2024 SICASH
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

Development of an ARIMA Model to Predict the Monthly Price of Bitcoin in USD

Kapukotuwa, R.W.M.C.L.B.*[1], Muthuranwela, M.M.P.L[2], Samarakoon, H.G.I.L.[3], Dilshan, P.G.S.[4], Sajeewani, A.K.R.K.[5], Peiris, T. S. G.[6]

[1,2,3,4,5,6] *Sri Lanka Institute of Information Technology*

Email address of the corresponding author - *chamindukapukotuwa18@gmail.com

Abstract

This study examines the bitcoin price in USD in the world by developing a suitable time series model to identify its future trends. This data set consists of monthly bitcoin prices from August 2010 to July 2024. It was found that the original series is not stationary and not seasonality. The stationary was achieved by the first difference. Of the parsimonious models identified based on the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) of the stationary series, an auto-regressive integrated moving average (ARIMA) (2,1,2) model was identified as the best-fitted model. The significance of the model and its parameters and information criteria such as the Akaike Information Criterion (AIC), Schwarz Criterion, and log-likelihood was used to identify the best-fitted model. The model was trained using data from August 2010 to March 2024. The residuals of the model were found to be white noise. The mean absolute percentage error (MAPE) for validation data is 7.09%. The percentage errors for the validating set are all positive and varied from 3.5% to 12.9%. The predicted Bitcoin price (USD) from August to October 2024 are $59947.88, $60308.7, and $60669.53. Bitcoin price can be utilized by market demand and supply, regulatory environment, and technology development.

Keywords: ACF; ARIMA models; Bitcoin price; Forecasting; PACF; Time series analysis

Introduction

Bitcoin, the first decentralized cryptocurrency invented by an unknown person Satoshi Nakamoto, is currently the highest-valued cryptocurrency. Due to its scarcity and decentralized nature, Bitcoin is resistant to inflation and manipulation. Transactions in the peer-to-peer Bitcoin network are verified using cryptography and recorded on a publicly distributed ledger. The total supply of Bitcoin is 21 million (Bitcoin - BTC Price, Live Chart, and News | Blockchain.com, n.d). Over time, bitcoin became popular worldwide, and a prominent financial asset, and its price also began to increase with high volatility. The high volatility of bitcoin price is affected by factors like market demand, limited supply, microeconomic trends, regulatory news, and technical developments (blockchain technology) a compelling subject for time series analysis (Reiff, 2024).

According to Fig. 1 from November 2017 to December, the Bitcoin price increased dramatically from around $5,000 to $20,000. From September to December 2020, the price of Bitcoin increased by nearly $10,000 to $30,000. From April to July 2021, the bitcoin price was around $64,000 to $49,000 in the market, it reached its highest value in this period. However, it fell to around $30,000 by July 2021 because of a regulatory crackdown in China and Elon Mask's ecological concerns. According to crypto miners, there will be a collapse in bitcoin prices (Shanghai, 2021). Building a model through time series analysis about the price of bitcoin leads us to predict statistically more accurate values in relevant periods. It will be an immense opportunity for crypto miners and people interested in their decision-making process. At the same time, the price of Bitcoin has been the subject of various research. There is limited research using a time series model that accurately captures the long-term price.

The significance of the study is due to its potential contribution to both academic and real-world applications. An effective prediction model based on time series analysis could significantly enhance the decision-making process for investors and policymakers. Considering the volatility of Bitcoin price and growing influence in the global market, creating a reliable prediction model might lower investment risk and promote more stable market conditions.

Materials and Methods

Secondary Data

The data on the monthly bitcoin price in US dollars was collected from August 2010 to March 2024 including 164 observations (Bitcoin - BTC Price, 2024).

Statistical Analysis

In this study, the time series Auto Regressive Moving Average (ARIMA) model was used which contains both autoregressive (AR) and moving average (MA) parameters. This method is mostly useful for analysing and predicting future values in time series data. It is also known as The Box – Jenkins ARMA (p, q) model and is expressed as follows:

$$Y_t = \mu + \emptyset_1 Y_{t-1} + + \ldots\ldots\ldots +$$
$$\emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} - -\theta_3 e_{t-3} - \ldots\ldots\ldots - \theta_q e_{t-q}$$
----------(1)

The coefficients $\emptyset_i(i = 1,2,$ ...., p) are coefficients of the autoregressive (AR) process, $\theta_i(\ i =$ 1,2,3........., q) are coefficients of the moving average (MA) process, { e_t } is a purely random process with mean zero and constant variance and $\{Y_t\}$ is the observed time series. The p and q are the order of the Autoregressive (AR) process and the order of the moving average (MA) process respectively. Autoregressive Moving Average (ARMA) (p, q) models are used when time series are stationary, but if they are non-stationary, ARIMA (p, d, q) models are used by differencing non-stationary from stationary where d is the order of difference (Box et al, 2015).

