International Journal of Innovation in Science and Mathematics
Volume 10, Issue 6, ISSN (Online): 2347–9051

Gold and Bitcoin Price Forecasting Based on Multiple Models

Zimu Lin[1*], Han Wang[1], Xiang Zheng[1], Rongrong Shentu[1] and Luyan Wang[1]
[1]School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China.
[*]Corresponding author email id: 1256928155@qq.com

Date of publication (dd/mm/yyyy): 09/11/2022

***Abstract* – Gold and Bitcoin are volatile assets traded in two markets. For centuries, people have been using gold as a wealth preservation tool. With the increasing role of the gold market, statistical modeling and forecasting of gold market fluctuations and prices have attracted widespread attention. With the development of data, modern currency presents the characteristics of virtualization, and some data represented by Bitcoin have achieved rapid development. However, due to the influence of policy, economic situation and many uncertain factors, the transaction price of Bitcoin shows the characteristics of non-linearity, non-stationarity and high volatility, which puts forward a great problem for the prediction of Bitcoin price. In this regard, we introduce the ARIMA model to predict the prices of gold and bitcoin in the next three days while considering the influence of time factors, and select some other prediction models to compare with the performance of the ARIMA model to explore our time series prediction method is good, which also shows that the model can be used as a reference investment model for traders.**

***Keywords* – Quantitative Trading, Time Series Analysis, Forecasting Model, ARIMA Model.**

I. INTRODUCTION

Gold and Bitcoin are volatile assets traded in two markets. Gold is a multi-faceted metal, because it can effectively avoid various risks. For centuries, people have used it as a tool to preserve wealth. In August 2015, the global stock market was hit, and the stock price plummeted. The S&P 500 fell by 10%, but the price of gold denominated in US dollars rose by 5%. This phenomenon also occurred in the 'Black Monday' in 1987 and the financial crisis in 2007.Many examples show that the value of gold continued to rise every year during the war and economic downturn. With the increasing role of the gold market, the statistical modeling and prediction of the volatility and price of the gold market has attracted widespread attention. In fact, gold can be used in portfolios during market shocks to protect global purchasing power [1].

Compared with gold, Bitcoin has high yield, high volatility, free supervision and tax exemption, and there will be huge development space in the financial field. In October 2008, the new concept of Bitcoin was first proposed. Bitcoin was the first decentralized cryptocurrency, while other digital currencies were created by people involved through a mechanism for cloning or tweaking Bitcoin. As of December 2020, Bitcoin remains the largest and most popular cryptocurrency market by market value. Bitcoin accounts account for more than 68 per cent of the total cryptocurrency market value on the market, according to data from the 'Coin Market Cap' website. For now, Bitcoin remains the largest leader in the cryptocurrency market. In the past decade, the price of Bitcoin has been very unstable, which may make its owners earn more than ten times the value of the purchase price overnight, creating more than 3000 % of the absurd profits, or may return to the origin overnight. Severe price fluctuations make the price of Bitcoin difficult to predict. However, as a currency, due to the limited time of Bitcoin, although price volatility has far exceeded the traditional fiat currency, it also provides a certain possibility for price forecasting [2].

Based on this, we use the time series analysis method to predict the price trend of these two financial assets a-

-nd help investors to have a more reasonable estimate of the price level of bitcoin and gold before reinvesting, so as to provide theoretical help and support for their choice.

II. Related Work

A. *Research on Gold Price Forecasting*

Back in 2010, Li used the US dollar index, crude oil prices, US dollar ten-year treasury bond yields, US CPI and other macroeconomic indicators and gold prices as samples, and used a neural network model based on genetic algorithm to predict. The simulation results show that the algorithm can better predict the trend of gold prices and achieve the stability of the prediction results [3].

In 2013, Liu combined with Markov theory to establish a grey-Markov model for predicting gold prices and improved the model. Through comparative analysis, the improved grey-Markov composite model has the best prediction effect. Finally, the composite model is selected to forecast the gold price in the next three years [4]. He introduced GARCH family models to deal with the conditional heteroscedasticity phenomenon in the gold price series, and introduced the dollar index, which has the greatest impact on the gold price, into the mean equation as an exogenous variable. Finally, an EGARCH-X prediction model with exogenous variables was established. The results show that the EGARCH-X model can not only effectively realize the black box prediction function of the neural network, but also greatly improve the prediction accuracy. It can intuitively grasp the relationship between volatility changes and achieve good results [5].

