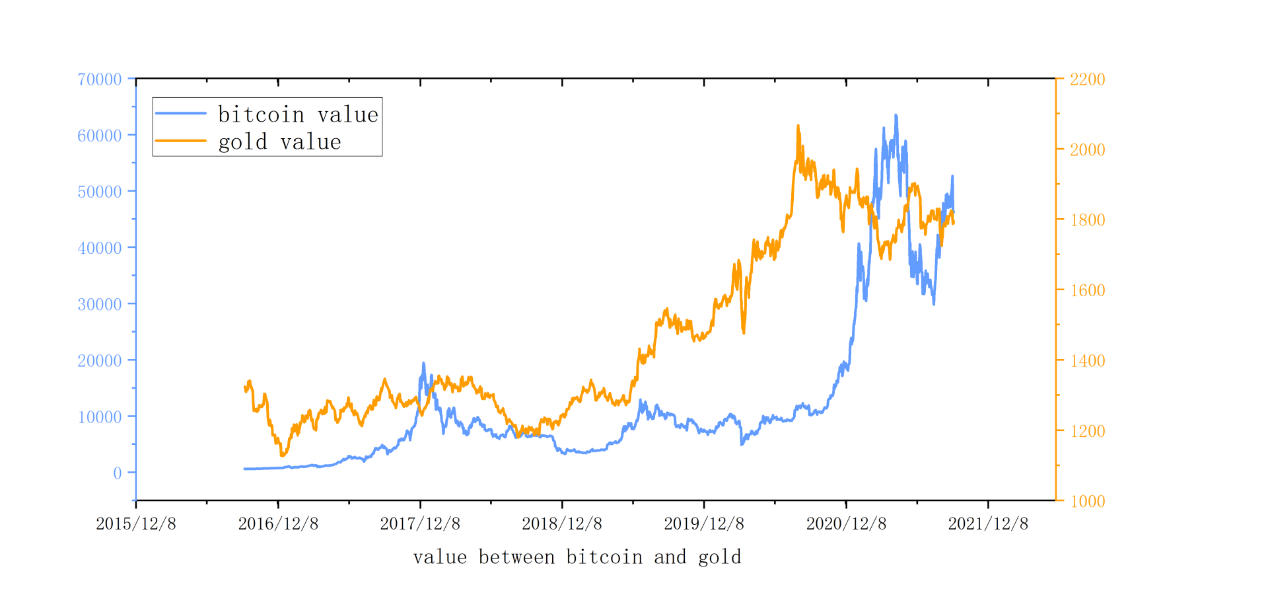
**4.2 Data cleaning**

**i. Data description**

BCHAIN-MKPRU.csv has a total of 1825 valid data. LBMA-GOLD.csv has a total of 1264 records, because gold is only traded on market open days, there are 10 missing values, and the method of adding the adjacent two days to take the average is used to fill in the missing values.

**ii. Data visualization**

The daily trading price comparison chart of Bitcoin and gold is drawn based on the data given by the question as follows:



As can be seen from the figure, compared with Bitcoin, the price fluctuation range of gold is smaller, but at the same time, its rising space is small, which is due to the strong function of gold preservation; while the price range of Bitcoin is very large, and its highest The price can reach 100 times the initial price, and at the same time, it is accompanied by greater risk. The trade-off between bitcoin and gold prices is due to the tendency of consumers to convert bitcoin to gold when market confidence is low.

**4.3 Data processing and mining**

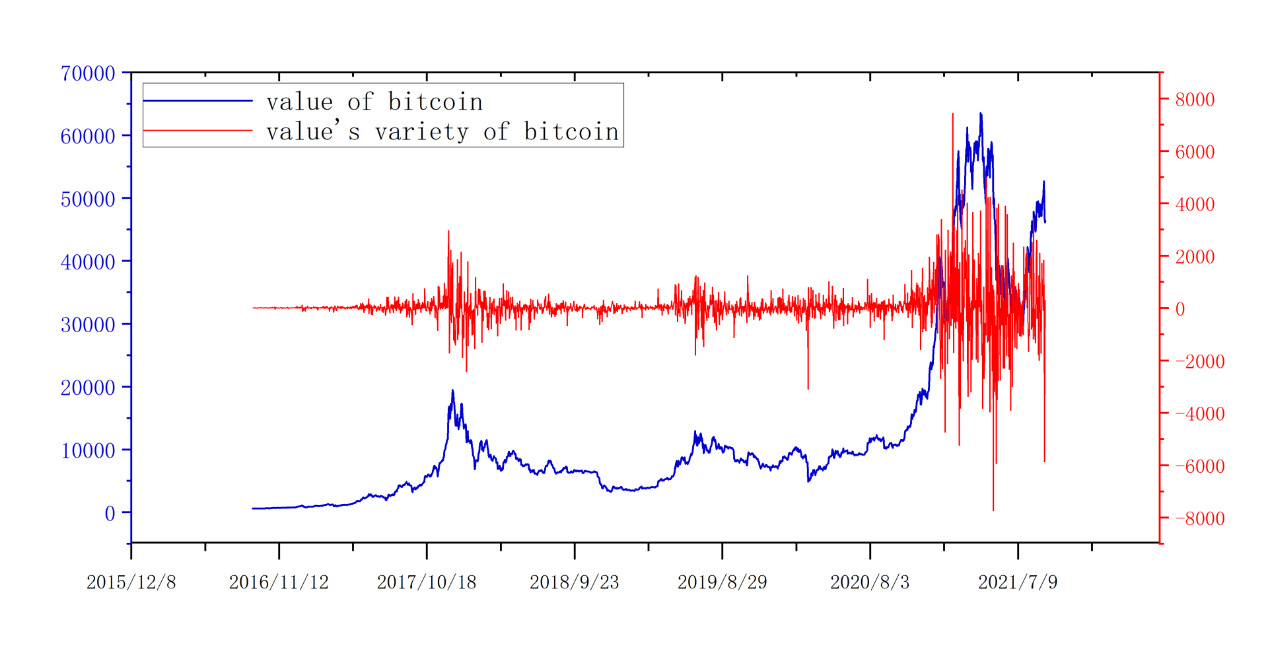
**i. Bitcoin (Gold) price and growth chart**

The daily change in the value of Bitcoin (δ\_bi) is defined as:

δ\_bi=b(i+1)-b(i),

where b(i) represents the bitcoin value on day i

Draw a double y-axis graph as follows, where the left axis represents the bitcoin value ($/bicoin), and the right axis represents the daily change in value ($/bicoin)