If the series is made stationary by taking d^{th} difference of the observed series such that $Z_t = Y_t - Y_{t-1}$ then

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots \emptyset_p Z_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots \theta_q e_{t-q} + e_t + \mu$$
------(2)

Comparing the pattern of sample ACF and SPACF with the corresponding theoretical ACF and PACF, several parsimonious models are postulated. The best-fitted model among the selected parsimonious models is selected by comparing the status of the significance of the model, parameters of the model and information criteria such as AIC, SC, and log-likelihood (Box et al, 2015).

Results and Discussion

Temporal Variability

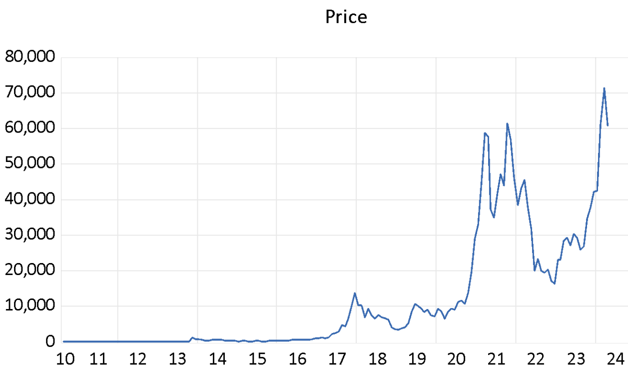


Figure 1. Temporal variability of original series

The time series plot in Figure 1 indicates the temporal variability of the monthly price of bitcoin (in USD) from August 2010 to March 2024. The monthly price fluctuated significantly, with a value of 0.1 in the year 2010 to 42272.50 in the year 2023. The average monthly price of Bitcoin is 12374.32. There has been a noticeable increase throughout time, which is reflective of growing interest in Bitcoin and its increasing usage. However, the standard deviation is 17948.94. The Augmented Dickey-Fuller (ADF) test indicates that the original series is non-stationary, with a p-value (0.9210) greater than 0.05, to ensure the stationary of the series' first difference was calculated and following ADF testing revealed that this first difference series is stationary, with a p-value (0.000) less than the significance level (0.05).

ACF and PACF of the 1st Difference Series

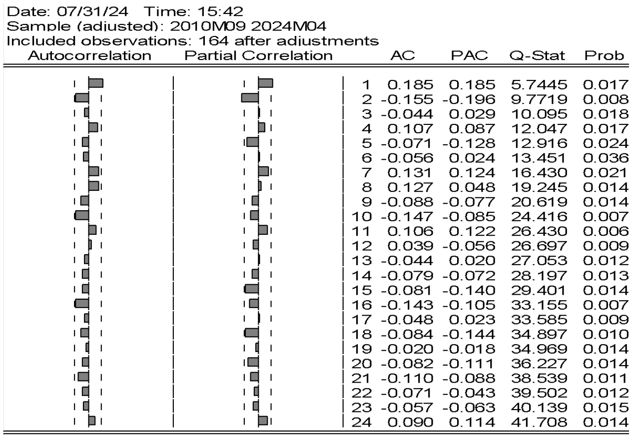


Figure 2. ACF and PACF of the First Difference Series

The first difference series of the original series' Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are shown in Figure 2. Thus, this correlogram is employed to identify orders of AR(p), and MA(q). According to the ACF, the 1st lag is significantly different from zero, so we can consider it can be MA 1, as well as the PACF, first two lags are significantly different from zero. so, it may be AR (1) or AR (2). Possible models that might fit the data include ARIMA (1,1,0), ARIMA (2,1,0), ARIMA (2,1,2), and ARIMA (0,1,2), ARIMA (1,1,1), ARIMA (2, 1, 1). These Models were the most appropriate models for the stationary series that could be obtained by identifying the first difference of the original series.

Identification of the Best-Fitted Model

Table 1. Comparison of Estimated Models

Model	MA (1)	MA (2)	AR (1)	AR (2)	AIC	Schwarz Criterion	Log Likelihood
ARIMA (1,1,0)	N/A	N/A	Significant	N/A	19.4314	19.4882	-1590.382
ARIMA (2,1,0)	N/A	N/A	Significant	Significant	19.3957	19.4713	-1586.448
ARIMA (0,1,2)	Significant	Significant	N/A	N/A	19.4126	19.4882	-1587.843
ARIMA (1,1,1)	Not Significant	N/A	Not Significant	N/A	19.452	19.527	-1620.292
ARIMA (2,1,1)	Not Significant	N/A	Not Significant	Not Significant	19.458	19.551	-1619.745
ARIMA (2,1,2)	Significant	Significant	Significant	Significant	19.1807	19.2660	-1528.456