In 2016, Si established the ARIMA model to forecast and analyze the London gold price in the first half of 2015. The results show that the predicted value of the model has a high degree of fitting with the actual data, and the forecast result is more accurate [6]. By logarithming the first-order difference of the daily gold price sequence from 2014 to February 2016, Mao obtained a stable gold price yield sequence, and successfully established the GARCH (1, 1) model, which fits the current price trend well. Then use the model to predict the gold price changes in the next month to make a reasonable forecast [7].

In 2020, Han combined the quotient space theory and support vector machine method to predict the price of gold in China according to the price factor of gold price, and compared the prediction results with the GM(1,1) prediction value and the actual gold price. It is proved that the prediction results of the model are within the allowable range of error and superior to the traditional price prediction method [8].

In 2022, Xu proposed an LM-BP model that combines macro international economic factors with gold price time series. The simulation results show that the algorithm can better predict the trend of gold price, and use the fixed weight threshold method to achieve the stability of the prediction results, so that it can be applied in practice [9]. Huang proposed an improved multidimensional GM(1,N) grey model combined with BP neural network model (IGM-BP), which is applied to the prediction and analysis of gold futures prices, which is conducive to improving the prediction accuracy of gold futures prices. The results show that in the prediction of gold price, compared with the traditional grey neural network model, the IGM-BP prediction model has smaller error and good fitting effect in the high noise nonlinear prediction system, showing higher application value [10].

B. *Research on Bitcoin Prices Forecasting*

In 2015, Liang applied the wavelet analysis method to the prediction of bitcoin price. Combined with the fun-

-ction of wavelet analysis, using the time series data of bitcoin price, the trend of bitcoin price change trend was predicted for the first quarter. The results show that the predicted results of bitcoin price based on wavelet analysis are basically the same as the actual results in the short term [11].

In 2016, Li used BP neural network to establish a bitcoin market forecasting model for the randomness and mutation characteristics of bitcoin market development. The data from August 18, 2010 to March 2, 2016 were selected as samples, and multiple prediction models were constructed. The data of different time periods were used to predict the single-day and one-week bitcoin market. The results show that the closer the data is, the more conducive it is to accurately predict the price of bitcoin, and the more short-term predictions can better fit the development trend of bitcoin [12].

In 2020, Zhang used the CEEMDAN decomposition method to decompose the price of Bitcoin, and used the NAR neural network model and the ARIMA model to predict, and compared the prediction results with the direct use of the two models. The results show that the prediction accuracy is higher after CEENDAN decomposition in the medium and long term, but the accuracy will be reduced in the short term. At the same time, compared with ARIMA model, the prediction accuracy of NAR neural network model is higher [13].

In 2021, Duan used the closing price of Bitcoin from October 2013 to April 2019 for analysis. The ARIMA model was used to first perform a stationarity test on the sequence of the original data. When the data is unstable, the data is smoothed and the unit root test is performed. When the sequence is stationary, the parameters are compared to establish a reasonable ARIMA model. At the same time, compared with the automatically generated ARIMA model, a better model is selected for short-term price forecasting [14].

In 2022, Mo established an ARIMA-Transformer combined model based on wavelet analysis to analyze the random fluctuations, cyclical changes, and periodic changes of time series from different dimensions, and performed a time window rolling prediction of the price of Bitcoin. The predicted results are roughly the same as the actual Bitcoin price trend, indicating that the model can be used as a reference investment model for traders [15]. Han proposed a new model based on graph neural network for bitcoin transaction prediction, which aggregates user's neighborhood information through time attention mechanism, and introduces a novel information feedback mechanism to make full use of network information. The experiment is carried out on two real data sets. The results show that the improved model is about 7%, 6% and 22% higher than the best comparison model under the auc, ap and f1 indicators, respectively, and can analyze and predict bitcoin transactions more accurately [16].