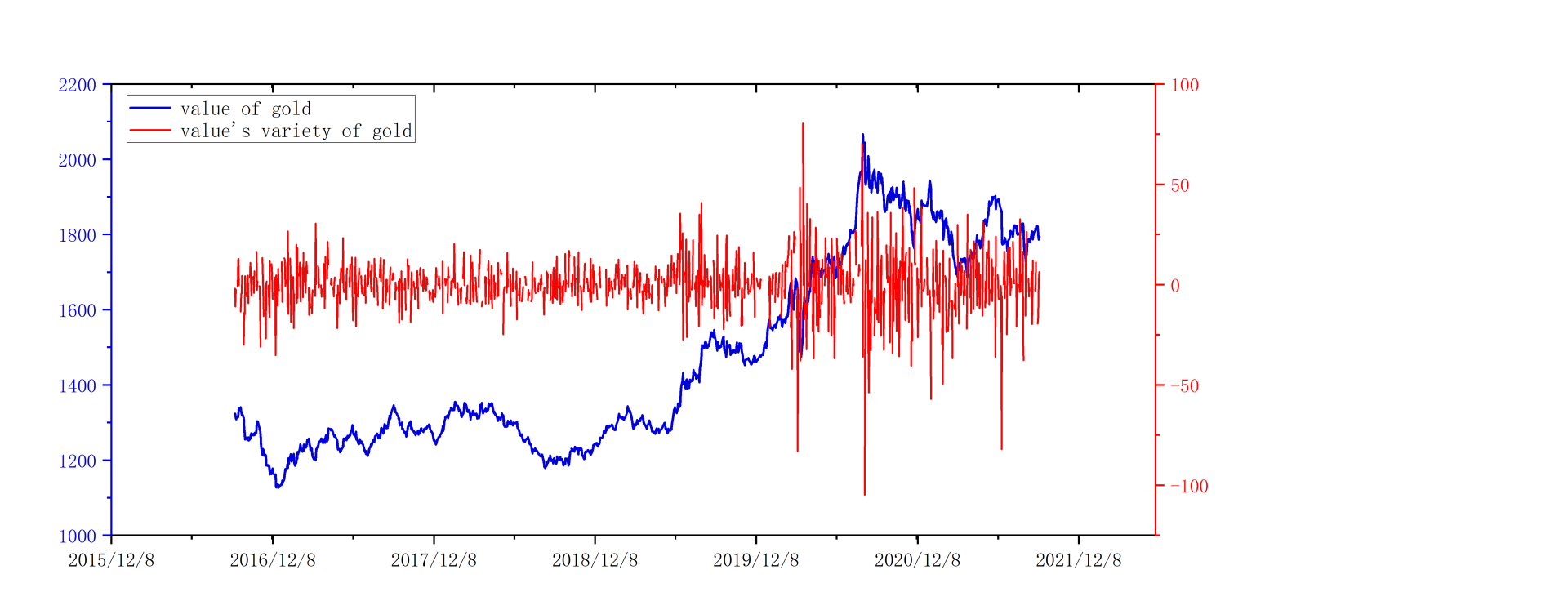


The daily change in the value of Gold (δ\_gi) is defined as:

δ\_gi=g(i+1)-g(i)

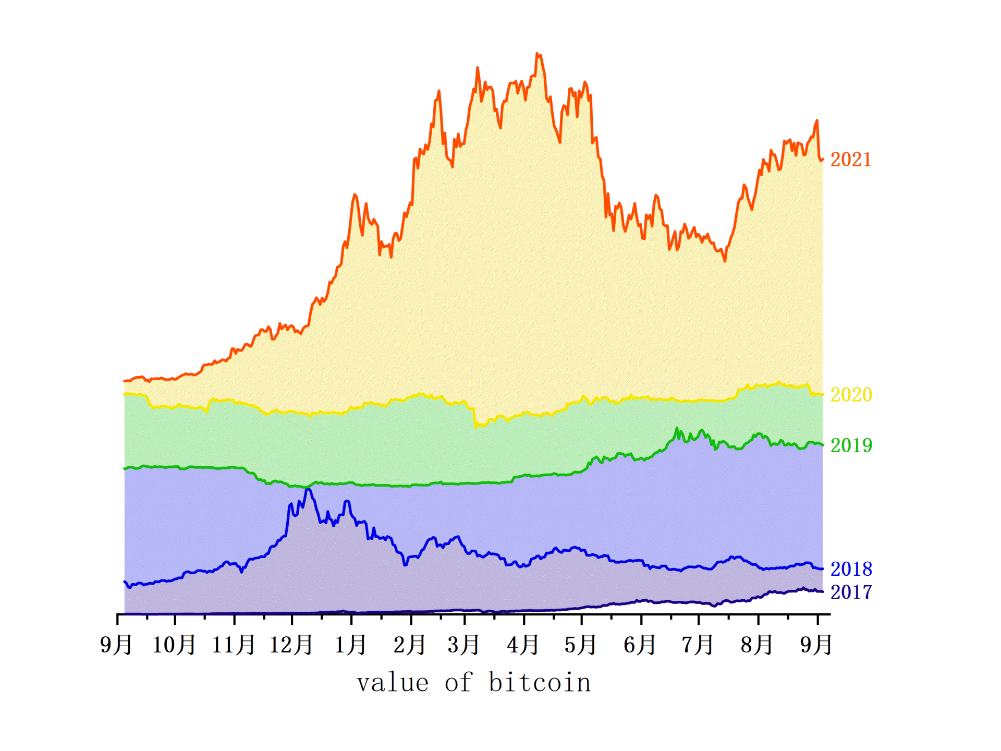
where g(i) represents the bitcoin value on day i