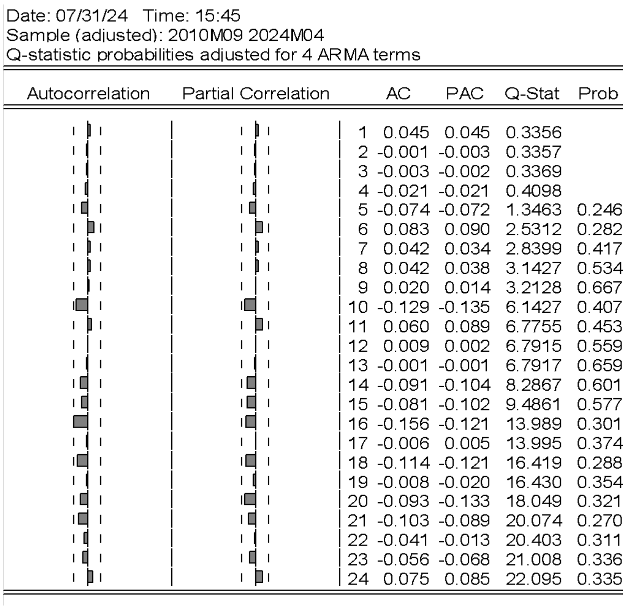
In order to reduce the heteroscedasticity of the variance log transformation was used and found that there was no improvement. Based on the significance of the parameters, lowest Akaike info criterion (AIC), Schwarz criterion (SC), and highest log-likelihood, among the four possible models shown in Table 1, the ARIMA (2,1,2) was found to be the best-fitted model. As a result, the following could be the best model fitted:

$$(1 - B)[1 - \emptyset_1 B - \emptyset_2 B^2]y_t = e_t[1 - \theta_1 B - \theta_2 B^2] + \mu$$

$$(1 - B)[1 - 0.548B - 0.77B^2]y_t = e_t[1 + 0.81B - 0.323B^2] + 360.82$$

Model Diagnostics

Correlogram of Residuals



Date: 07/31/24 Time: 15:45
Sample (adjusted): 2010M09 2024M04
Q-statistic probabilities adjusted for 4 ARMA terms

	AC	PAC	Q-Stat	Prob
1	0.045	0.045	0.3356	
2	-0.001	-0.003	0.3357	
3	-0.003	-0.002	0.3369	
4	-0.021	-0.021	0.4098	
5	-0.074	-0.072	1.3463	0.246
6	0.083	0.090	2.5312	0.282
7	0.042	0.034	2.8399	0.417
8	0.042	0.038	3.1427	0.534
9	0.020	0.014	3.2128	0.667
10	-0.129	-0.135	6.1427	0.407
11	0.060	0.089	6.7755	0.453
12	0.009	0.002	6.7915	0.559
13	-0.001	-0.001	6.7917	0.659
14	-0.091	-0.104	8.2867	0.601
15	-0.081	-0.102	9.4861	0.577
16	-0.156	-0.121	13.989	0.301
17	-0.006	0.005	13.995	0.374
18	-0.114	-0.121	16.419	0.288
19	-0.008	-0.020	16.430	0.354
20	-0.093	-0.133	18.049	0.321
21	-0.103	-0.089	20.074	0.270
22	-0.041	-0.013	20.403	0.311
23	-0.056	-0.068	21.008	0.336
24	0.075	0.085	22.095	0.335

Figure 3. Correlogr am of the Residuals of the Best-Fitted Model

Figure 3 indicates that the Q-Statistic for residual probabilities was not statistically significant (P value > 0.05), It can be concluded with 95% confidence that the errors are white noise. Jarque-Bera value is 476.26 and p value is 0.000; thus, residuals deviated from normality. Therefore, the ARIMA (2,1,2) model can be accepted.

Inverse Roots of AR/MA Polynomial

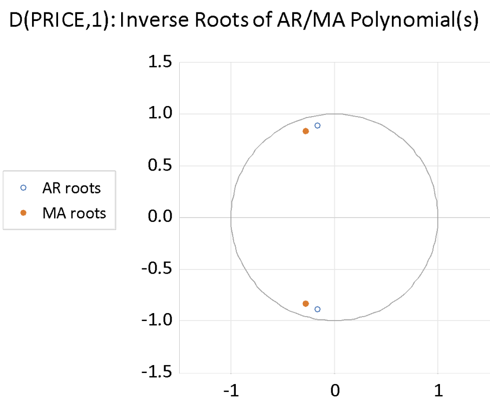


Figure 4. Inverse Roots of AR/MA Polynomial.