C. *Literature Review*

Most of the existing literature uses time series analysis and machine learning methods to forecast a single asset. However, few scholars have conducted portfolio research from the perspective of traditional assets and digital currencies, that is, forecasting and analyzing both assets at the same time. In order to solve the above problems, this paper first performs data preprocessing to process missing values and outliers in the data set; secondly, a time series model is constructed to model and predict the processed data. In addition, this paper compares ARIMA model with other forecasting models to verify the superiority of ARIMA model.

III. Empirical Analysis

A. *Data Sources*

After cleaning the dataset to remove missing values and some outliers, we divide the dataset into training set and test set. The training set uses the daily gold settlement price in the United States from January 2, 2008 to May 16, 2018, a total of 2290 days, and the daily bitcoin settlement price data from April 28, 2013 to August 14, 2018. The test set is valid from September 11, 2018 to September 10, 2021.

Table 1. Partial data display.

Date	Gold USD (PM)	Date	Bitcoin Value
2008/1/17	1270.95	2013/4/16	762.97
2008/1/18	1219	2013/4/17	11584.83
2008/1/19	1406.8	2013/4/18	3961.493333
2008/2/17	1269.6	2013/4/19	7296.77
2008/2/18	1215.45	2013/4/20	19454.54

Furthermore, to remove the dimensional effect between prices, we normalize the data to account for comparability between data. We use today's price divided by yesterday's price instead of today's price for subsequent forecasts.

B. *Model Overview*

Since the model required by this topic uses only the past daily price stream to determine whether a trader should change an asset's position on a daily basis. Therefore, for this problem, we introduce a time series model to predict the value of gold and Bitcoin in the next three days through curve fitting and parameter estimation. In addition, in order to reflect the superiority of time series models, we selected some models for comparative analysis with time series models.

Time series analysis is a theory and method of establishing mathematical models through curve fitting and parameter estimation based on time series data obtained from systematic observations. It is generally performed using curve fitting and parameter estimation methods such as nonlinear least squares.

1. Auto regression moving average (ARMA model): It is the most commonly used model for fitting stationary series. It can be subdivided into three categories: auto regression model, moving average model and auto regression moving average model.

2. ARIMA model: When the time series itself is not stationary, if its increment, that is, the first difference, is stable near the zero point, it can be regarded as a stationary sequence. Any non-stationary sequence can be fitted with ARIMA model as long as it is stationary after difference by the difference operation of appropriate order.

The detail can be described by by the following equation: $\hat{y} = \mu + \phi_1 \times y_{t-1} + ... + \phi_p \times y_{t-p} + \theta_1 \times e_{t-1} - \theta_q \times e_{q-1}$

Among them, Φ represents the coefficient of AR, and θ represents the coefficient of MA.

C. *Create Time Series Data Objects and Plot Time Series Diagrams*

The gold and bitcoin data are presented as time series, which means our data exists in a continuous time interval with equal intervals between each successive measurement. In R, we are able to use the ts() method to create time series objects for our data vectors. By calling our newly created time series object, store the time series data in the variable Value and the variable USD, and use the plot. ts function to draw the time series diagram.

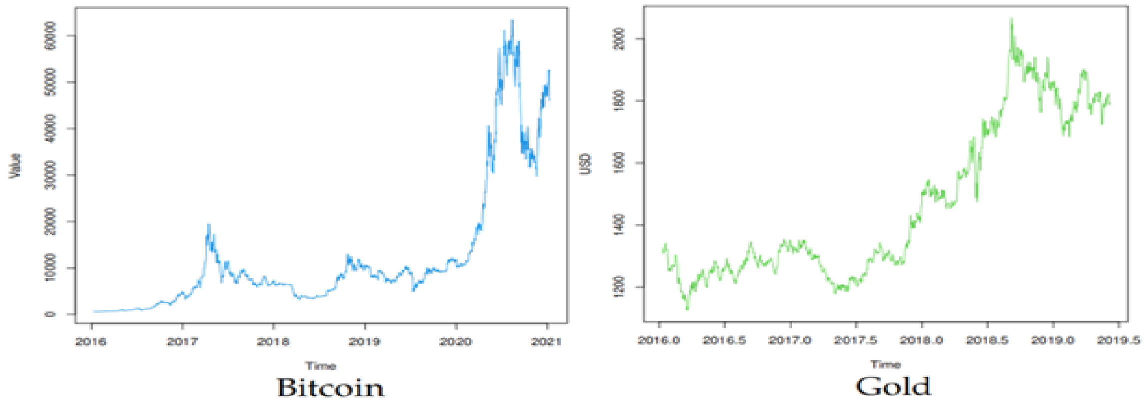

Fig. 1. Time series graph.