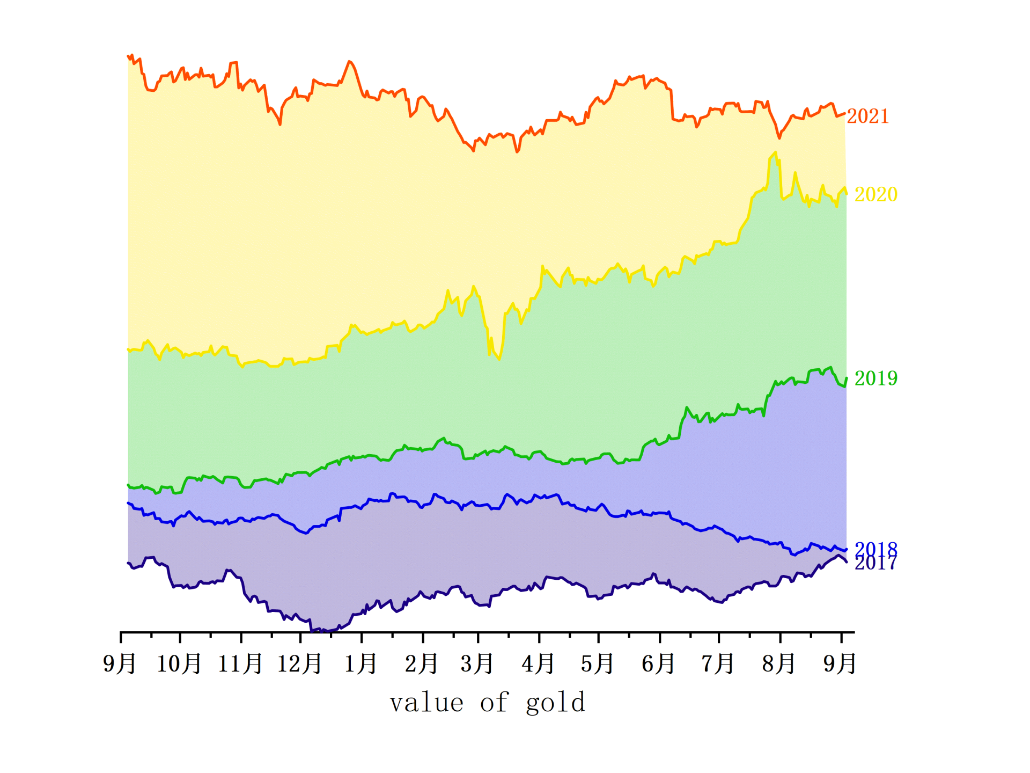
Draw a double y-axis graph as follows, where the left axis represents the value of gold ($/ounce), and the right axis represents the daily change in value ($/ounce)



According to time observation, it can be seen that the data shows the characteristics of "volatility clustering", that is, the variance is relatively small in a certain period, but relatively large in another period, large fluctuations are often followed by large fluctuations, and small fluctuations are often followed by large fluctuations. Volatility is often followed by small fluctuations.

**ii. Year-to-year accumulation trend chart**





Observing the trend stacking chart of the value of Bitcoin (left) and gold (right) from 2016 to 2021, it can be seen that although the two fluctuate greatly in the short term, the overall trend is increasing year by year. Since September 2019, the change of Bitcoin has increased sharply. The reason may be that on the one hand, Bitcoin has continued to fluctuate at a low level, and in the context of the strengthening of the US dollar and the continuous rise of global risk assets such as the stock market, Bitcoin still has no A new low is reached, which shows that the holders of the Bitcoin market have been inclined to reluctant to sell. As long as there is a little news or demand stimulation, the market will rapidly increase buying, leading to price increases; on the other hand, the US Congress has introduced two The cryptocurrency-related bill aims to address the market’s persistent concerns about price manipulation and improve the U.S.’s competitive advantage in the nascent cryptocurrency industry.

**5. Bitcoin and Gold Value Prediction Model**

After consulting the literature, it is found that the quantitative methods used for economic forecasting in the world can basically be classified into the following two types of models: one is the traditional forecasting model based on statistical mathematical modeling, including exponential smoothing method, ARIMA model and Grey forecasting model, etc.; another type is an innovative forecasting model based on the idea of simulation and based on neural network. Each forecasting model has its advantages and disadvantages, such as ARIMA is suitable for stationary time series, neural networks are slow to converge, etc. This paper attempts to use three models to predict the daily value of Bitcoin and gold, and obtains the optimal prediction model through comparison, and then provides support for the formulation of subsequent trading strategies.

**5.1 The Long Short-Term Memory (LSTM)**

**5.1.1 Model introduction**

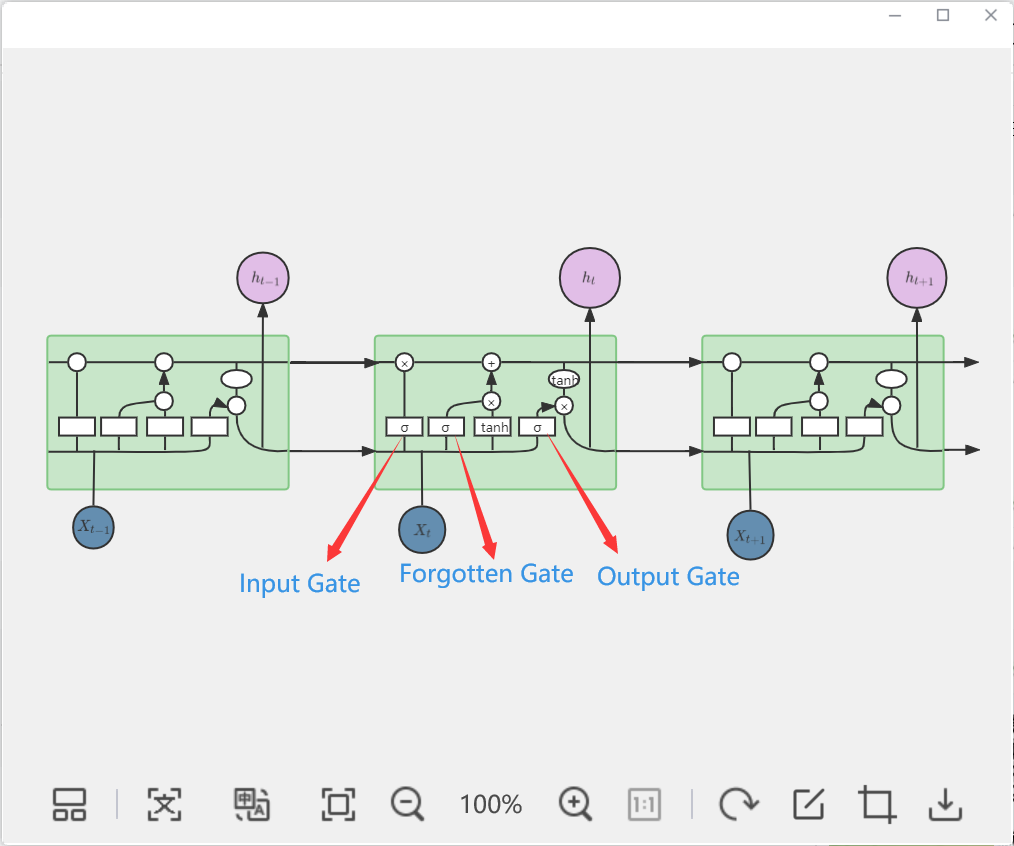
Due to the randomness and non-stationarity of financial time series data, it is difficult for traditional forecasting methods to analyze them accurately. Deep learning methods have more advantages than traditional econometric models in identifying the structure and patterns of data, and can effectively learn the nonlinear and non-stationary characteristics of time series, which is suitable for financial time series forecasting problems. As a deep learning model, the long-term and short-term memory model has long-term memory ability. Each layer of its network layer structure is connected to each other and consists of one or more units with forgettable and memory functions. Can handle time series data better than other neural network models. Conveyor belt transport mechanism, information travels along the entire chain with only some minor linear interactions.

