Bitcoin price forecasting using ARIMA model

Haolin Tian

College of Alameda, Peralta Community College District, Oakland, CA 94606, United States

30032933@cc.peralta.edu

Abstract. The Bitcoin price was chosen as the research subject, and the observation period was set from January 2015 to September 2023. An ARIMA time series model was constructed to forecast the trading price. The results indicate that the optimal model for fitting the trading price is ARIMA (3, 2, 8). This model takes into account trends, seasonality, and other factors that may impact the price of Bitcoin. By analyzing the historical data, the model was able to accurately predict the short-term fluctuations in Bitcoin's trading price. Based on this, short-term predictions were made for Bitcoin's trading price in the next year. Recommendations were then provided by combining the forecast results with the economic development situation in the post-pandemic era. The recommendations suggest that Bitcoin has become a low-quality asset and is no longer suitable for diversifying one's investment portfolio, but rather focus on the development of physical industries and adjust one's investment portfolio in a timely manner.

Keywords: Bitcoin, ARIMA, financial modeling, cryptocurrency, time series analysis.

1. Introduction

As a cryptocurrency, Bitcoin has attracted wide attention in the financial field. Nakamoto invented Bitcoin and expressed it as a decentralized digital currency that uses blockchain technology for transaction records and verification [1]. Over time, Bitcoin has become more widely known, and demand for Bitcoin has also begun to increase, especially during the pandemic. Rebucci et al. pointed out that to drastically reduce the impact of the COVID-19 virus on the economy, central banks and governments around the world have introduced loose monetary and fiscal policies [2]. However, under the loose monetary and fiscal policies and the pessimism shown in the markets, global capital has sought safe-haven assets. Zhao et al. have shown that Bitcoin, as an alternative asset for value storage, is also increasingly favored by safe-haven funds [3].

Ciaian et al. pointed out in their research that the high volatility and uncertainty of the Bitcoin market provide investors and traders with enormous risks and opportunities [4]. Therefore, the ability to accurately predict Bitcoin prices is crucial for individuals and investors. So, if we can accurately predict the future price trend of Bitcoin transactions, it will not only help individuals make judgments in advance and achieve their diversified investment portfolios; it can also provide investors with short-term market direction indicators and issue warnings before turning points as a supplement to investment strategies.

However, predicting Bitcoin prices is a challenging task. Kristoufek showed that the Bitcoin market is influenced by many factors, such as market sentiment, trading volume, and market liquidity [5]. In previous studies, many scholars have tried to use various methods to predict Bitcoin prices. For example, Karasu used machine learning-based methods, such as SVM technology, to predict Bitcoin prices [6]. Kim et al. compared the performance advantages and disadvantages of various prediction methods based on deep learning technology [7]. However, these methods may have limitations in dealing with nonlinear and non-stationary data.

This article will use time series prediction methods to predict Bitcoin prices, for the following two reasons. On the one hand, the Bitcoin market has some highly similar characteristics to the stock trading market, which makes time series analysis more applicable in this market. Bitcoin market trading activities are more continuous and uninterrupted, with large numbers of transactions occurring daily. This continuity and high frequency of trading causes the Bitcoin market price data to form a continuous time series, which contains rich historical information. At the same time, the Bitcoin market has high price volatility, which means that prices may change significantly in a short period. The openness and transparency of the Bitcoin market, allow market participants to extensively access market data, including trading volume, price fluctuations, and market depth.

On the other hand, time series analysis methods are quite mature and have good performance in various fields. A. Ariyo and his research partners used the ARIMA model for stock price prediction, proving that the ARIMA model has great short-term predictive potential [8]. Mbachu et al. researched house price prediction. They compared the predictive performance of autoregressive integrated moving average (ARIMA) models and multiple linear regression (MLR) models. They concluded that ARIMA models are generally superior to regression models and more reliable in accurately monitoring and predicting housing prices [9]. Kriechbaumer et al. proposed an improved wavelet LET-ARIMA method to predict metal prices. This method first decomposes the original price time series using wavelet transforms, then establishes ARIMA models for each component time series for prediction, and finally adds the predicted results. Empirical results show that compared with the classical ARIMA model, this method can significantly improve the prediction accuracy of aluminum, copper, lead, and zinc prices [10].

In summary, due to the continuity, high frequency of transactions, price volatility, and openness and transparency of the Bitcoin market, time series analysis methods may have special applicability in the Bitcoin market, which can help everyone reveal price trends and make decisions accordingly. By using a large amount of historical Bitcoin price data for model training and validation, this article hopes to provide an accurate and reliable method to predict Bitcoin price volatility.