From the time series diagram of gold and bitcoin, we can see that whether it is gold or bitcoin, its time series is non-stationary, and generally shows an upward trend from 2016 to 2021. Among them, from 2016 to the beginning of 2017, the development of Bitcoin was relatively slow, and its price trend was relatively flat. From 2017 to early 2018, Bitcoin price started to grow rapidly until it reached its first peak. After a wild rise, the price of Bitcoin began to fall sharply, and by the beginning of 2019, the price of Bitcoin began to recover somewhat. After a relatively stable trend, bitcoin prices have experienced a dramatic recovery in 2020, with prices breaking past peaks to new all-time highs. Gold, on the other hand, has experienced a sharp decline since 2016. After a brief trough, the price rose rapidly and fluctuated slightly. After reaching a second trough in May 2017, it began to gradually increase and peaked in May 2018.

D. *Data Visualization and Time Series Decomposition*

We gain insight into the trend, seasonality, and stationarity of this data by plotting it.

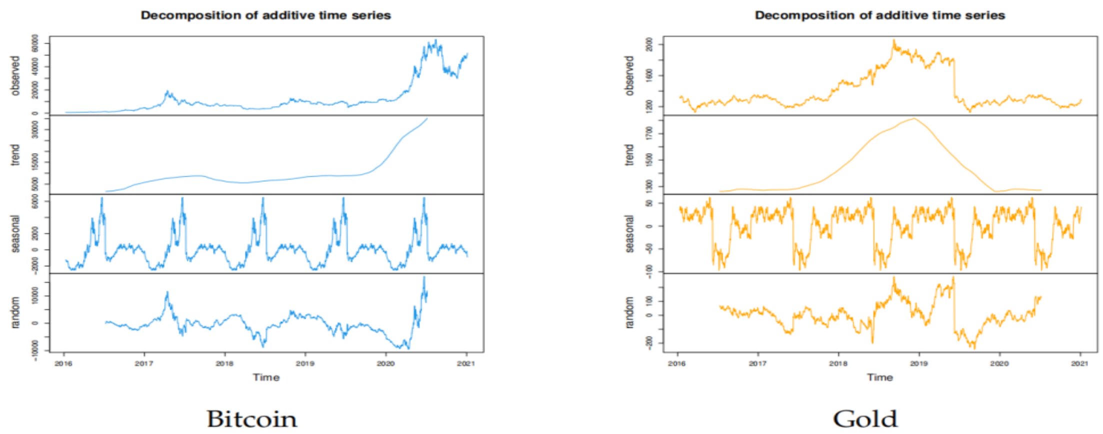

Fig. 2. Decomposition of additive time series.

1. Trend: Refers to the long-term development trend of the data set, such as rising or falling.
2. Seasonality: A dataset has seasonality when it has a pattern that repeats over a known, fixed period of time (eg monthly, quarterly, yearly).

3. Heteroskedasticity: A data is heteroskedastic when its variability is not constant (that is, its variance increases or decreases as a function of the explanatory variable).

4. Stationarity: If the mean and variance are constant (i.e. their joint distribution does not vary with time), then the stochastic process is called stationary.

From the Trend in the above figure, it can be seen that the overall price of Bitcoin is on a continuous upward trend, while the price of gold first rises steadily to a peak and then shows a steady downward trend. As can be seen from the seasonal in the figure, both Bitcoin and gold exhibit cyclical changes and thus have seasonality.