The LSTM has three gates (pictured below) that protect and control the cell state.

**·Input gate:** Determine how much of the input data of the network at the current moment needs to be saved to the unit state.

**·Forget gate:** Determine how much of the unit state at the previous moment needs to be retained to the current moment.

**·Output gate:** Controls how much of the current cell state needs to be output to the current output value.



With this, information can be removed or added to the cell state, whereby structures called gates are carefully regulated. The specific implementation makes use of the sigmoid(σ) function, the sigmoid layer outputs a number between 0 and 1, with a value of 0 meaning "don't let anything through" and a value of 1 meaning "let everything through."

**5.1.2 Model implementation**

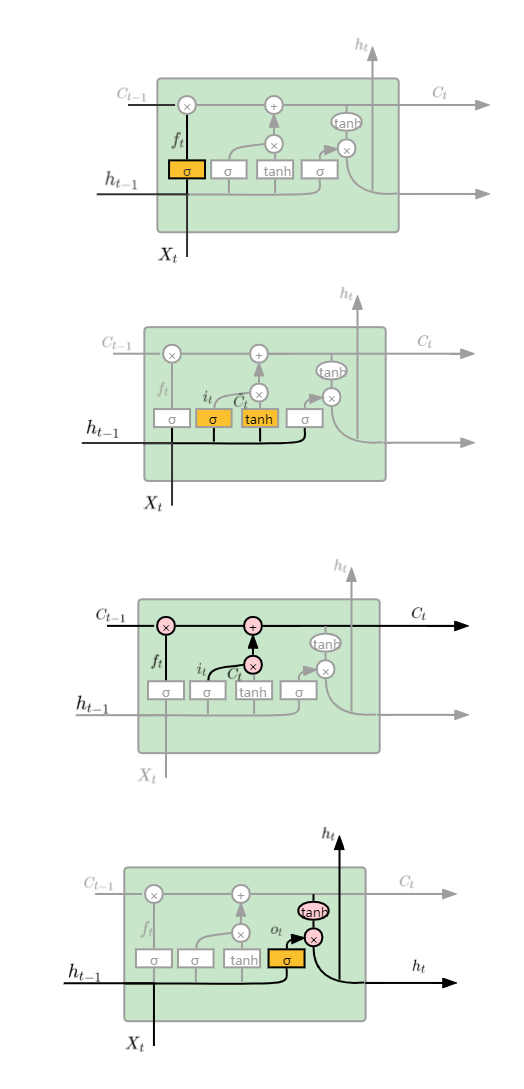
The specific implementation steps of the model (as shown in the following figure) are:

**STEP1:** Apply the output h\_t-1 of the previous moment and the current data input x\_t, and obtain f\_t through the forget gate.

**STEP2:** Apply the output h\_t-1 of the previous moment and the current data input x\_t, obtain i\_t through the input gate, and obtain the temporary state C~t at the current moment through the unit state.

**STEP3:** Apply the cell state C\_t-1 of the previous cell structure, the forget gate output f\_t, the input gate output i\_t, and the cell state output C~t to obtain the current cell state C\_t.

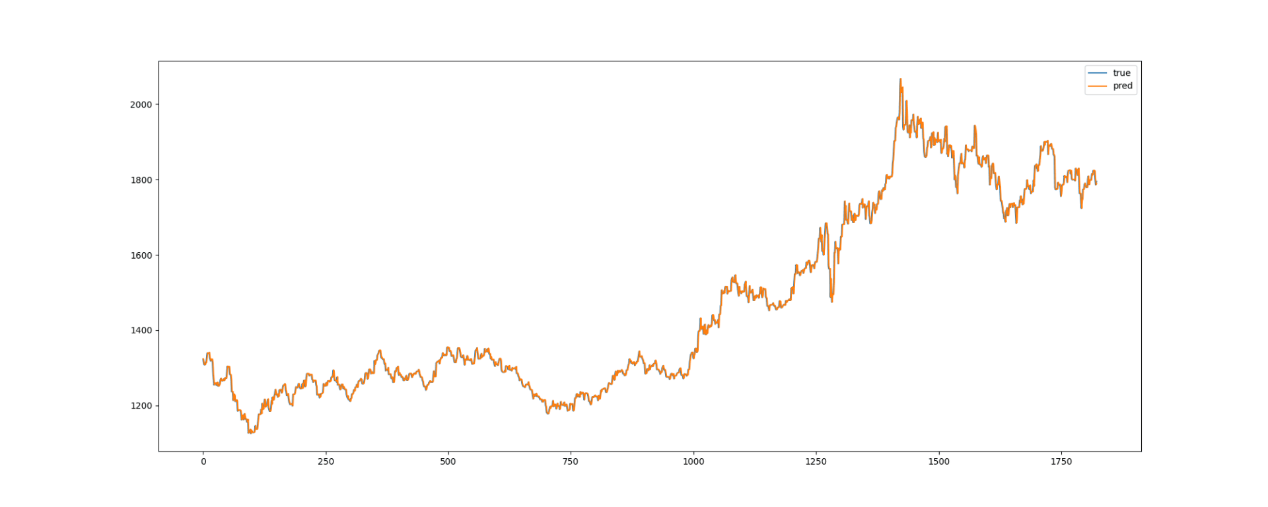
**STEP4:** Apply the output h\_t-1 of the previous moment and the current data input x\_t, obtain the process of o\_t through the output gate, and combine the cell state C\_t and o\_t of the current cell to obtain the final output h\_t.