2. Methodology

This section of the article has three objectives. Firstly, it introduces the knowledge of the ARIMA model used for time series forecasting. Secondly, it provides a detailed explanation of the modeling process and approach using diagrams. Lastly, it discusses the parameter metrics used to assess the effectiveness of the predictions.

2.1. ARIMA model approach

The ARIMA (Autoregressive Integrated Moving Average) model is a widely used time series model for forecasting future values based on historical data. It was proposed by George E. P. Box and Gwilym M. Jenkins in 1976.

The development of the ARIMA model can be traced back to the earlier work on autoregressive (AR) and moving average (MA) models. The AR model, also known as the autoregression model, studies the dependence between a variable of interest and its past values. It represents regressing the variable against itself. The AR model of order p, denoted as AR(p), considers the dependencies with lagged values up to p. The formula expression for AR(p) is:

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + x + \emptyset_p y_{t-p} + \varepsilon_t + c \quad (1)$$

Here, y_t is the stationary variable, $\emptyset_i$ represents the autocorrelation coefficient at lag i, ε_t is the normally distributed white noise with mean zero and variance one, and c is a constant term.

On the other hand, the MA model, or moving average model, forecasts the variable of interest based on past forecast errors. It represents the regression of the variable against the error terms. The MA model of order q, denoted as MA(q), considers the dependencies with lagged error terms up to q. The formula expression for MA(q) is:

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + x + \theta_p \varepsilon_{t-p} + \varepsilon_t + \mu \quad (2)$$

Here, θ_i represents the moving average coefficient at lag i, ε_t and ε_{t-i} are the white noise error terms, and μ is the mean of the series.

ARIMA (p, d, q) represents a composite model that incorporates differencing (d), autoregression (AR), and moving average (MA). The ARIMA model is a further extension of the ARMA model that retains its advantages while overcoming its limitations. It has two main improvements:

First, handling non-stationary time series data: The ARMA model has limited effectiveness in modeling and predicting non-stationary time series. The ARIMA model adds a differencing step to the ARMA model, where the original data is differenced to transform the non-stationary time series into a stationary one, thus better handling non-stationary data.

Secondly, seasonality elimination: The ARMA model lacks flexibility in dealing with time series data with trends and seasonality. The ARIMA model introduces seasonal differencing to remove the seasonal effects from the data, resulting in more accurate modeling and prediction of time series with seasonality.

In summary, the ARMA model addresses the limitations of the AR and MA models, while the ARIMA model further extends the capabilities of the ARMA model by handling non-stationary data and eliminating seasonality. The ARIMA model retains the flexibility and predictive accuracy of the ARMA model, making it a valuable tool in time series analysis and forecasting.

When using the ARIMA model, it is important to preprocess the data by applying differencing to stabilize the mean and eliminate trends. If the data exhibits seasonality, seasonal differencing can be applied. The ARIMA model's formula expression is:

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-p} + \varepsilon_t + c \quad (3)$$

Here, y_t represents the differenced time series. The model parameters ($\emptyset_i$ and θ_i) are estimated based on the data, and the white noise assumption is used for the error terms (ε_t).

2.2. Verify accuracy

Due to the fact that data forecasting involves predicting future data based on historical data changes, it is impossible for anyone to accurately predict data and guarantee the absolute accuracy of the results. This means that we need to consider the level of error present in the forecasts. The most important criterion for measuring forecast accuracy is prediction error. Therefore, in this paper, we will use Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) to evaluate the forecast accuracy of different ARIMA models. The formulas for calculating MAE, MAPE, and RMSE are listed in the table below. In the formulas, y_t represents the actual values, y t̂ represents the predicted values, and n represents the number of forecast periods. By analyzing the prediction errors, we can select the ARIMA model with the smallest prediction error, thereby determining the best-fitting model.

$$Mean\ Absolute\ Error\ (MAE) = \sum_{t=1}^{n} \left| \frac{yt - \hat{y}t}{n} \right| \quad (4)$$

$$Mean\ Absolute\ Percentage\ Error\ (MAPE) = \left(\sum_{t=1}^{n} \left| \frac{yt - \hat{y}t}{n} \right| \right) \frac{100}{n} \quad (5)$$

$$Root\ Mean\ Square\ Error\ (RMSE) = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(yt - \hat{y}t)^2} \quad (6)$$

3. Results and discussion

3.1. Data sources and analysis environment

First, the selected Bitcoin data consists of 3,171 daily adjusted closing prices from January 1, 2015, to September 6, 2023, sourced from Kaggle. Secondly, all empirical analyses in this paper were conducted using RStudio software.