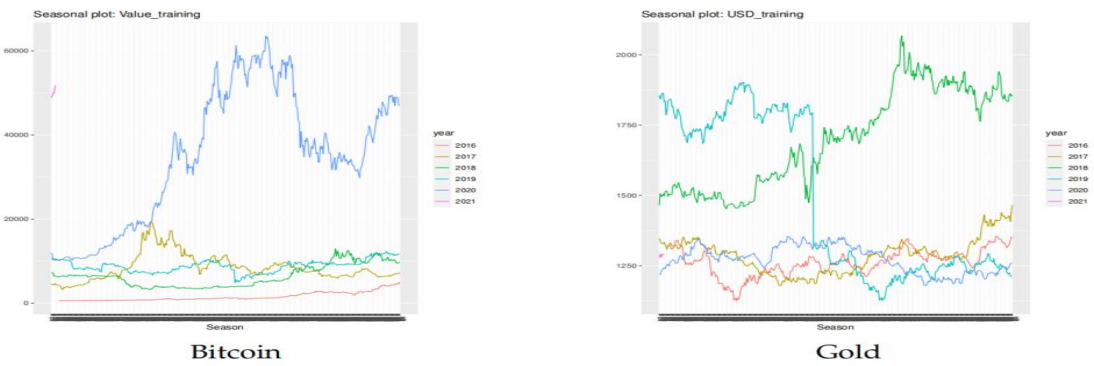

Fig. 3. Seasonal plot: Value_training.

For such obviously non-stationary time series, we can convert them into stationary time series by difference method. Also, to be on the safe side, we still need to test it for randomness and stationarity.

E. *Data Pure Randomness and Stationarity Test*

We have drawn the autocorrelation and partial autocorrelation graphs of gold and Bitcoin respectively. We find that the autocorrelations of the two are tailed, and the partial autocorrelations are all first-order truncations.

Then, we use the Box-Ljung LB test and adf.test of the R language to test, and the results are shown in the following table.

After testing, we can know that the p-value of the adf test of gold and bitcoin is less than 0.05, and the series is stationary.

F. *ARIMA Time Series Model Modeling*

After converting the non-stationary time series to a stationary series, we can use auto.arima to automatically select parameters for it, the parameters of Bitcoin are c(0, 1, 2), and the parameters of gold are c(0, 1, 0). Therefore, we construct the ARIMA model of Bitcoin according to c(0, 1, 2) and the ARIMA model of gold according to c(0, 1, 0).

Table 2. Series: Value_training ARIMA (0, 1, 2) with drift of bitcoin.

Coefficients	**ma1**	**ma2**	**Drift**	**Sigma^2**	**Log Likelihood**	**AIC**	**AICc**	**BIC**
s.e.	-0.0210 0.3911	0.08080.0230	28.086818.5759	626758	-14736.04	29480.07	29480.1	29502.1
ME	RMSE	MAE	MAE	MAPE	MASE	ACF1	ME	
-0.00204981	790.8113	366.0442	-0.6083633	2.913608	0.0362006	0.0019529	-0.00204981	

Table 3. Series: Value_training ARIMA(0, 1, 2) with drift of gold.

Coefficients	drift	ma2	Sigma^2	Log Likelihood	AIC	AICc	BIC
s.e.	-0.0210	0.3911	278.6	-7705.08	15414.15	15414.16	15425.17
ME	RMSE	MAE	MAE	MAPE	MASE	ACF1	ME
0.000727414	16.68229	8.633029	-0.006200354	0.5995043	0.0297419	0.01217508	0.000727414

From the above results, it can be seen that the time series model coefficients of Bitcoin are -0.0786, 0.080, 28.0868, and the standard deviations are 0.0235, 0.0230, respectively. The t-test values obtained by T = coefficient/se are 3.42, and 3.51 is greater than 1.96, respectively. That is, with a 95% confidence interval, the model fits well. Similarly, the golden model fits better.

G. *Model Residual Test*

If the model works well, its residuals should be approximately normally distributed white noise, fluctuating randomly above and below the model's predicted value. Therefore, in order to diagnose the model, we need to perform a residual test on it.