**5.1.3 Results and Analysis**

Since gold is only traded on working days, the discontinuity of the gold price will increase the difficulty of its price forecasting, so we make up for it to facilitate subsequent forecasting and strategy formulation. Since we can only use the past daily price stream to determine whether to trade, we use the previous day's gold price to fill in missing values.

The prediction results obtained by using the LSTM model to train Bitcoin and gold prices are shown in the following figure.



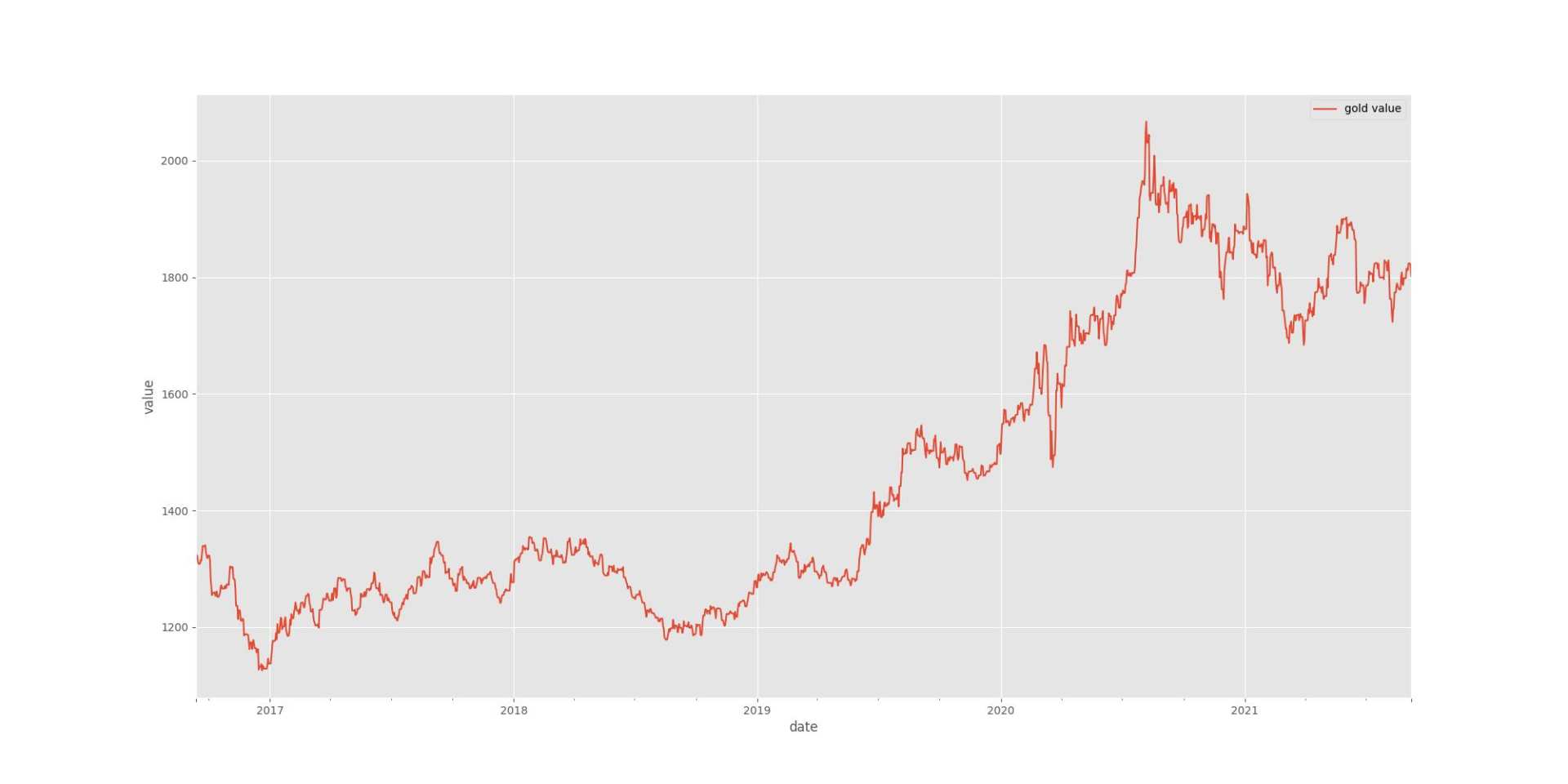
It can be seen from the figure that the predicted value overlaps the actual value with a high degree, and the prediction is more successful.

**5.2 ARIMA time series forecasting model**

**5.2.1 Model introduction**

The full name is Autoregressive Integrated Moving Average Model, abbreviated as ARIMA, where ARIMA(p, d, q) is called differential autoregressive moving average model, AR is autoregressive, p is autoregressive term; MA is moving average, q is moving average The number of terms, d is the number of differences made when the time series becomes stationary. The so-called ARIMA model refers to a model established by converting a non-stationary time series into a stationary time series, and then regressing the dependent variable only on its lag value and the present value and lag value of the random error term.