3.2. Descriptive statistical analysis

Before conducting the analysis, descriptive statistical analysis was performed on the daily closing prices of Bitcoin in this study. The first step was data preprocessing. By calling the summary function, can prove that the data is very clean. The next step was to convert the Bitcoin closing price data into a time series. Figure 1 shows the time series plot of the Bitcoin closing prices. It can be visually observed from the plot that there is a trend and seasonality. To avoid subjective errors, the “stl” function was used in RStudio to decompose the time series into seasonal, trend, and irregular components to determine the stationarity of the time series. As shown in Figure 2, the time series exhibits strong seasonality and a significant downward trend after 2017.

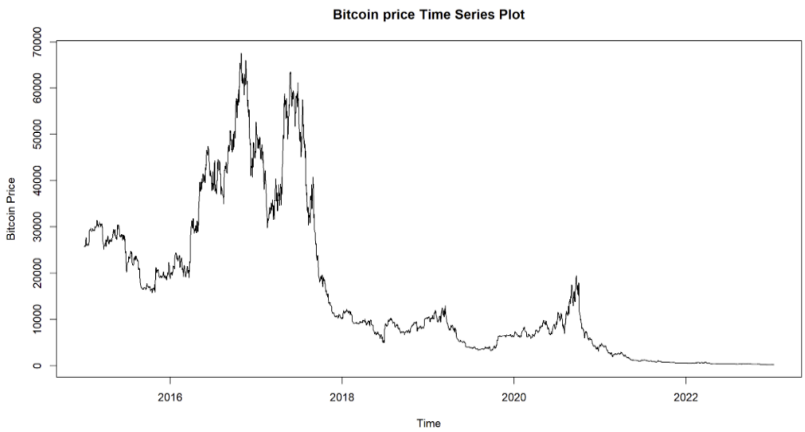

Figure 1. Time series chart of Bitcoin’s adjusted closing price.

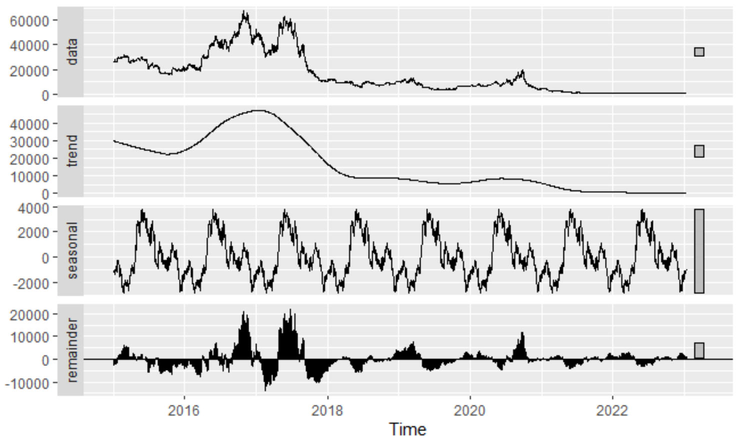

Figure 2. Decompose the time series

3.3. Data processing

Next, the original time series was subjected to first-order differencing to obtain the first-differenced time series plot (as shown in Figure 3). Subsequently, the Augmented Dickey-Fuller (ADF) test was performed, and the results are presented in Table 1.

From Figure 3, it can be observed that the first-differenced time series exhibits randomness in the temporal dimension, without any noticeable trend or periodicity. Table 1 shows that the p-value of the ADF test is less than 0.01, which is below the significance level of 0.05. The p-value in the KPSS test is 0.1, which is greater than 0.05, so the time series is stationary. This suggests that the first-order differencing has been effective in eliminating the non-stationarity of the original series.

Based on the comprehensive analysis above, it can be concluded that the first-order differencing has achieved the desired effect. Consequently, further analysis can be conducted.