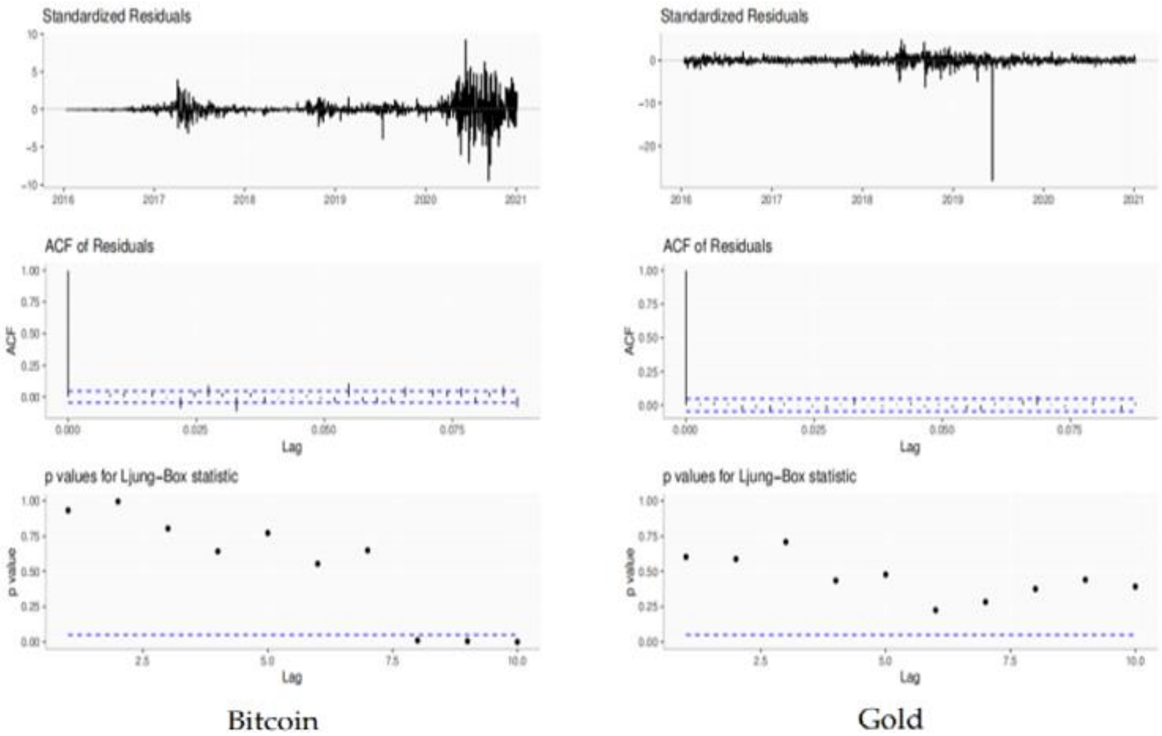

Fig. 4. Model results.

It can be seen from the above results that the time series model coefficients of Bitcoin are -0.0786, 0.080, and 28.0868, and the standard deviations are 0.0235 and 0.0230, respectively. The t-test values obtained by T = coefficient/se are 3.42, 3.51, and both are Greater than 1.96, that is, under the 95% confidence interval, the model fits well. Similarly, the golden model fits better.

H. *Model Prediction*

Import the data of the test set into the variable sp500_TR, fit the time series model trained on the previous training set, and use the forecast library to predict.

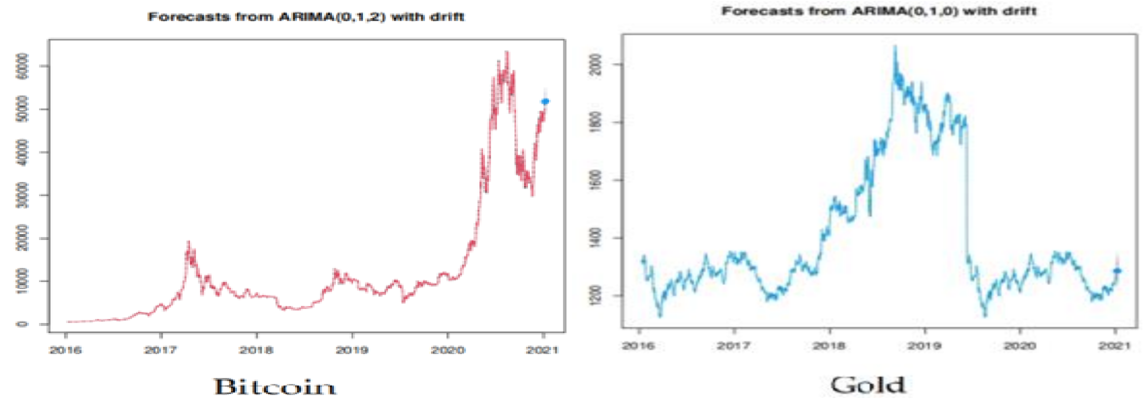

Fig. 5. Forecasts from ARIMA(0, 1, 0) with drift.