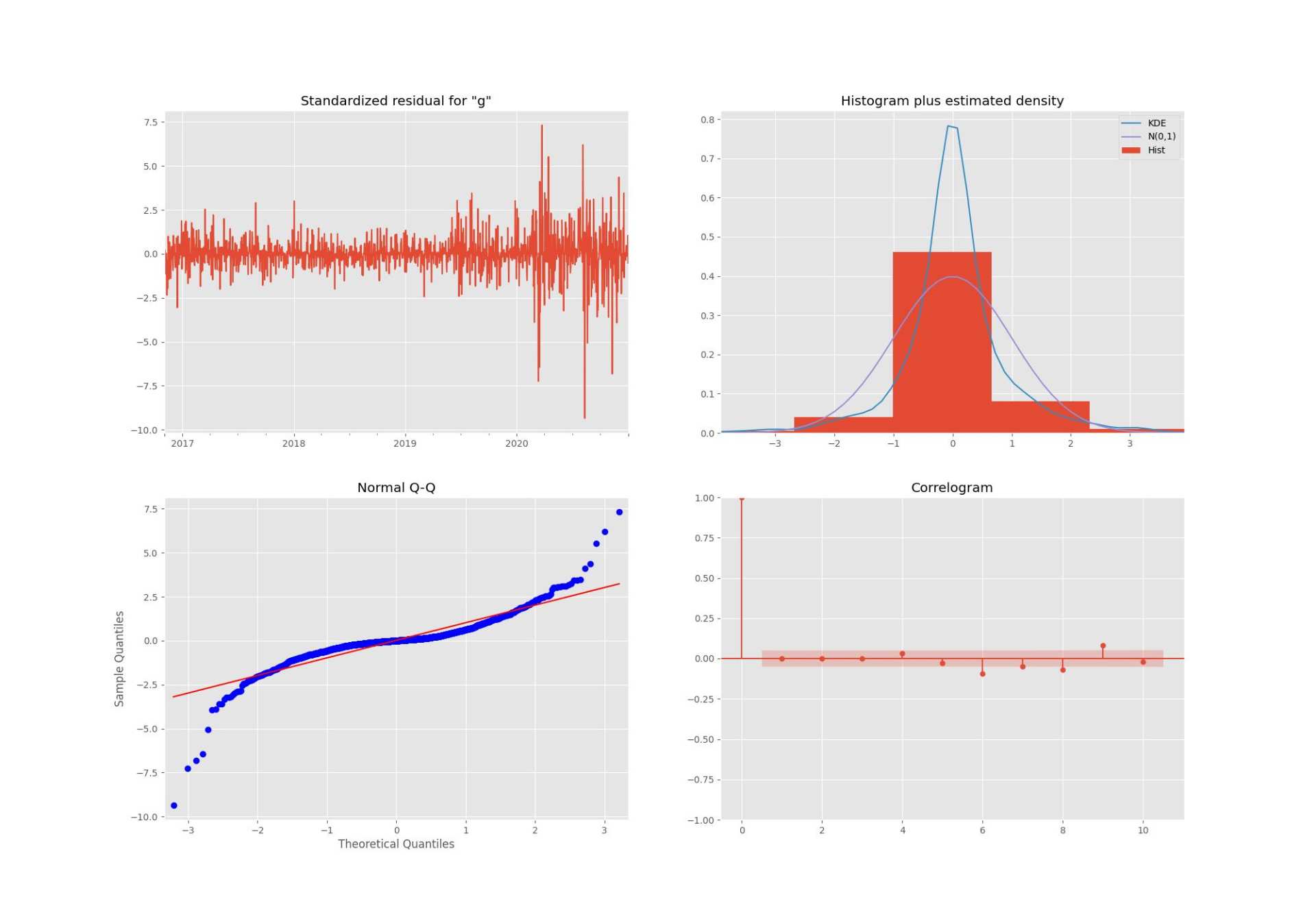
**5.2.2 Model implementation (taking gold price as an example)**

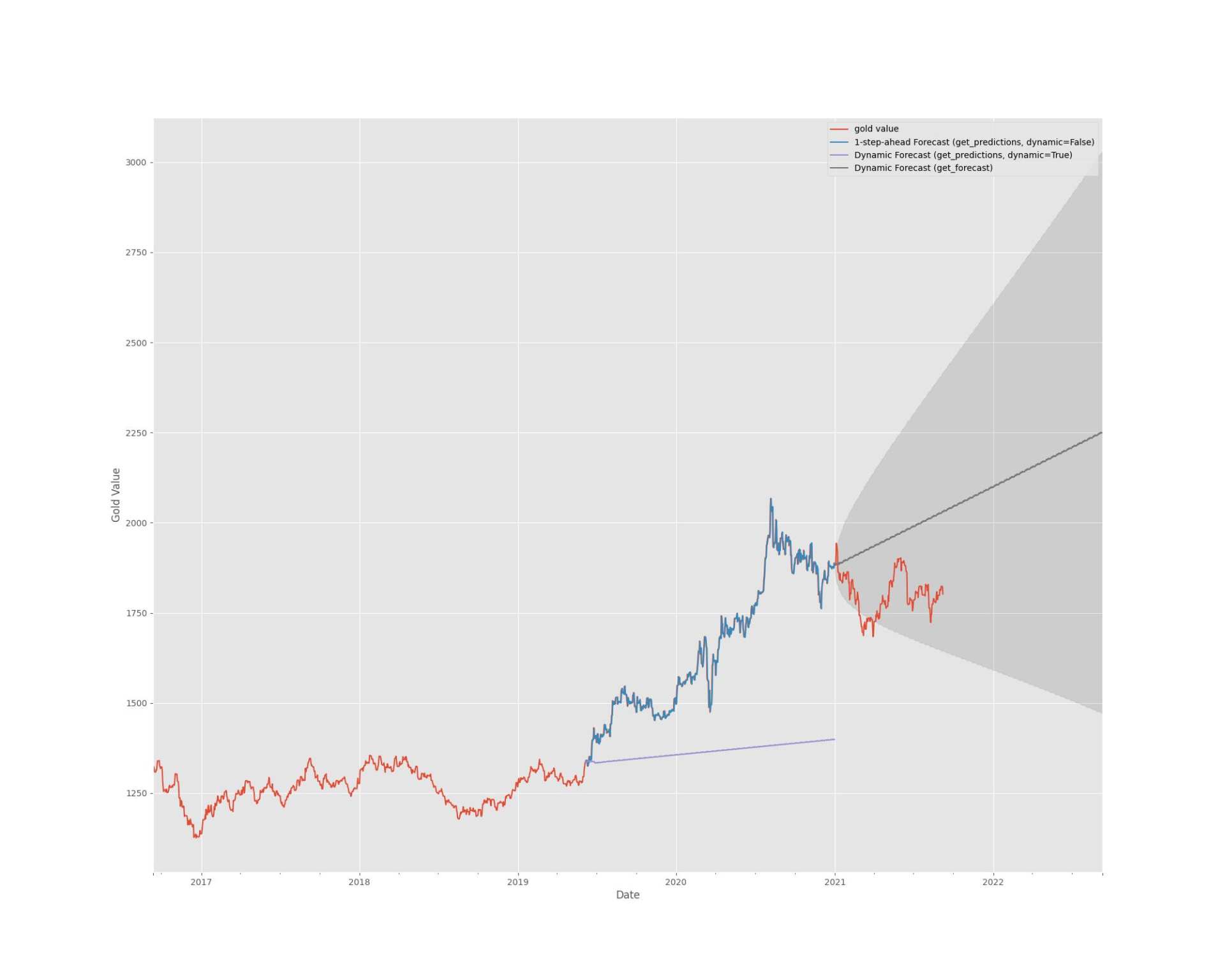


Using the ARIMA model to predict time series data must be stable. If the data is unstable, it is impossible to capture the law. Use differential analysis to process the data stability.

We choose a subset of the data sequence as training data, and our goal is to predict the last day of the series based on this input sequence.