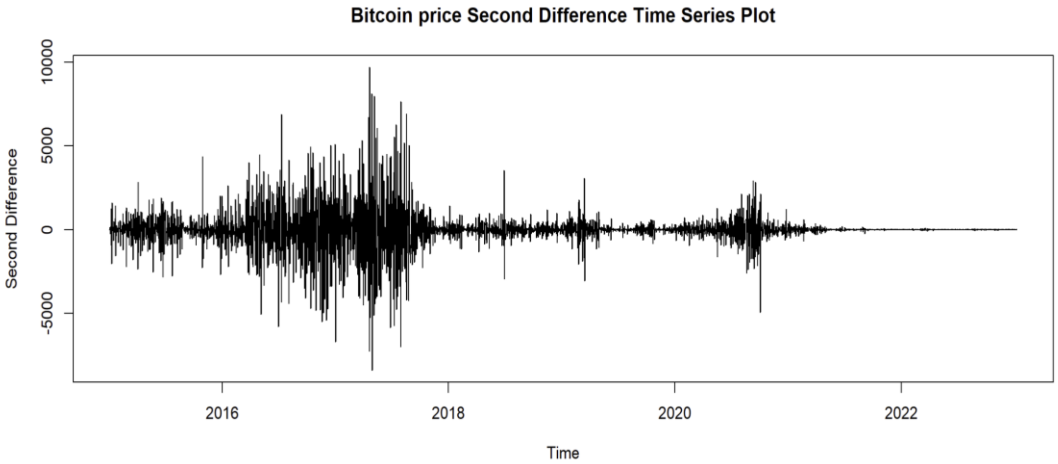

Figure 3. Second-order differential time series.

Table 1. ADF test results for first-order differences

Method	P-Value
Dicky-Fuller	0.01
KPSS	0.1

3.4. Model recognition and fitting

Using first-order difference sequences, autocorrelation and partial autocorrelation tests were conducted to determine the order of the model. The results of the ACF and PACF analyses are presented in Figure 4 and 5, respectively. The ACF displays the correlation between each observation and its lagged values, while the PACF represents the correlation between an observation and its lagged values, excluding the influence of intermediate lags. These visualizations aid in determining the number of significant lags to include in the model. Furthermore, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were employed to select among the different models. These criteria balance the goodness of fit with the complexity of the model. By comparing the AIC and BIC values for the 18 models under consideration, the model with the order (3,2,8) was ultimately chosen. This result provides important evidence for subsequent model fitting.