Comparison of Observed and Predicted Values of Training Data Set

There is a very strong positive relationship between the predicted price and the actual price of Bitcoin (USD) (r=0.974) (p-value= 0.00) which is closer to 1

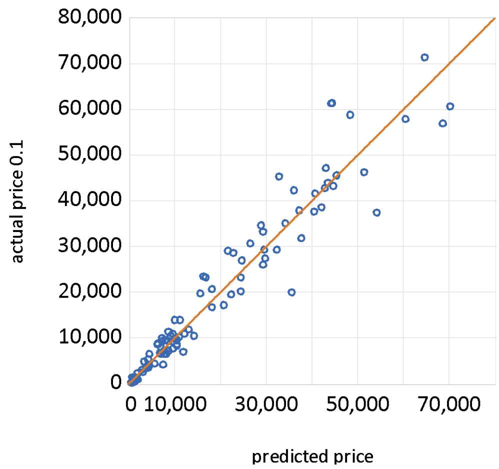


Figure 6. The Actual and Predicted Price of the Training Data Set (Aug. 2010- Apl. 2024)

Validation of the Model for the Independent Data Set

The fitted model was tested for an independent data set from April 2024 to July 2024 (Table 2). The percentage error for the validating set is positive and it varied from 3.5% to 12.9%

Table 2. Comparison of the Best-Fitted Model for an Independent Data Set

Period	Actual	Forecast	Percentage Error
4/1/2024	60660.60	58504.57	3.56%
5/1/2024	67187.00	58865.4	12.39%
6/1/2024	62754.30	59587.05	5.62%
7/1/2024	63935.10	59587.05	6.80%

The percentage error numbers vary from 3.56% to 12.39%, and the mean absolute percentage error (MAPE) is 7.09%. These results imply that the values may be used to validate the model. Furthermore, the Theil Inequality Coefficient is closer to 0 (0.388), reflecting the model's strong predictive ability (Figure 7).

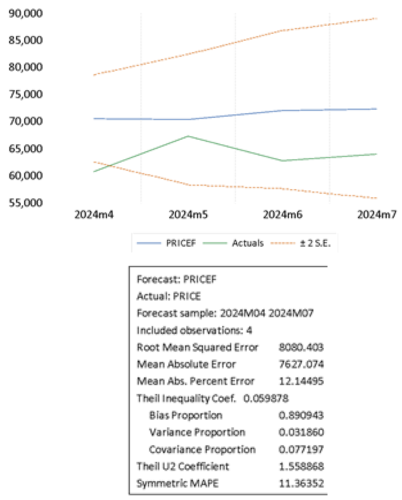


Figure 7. Plot of Actual and Predicted Price for Bitcoin Price (April 2024 to July 2024)

Short Term Prediction

The model was used to predict future values from August 2024 to October 2024 and the predicted values are shown in Table 3

Table 3. Predicted Values of the Subsequent Months

Month	8/1/2024	9/1/2024	10/1/2024
Predicted values	59947.88	60308.7	60669.53

The best fitted model predicted Bitcoin price (USD) for the next 3 months. According to these results, appropriate decisions can be made.

Conclusion

In this study, we provide an accurate model and forecast the price of Bitcoin (USD) using the Box Jenkins ARIMA approach, depending on monthly data from August 2010 to July 2024. The ARIMA (2,1,2) model was found as the best-fitted model based on the significance of the parameters and the model selection criteria including, Akaike Information Criterion (AIC), Schwarz Criterion (SC), and log-likelihood. The model's residuals were confirmed to be white noise. The ARIMA (2,1,2) model was validated with data from the first four months of 2024, showing a mean absolute percentage error (MAPE) of 7.09%, suggesting a strong predictive capability. The model is derived as $(1-\beta)[1-0.548\beta-0.77\beta^2]y_t = e_t[1+0.81\beta-0.323\beta^2]+360.82$.

The predicted Bitcoin prices for August, September, and October 2024 are $59,947.88, $60,308.7, and $60,669.53, respectively. Furthermore, Bitcoin prices may be influenced by a variety of factors, like legislation, technological developments, market attitudes, and macroeconomic causes. Predictions are therefore valuable but should be interpreted with caution.

References

Bitcoin - BTC price (2024). live chart, and news. https://www.blockchain.com/explorer/assets/BTC

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. Wiley.

Reiff, N. (2024, January 13). Why is Bitcoin volatile? Investopedia. https://www.investopedia.com/articles/investing/052014/why-bitcoins-value-so-volatile.asp

Shanghai, C. C. /. C. A. (2021, June 2). Why China is cracking down on bitcoin mining and what it could mean for other countries. *TIME*. https://time.com/6051991/why-china-is-cracking-down-on-bitcoin-mining-and-what-it-could-mean-for-other-countries/