Once the model is fitted, we can check if it is as we expected and if the assumptions we have made are violated. For this, we can use the plot\_diagnostics method. The calculation results are shown below.





**5.2.3 Results and Analysis**

ARIMA essentially only captures linear relationships, not nonlinear relationships. That is to say, to use the ARIMA model to predict time series data, it must be stable. If the data is not stable, it is impossible to capture the law. The price data of gold and bitcoin in this question are unstable and often fluctuate due to the influence of policies and news, so the effect is not good.

**5.3 Grey System Prediction Model**

**5.3.1 Model introduction**

Grey forecasting is a method for forecasting a system with uncertain factors. The gray system is a system between the white system and the black system. It studies a system with few data and uncertainties with "partial information known and partial information unknown". In the currency market, due to factors such as interest rate policy, financial policy, international market, political environment and corporate reform, it is difficult for even the best analysts to accurately grasp it. Therefore, the grey system theory can generate information through the generation of known information. to study these gray uncertain systems.

**5.3.2 Model implementation**

**5.3.3 Results and Analysis**

The GM model has a better prediction effect for data with a clear trend, but for data with frequent oscillations, lack of consideration of the internal mechanism of the system, the prediction effect is significantly affected, and there may be large errors. This is due to GM It is determined by the mechanism of the model itself.

**5.4 Summary**

|  |  |  |
| --- | --- | --- |
| **Model** | **Bitcoin Mean Squared Error** | **golden mean squared error** |
| **LSTM** |  |  |
| **ARIMA** |  |  |
| **GM** |  |  |

Among the three models, the LSTM model has the smallest price prediction error MSE for gold and Bitcoin, indicating that the LSTM model has a stronger advantage in financial time series forecasting. At the same time, because the GM algorithm requires less modeling information and is easy to calculate, it can be used to process short-term predictions of small sample data, that is, predict the short-term prices of Bitcoin and gold, assist LSTM to search for hyperparameters, and then practice to a certain extent. Optimal LSTM model to improve prediction accuracy.

**6. Quantitative Trading Decision Model**

This question requires us to give the best day trading strategy. First define dummy variables to quantify the timing of buying and selling Bitcoin and gold, then assign different weights to different dummy variables, and finally determine whether to buy or sell Bitcoin or gold and the amount of buying and selling. Since investors did not fully understand the market at the beginning, in the first 100 days, no actual investment was made, but the optimal weights of dummy variables were determined as a training set. For the determination of the optimal weights, we take the different weights of the dummy variables as the decision variables, take the maximization of the funds owned on the 100th day as the optimization goal, and take the sum of the weights equal to the first-class limit as the constraint condition to establish a single-objective optimization. The model is solved by genetic algorithm, and the best daily trading strategy is finally obtained. At the same time, we realize the construction of the investment trading platform through programming, only need to input the initial [C, G, B], the bitcoin and gold prices before the trading day and the predicted price on the trading day can realize intelligent investment and obtain ideal returns.

**6.1 Data feature extraction**

**6.1.1 first-level indicator**

Define the daily price of Bitcoin as b\_i, i=1,...,1826; the daily price of Bitcoin predicted by the LSTM model is vb\_i, i=41,...,1826, then the second day of Bitcoin is predicted The difference between the price and the price of the day is

eb\_i=vb\_(i+1)-b\_(i) ,

eb\_i>0, indicates that the price of Bitcoin will rise on the second day, eb\_i<0, indicates that the price of Bitcoin will fall on the second day.

Similarly, the daily price of gold is defined as g\_i, and the daily price of gold predicted by the LSTM model is vg\_i, then the difference between the predicted price of gold on the second day and the price of the current day is

eg\_i=vg\_(i+1)-g\_(i)

Since the volatility of the currency market is too large, it is not conducive to the making of investment decisions, so we make the first-order difference and the second-order difference for Bitcoin and gold as follows:

δ\_bi=b\_(i+1)-b\_(i),

Δ\_bi=δ\_b(i+1)-δ\_bi

i=1,…,1826

Calculate the mean, maximum, and minimum b\_mean, b\_max, b\_min; g\_mean, b\_max, b\_min of the bitcoin and gold prices for the 99 days before the trading day.

To characterize market volatility, we define the rate of change in Bitcoin

bv\_i=(b\_i-b\_(i-1))/b\_(i-1),

gold change rate

gv\_i=(g\_i-g\_(i-1))/g\_(i-1)

Calculate the consecutive rising days BI and GI of Bitcoin and gold, and the consecutive falling days BD and GD before the trading day.

In order to study the periodic changes of Bitcoin and gold prices, we calculate the difference between Bitcoin every 10 days as follows:

δ\_bi(10)=b\_(i-1)-b\_(i-10)

At the same time, the difference between the maximum and minimum values of the ten days is calculated as follows:

Δt\_bi=argmax(i-10<=k<=i-1)b\_k- argmin(i-10<=k<=i-1)b\_k

Similarly, the corresponding variables of gold are as follows:

δ\_gi(10)=g\_(i-1)-g\_(i-10)

Δt\_gi=|argmax(i-10<=k<=i-1)g\_k- argmin(i-10<=k<=i-1)g\_k|

**6.1.2 Trading Signal Construction**

**A. buy or sell bitcoin**

After consulting the relevant information, define dummy variables to construct trading signals and describe the time of buying and selling bitcoins. 1 means buying bitcoins, -1 means selling bitcoins, and 0 means no operation.

**a.** If the daily price of bitcoin is greater than the average value of the previous 100 days, buy bitcoin, and sell if it is less than that, namely, If b\_i<b\_mean, output 1; if **b\_**i>b\_mean, output -1; if not satisfied, output 0

**b.** If eb\_i>0, output 1, if eb\_i<0, output -1; if not satisfied, output 0; if not satisfied, output 0

**c.** If 3<BI<6, output 1; if 3<BD<6, output -1; if not satisfied, output 0

**d.** If BD>6, output 1; if BI>6, output -1; if not satisfied, output 0

**e.** If δ\_b(i+1)(10)<δ\_bi(10), and Δt\_b(i+1)> Δt\_bi, then output 1; if δ\_b(i+1)(10)>δ\_bi(10), and Δt\_b(i+1)< Δt\_bi, output -1; if not satisfied, output 0

**f.** If g\_i>g\_mean, output 1; if g\_i<g\_mean, output -1; if not satisfied, output 0

**g.** If eg\_i<0, output 1, if eg\_i>0, output -1; if not satisfied, output 0

**h.** If 3<GD<6, output 1; if 3<GI<6, output -1; if neither is satisfied, output 0

**i.** If GI>6, output 1; if GD>6, output -1; if not satisfied, output 0

**j.** If δ\_g(i+1)(10)>δ\_gi(10), and Δt\_g(i+1)< Δt\_gi, output 1; if δ\_g(i+1)(10)<δ\_gi(10), and Δt\_g (i+1)> Δt\_gi, output -1; if not satisfied, output 0

**k.** If bv\_i>0.08, output 1; if bv\_i<-0.08, output -1, if not satisfied, output 0

**l.** If gv\_i<-0.08, output 1; if gv\_i>0.08, output -1; if not satisfied, output 0

**B. buy or sell gold**

Define dummy variables to construct trading signals and describe the time of buying and selling gold. 1 means buying gold, -1 means selling gold, and 0 means no operation.