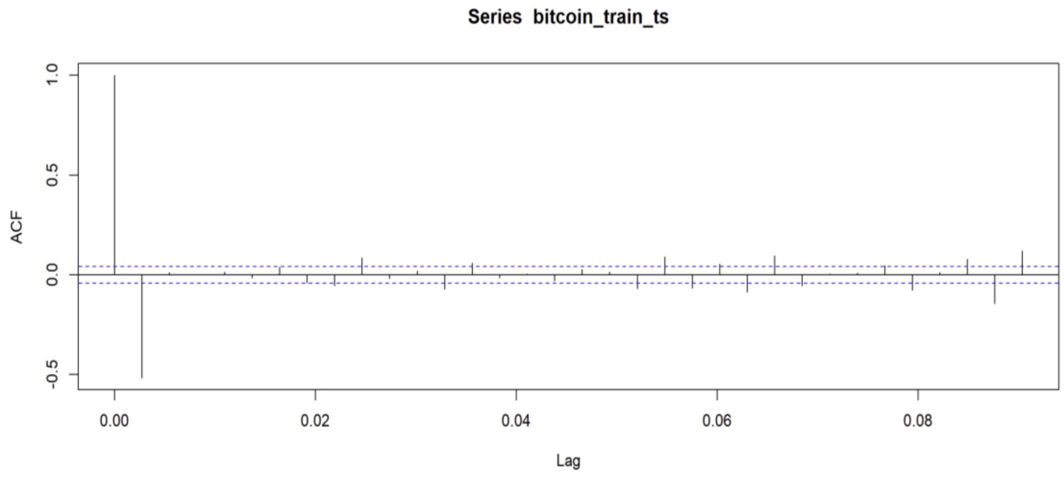

Figure 4. ACF Plot

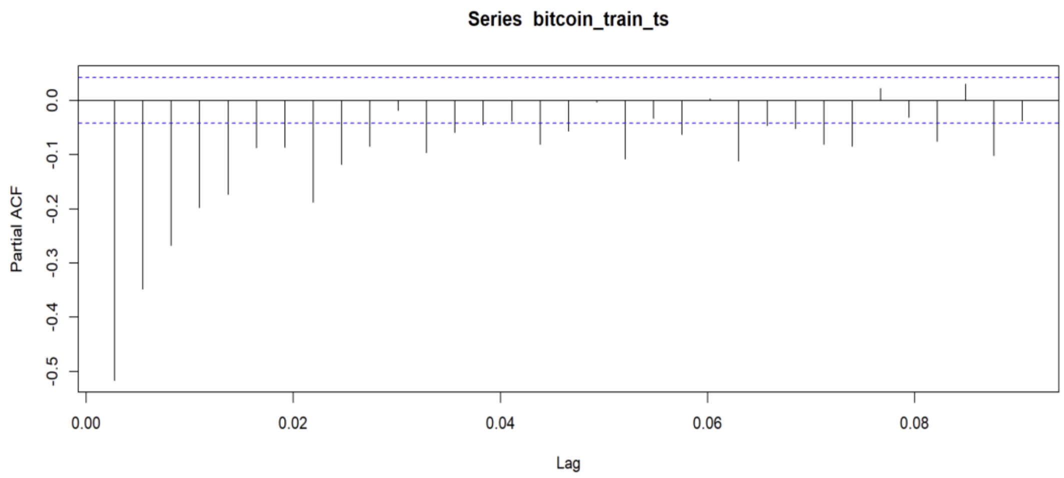

Figure 5. PACF Plot

By invoking the summary function, specific numerical values for each parameter can be obtained. The forecasting equation for ARIMA(3,2,8) is as follows:

$$y_t = -0.506y_{t-1} - 0.769y_{t-2} - 0.352y_{t-3} - 1.5\varepsilon_{t-1} + 0.74\varepsilon_{t-2} - 0.586\varepsilon_{t-3} - 0.03\varepsilon_{t-4} + 0.31\varepsilon_{t-5} + 0.037\varepsilon_{t-6} - 0.2\varepsilon_{t-7} + 0.14\varepsilon_{t-8} + 907636 \quad (7)$$

3.5. Model evaluation and diagnostic checking

The data from the test set was predicted using the ARIMA(3,2,8) model. After computation, the percentage error between the predicted data and the test data in the test set is only -0.002044555, indicating a high level of accuracy. Furthermore, the metrics MAE=409.5839, MAPE=376.2171, and RMSE=826.881 also demonstrate the excellent performance of the ARIMA(3,2,8) model. This further confirms that the model can be used to forecast the trading price of Bitcoin (Figure 6).

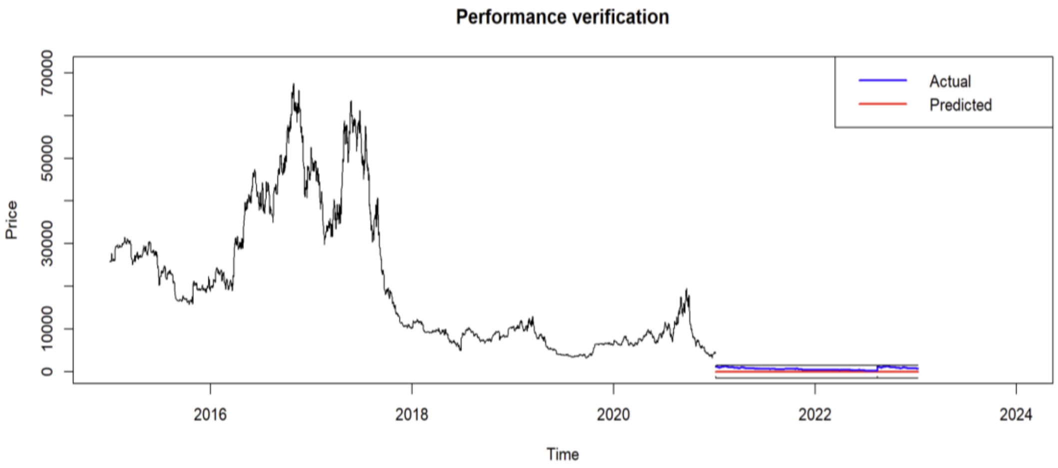

Figure 6. Performance verification

3.6. Predicting the future

Using data from ARIMA (3, 2, 8), predict the 2023/10-2024/December trading price. As shown in Figure 7, the predicted value and 95% confidence interval are indicated.