As can be seen from the figure, the prediction curve of the test set is well fitted and has a high degree of coincidence with the training set. As a result, the model performs well within the 80% and 95% confidence intervals with high prediction accuracy.

I. *Model Comparison*

In order to explore the effect of time series model construction, we selected model methods such as Box-Cox, Exponential Smoothing, Mean Forecast, Naive Forecast, Seasonal Naive Forecast, Neural Networks to compare with our time series model. The results are shown in the following table:

Table 4. Model comparison chart of bitcoin

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	46697.01	46776.30	46697.01	97.31	97.31	5406.82
Box- CoxTransformation	46696.44	46775.75	46696.44	97.31	97.31	5406.76
ExponentialSmoothing	46696.95	46776.25	46696.95	97.31	97.31	5406.82
MeanForecastMethod	46583.21	46662.69	46583.20	97.07	97.07	5393.65
NaiveForecastMethod	46696 96	46776 25	46696 96	97 31	97 31	5406 82
SeasonalNaiveForecast	46737.17	46816.56	46737.17	97.39	97.39	5411.47
NeuralNetwork	46698.73	46778.01	46698.73	97.31	97.31	5407.02

From the Bitcoin model comparison table, although the mean square error and root mean square error of Mean Forecas are the smallest, in general, the mean square error and root mean square error of several models are not much different, and MPE and MAPE are also relatively close., indicating that these model methods are suitable for Bitcoin.

Table 5. Model comparison chart of gold.

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-3814 96	4737 22	4328 18	-8 28	9 26	11 88
Box- Cox Transformation	-3018.50	3891.08	3627.48	-6.58	7.73	9.95
Exponential Smoothing	-3926.73	4827.00	4380.20	-8.52	9.38	12.02
Meanorecast Method	35855.89	35959.10	35855.88	74.64	74.64	98.41

Naïve Forecast Method	-3785.65	4662.97	4239.82	-8.21	9.07	11.63
Seasonal Naive Forecast	37740.40	37837.49	37740.40	78.59	78.59	103.58
Neural Network	-3783.07	4660.34	4237.83	-8.21	9.07	11.63

From the model comparison table of gold, although the mean square error and root mean square error of Box-Cox, Mean Forecast, and Seasonal Naive Forecast are smaller, the MAPE and MPE of Mean Forecast and Seasonal Naive Forecast are larger and perform poorly. In addition, the Box-Cox transformation is often used to transform non-normally distributed data into an approximately normal distribution, and Seasonal Naive Forecast works better for very seasonal data, so it is not used here. The remaining ARIMA models, Exponential Smoothing Forecast, Naive Forecast and Neural Networks have similar performance in terms of mean square error and root mean square error, but the MAPE of the other models is smaller than Exponential Smoothing Forecast, and the maximum error allowed by ARIMA is higher than that of Exponential Smoothing Forecast. For the subject we study, the price trends of gold and Bitcoin change over time. In the given dataset, the time factor is again particularly pronounced for gold and Bitcoin. Therefore, compared to The Nave Forecast and Neural Networks, we chose the ARIMA model for data analysis, which can take into account the impact of time factors and predict based on past gold and Bitcoin price changes, which is conducive to our trading decision choices.

IV. Conclusion

In order to predict the price trends of bitcoin and gold and thus provide investors with an investment basis, we introduce the ARIMA model to predict the prices of gold and bitcoin in the next three days while considering the influence of time factors, and select some other prediction models to compare with the performance of the ARIMA model to explore our time series prediction method is good, which also shows that the model can be used as a reference investment model for traders. In this regard, we reach the following conclusions: (1) The ARIMA model uses the price trends of gold and bitcoin in the past to predict future prices, and the predicted results are consistent with the laws of past gold and bitcoin price changes. In the process of constructing the time series model, we also pay attention to the influence of seasonal and periodic changes on specific time points while considering the development trend, so as to make the prediction results of gold and bitcoin more accurate. (2) ARIMA model can study the law of the development of things through time series without establishing a causal relationship model, and predict the future development of things accordingly. We use the ARIMA model to analyze the price changes of gold and bitcoin in the past stage. It can be seen that the prices of gold and bitcoin are generally on the rise and show cyclical changes. After the time series is stabilized, the prediction results of the model are more accurate.(3) In the comparative analysis of ARIMA model and other prediction models, we can see that ARIMA model and other models perform well when constructing prediction models based on bitcoin data sets. When constructing a prediction model based on the gold data set, although the performance of other models may be due to the ARIMA model in some values, in general, the values and models of the ARIMA model perform better.