**a.** If b\_i>b\_mean, output 1; if b\_i<b\_mean, output -1; if not satisfied, output 0

**b.** If eb\_i<0, output 1; if eb\_i>0, output -1; if not satisfied, output 0

**c.** If 3<BD<6, output 1; if 3<BI<6, output -1; if not satisfied, output 0

**d.** If BI>6, output 1; if BD>6, output -1; if not satisfied, output 0

**e.** If δ\_b(i+1)(10)>δ\_bi(10), and Δt\_b(i+1)< Δt\_bi, output 1; if δ\_b(i+1)(10)<δ\_bi(10),

and Δt\_b (i+1)> Δt\_bi, then output -1; if not satisfied, output 0

**f.** If g\_i<g\_mean, output 1; if g\_i>g\_mean, output -1; if not satisfied, output 0

**g.** If eg\_i>0, output 1; if eg\_i<0, output -1, if not satisfied, output 0

**h.** If 3<GI<6, output 1; if 3<GD<6, output -1; if not satisfied, output 0

**i.** If GD>6, output 1; if GI>6, output -1; if not satisfied, output 0

**j.** If δ\_g(i+1)(10)<δ\_gi(10), and Δt\_g(i+1)> Δt\_gi, output 1; if δ\_g(i+1)(10)>δ\_gi(10), and Δt\_g (i+1)< Δt\_gi, output -1; if not satisfied, output 0

**k.** If bv\_i<-0.08, output 1; if bv\_i>0.08, output -1; if not satisfied, output 0

**l.** If gv\_i>0.08, output 1; gv\_i<-0.08, output -1; if not satisfied, output

**6.2 Establishment of trading decision-making model - solution based on genetic algorithm**

**6.2.1 Single Objective Optimization Model**

**6.2.2 Model solving**

**A. Introduction to Genetic Algorithms**

Genetic Algorithm (GA) originated from the computer simulation study of biological systems. It is a random global search and optimization method developed by imitating the biological evolution mechanism in nature, drawing on Darwin's theory of evolution and Mendel's theory of genetics. Its essence is an efficient, parallel, global search method, which can automatically acquire and accumulate knowledge about the search space during the search process, and adaptively control the search process to obtain the best solution.

**B. Algorithm principle**

**C. Solution result display**

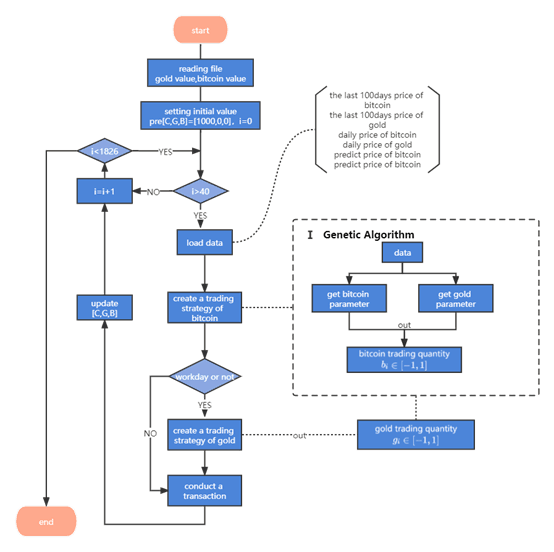
The results of solving the corresponding weights of each trading signal are as follows:

5.214285714 4.571428571 6.571428571 5.142857143 6.928571429 5.357142857 4 5.285714286 6 6 5.428571429 6.785714286 2 3.285714286 7.571428571 4.571428571 7.714285714 2.5 4.5 5.642857143 6.714285714 2.071428571 4.785714286 7.071428571

We use it to implement daily updates of trading strategies, and each day gives whether to buy or sell Bitcoin or gold and how much to buy or sell based on its first 100-day Bitcoin and gold prices, holdings, and predicted prices thereafter.

**6.3 Build a trading platform**

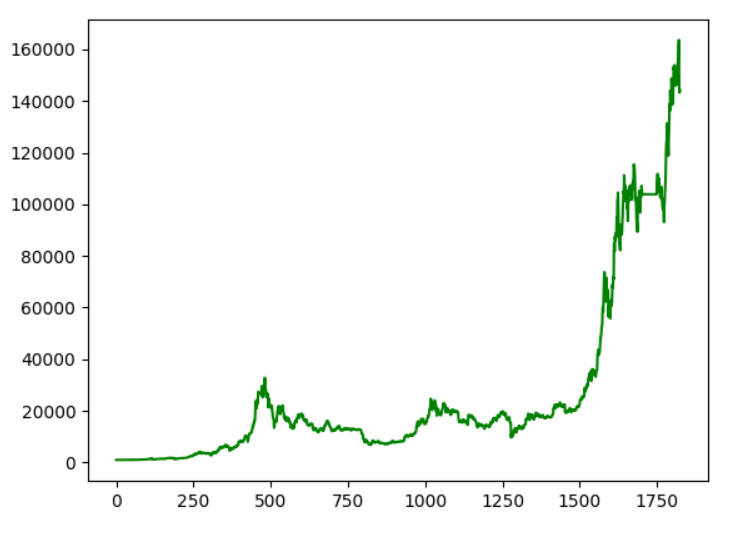
We build the following investment trading platform through programming, just enter the initial [C, G, B], the bitcoin and gold prices before the trading day and the predicted price on the trading day, to achieve intelligent investment and obtain ideal returns.



**i.gate analysis**

**6.4 Final income**

We use the above platforms to arrive at the best investment strategy as follows:



The final income is 142799.045