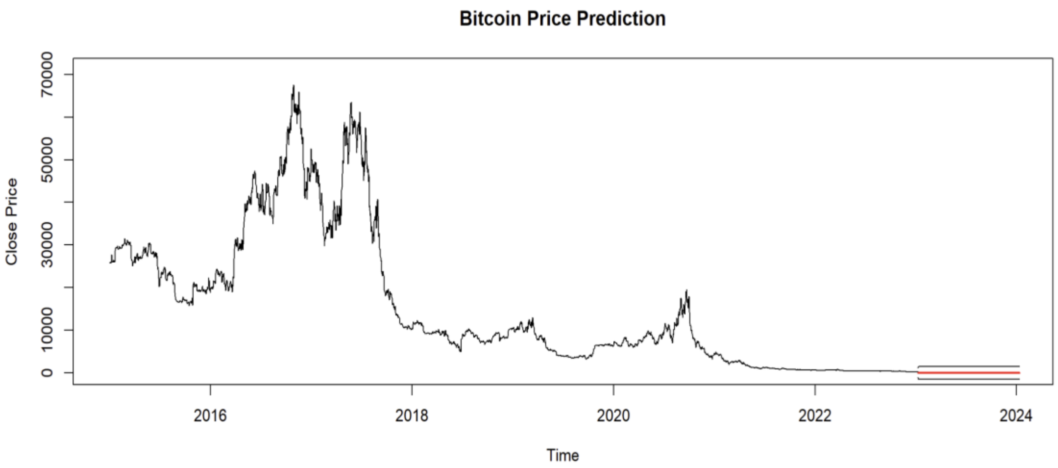

Figure 7. Bitcoin Price Prediction

4. Conclusion

This paper analyzes the prediction problem of the adjusted trading price of Bitcoin using the ARIMA model. The original series of the adjusted Bitcoin trading price exhibited non-stationarity, but after second-order differencing, the trend was eliminated and stationarity was achieved. The differenced stationary series was tested for pure randomness and found to be non-white noise, indicating that the model has statistical significance. Based on the principle of minimizing the Bayesian information criterion (BIC) and the Akaike information criterion (AIC), as well as the significance tests, the optimal fitting model was determined to be ARIMA(3, 2, 8). Performance tests demonstrated that ARIMA(3, 2, 8) exhibits excellent statistical performance. Such research fills a gap in using time series models to predict and analyze the price trends of Bitcoin in 2023. This paper provides guidance for those who are hesitant about the Bitcoin market after the epidemic.

Based on the ARIMA model, this paper successfully predicts the adjusted trading price of Bitcoin from October 2023 to October 2024. It can be understood that Bitcoin has completely lost its potential for appreciation. Despite being considered a high-quality safe-haven asset during the pandemic, this paper is pessimistic about the future prospects of Bitcoin. At least until the next mining boom, Bitcoin has become a low-quality asset and is no longer suitable for diversifying one's investment portfolio.

Therefore, the recommendation given in this paper is to reduce investments in Bitcoin. With the end of the global pandemic, the recovery of global physical industries becomes an inevitable outcome. It is advised not to increase holdings of Bitcoin in the future, but rather focus on the development of physical industries and adjust one's investment portfolio in a timely manner.

References

[1] Ariyo A A, et al. 2014 Stock Price Prediction Using the ARIMA Model. 2014 UK Sim-AMSS 16th International Conference on Computer Modelling and Simulation.

[2] Ciaian P, et al. 2015 The economics of BitCoin price formation. Applied Economics, 48(19), 1799–1815.

[3] Ji S, Kim J and Im H 2019 A Comparative Study of Bitcoin Price Prediction Using Deep Learning. Mathematics, 7(10), 898.

[4] Karasu S, et al. 2018 Prediction of Bitcoin prices with machine learning methods using time series data. 2018 26th Signal Processing and Communications Applications Conference (SIU).

[5] Kriechbaumer T, et al. 2014 An improved wavelet–ARIMA approach for forecasting metal prices. Resources Policy, 39, 32–41.

[6] Kristoufek L 2015 What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. PLOS ONE, 10(4).

[7] Nakamoto S 2008 Bitcoin: A Peer-to-Peer Electronic Cash System. Assets Pub.

[8] Rebucci A, et al. 2022 An Event Study of COVID-19 Central Bank Quantitative Easing in Advanced and Emerging Economies. Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling, Advances in Econometrics, 43, 291–322.

[9] Zhao L, Mbachu J and Liu Z 2019 Exploring the Trend of New Zealand Housing Prices to Support Sustainable Development. Sustainability, 11(9), 2482.

[10] Zhao Y, et al. 2023 The effects of quantitative easing on Bitcoin prices. Finance Research Letters, 57, 104232.