Of course, there are still deficiencies in this study. For example, the development and change of the predicted object is affected by many factors, but in the process of using the time series model to predict in this paper, all the influencing factors are attributed to the time factor, only the comprehensive effect of all the influencing factors is recognized, and the causal relationship between the predicted object and the influencing factors is not analyzed. In addition, in the process of constructing the time series model in this paper, we only predict according to historical data, without considering the possibility of market changes. In the future, it is hoped that the data can be expanded to explore the price trends of gold and bitcoin more comprehensively.

REFERENCES

[1] Li Haohuo.Quantitative trading strategy design based on genetic programming algorithm [D].Shanghai Normal University, 2021.
[2] Duan Huan. Research on gold futures price prediction based on time series [D].Harbin University of Technology, 2021.
[3] Li Jin. Application of GA-BP algorithm in gold price forecasting [J].Financial Economics, 2010 (02) : 73-74.
[4] Liu Chengjun, Yang Peng, Lv Wensheng, Huang Ping. Application of Grey-Markov Composite Model in Gold Price Forecasting [J]. Nonferrous Metals (Mining Part), 2013,65 (01) : 7-11.
[5] He Guoling. Research on gold spot price fluctuation based on GARCH family model [D]. Tianjin University of Finance and Economics, 2013.
[6] King. Short-term prediction of gold price based on ARIMA model [J]. Time Financial, 2016 (17) : 221-222.
[7] Mao Congqin.Gold price prediction analysis and mechanism research based on GARCH model [D].Suzhou University, 2016.
[8] Han Xu, Yang Shan, Wang Ximei. Gold Price SVM Model Prediction Based on Quotient Space [J]. Gold Science and Technology, 2020, 28 (01) : 148-157.
[9] Xu Jingyi, Kong Mengqi. Research on the price of COMEX gold futures in New York based on ARIMA model [J]. China, 2022(18): 123-125.
[10] Huang Qian, Liu Sheng. Gold price prediction based on IGM-BP model [J]. China Price, 2022 (10) : 87-89.
[11] Liang Qiu, Wang Fanbin, Yu Longxiu, Li Qiang, Wu Bingjie. Bitcoin price forecasting based on wavelet analysis [J]. Neijiang Technology, 2015, 36 (05) : 129-130.
[12] Li Jing. Using BP neural network to build a bitcoin market forecasting model [J].Accounting Monthly, 2016 (21) : 33-36.
[13] ZHANG Ming. Bitcoin price prediction based on NAR neural network under CEEMDAN decomposition [J]. Journal of Northeast Bay University, 2020, 35 (11) : 54-62.
[14] Paragraph grid. Research on bitcoin price prediction based on ARIMA model [J]. Modern Marketing (Late Edition), 2021(01): 27-29.
[15] Mo Shibing, Lin Huijun, Chen Yunwei, Liang Yi, Wang Zhenkun. Bitcoin price prediction of Arima-Transformer combination model based on wavelet analysis [J]. Modern Information Technology, 2022,6 (03) : 32-35.
[16] Han Zhongming, Wang Yuhang, Mao Yajun, Chen Fuyu. Bitcoin trading prediction based on graph neural network [J / OL]. Computer application research: 1-8 [2022-10-27]. DOI : 10.19734.

AUTHOR'S PROFILE

First Author
Zimu Lin, School of Accounting (Corresponding author), Zhejiang University of Finance and Economics, Hangzhou, China.

Second Author
Han Wang, School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China.
email id: 3047470075@qq.com

Third Author
Xiang Zheng, School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China.
email id: 3412201772@qq.com

Fourth Author
Rongrong Shentu, School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China.
email id: 228620027@qq.com

Fifth Author
Luyan Wang, School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China.
email id: 2083147472@qq.com