

Using ARIMA model to analyse and predict bitcoin price

Yang Si*

Department of Business, University of New Souths Wales, Sydney, Australia

*Corresponding author: z5270579@ad.unsw.edu.au

Abstract. In this paper, Autoregressive Integrated Moving Average (ARIMA) model is used for analysing and forecasting the adjusted closing price of bitcoin. The whole dataset used is daily bitcoin closing price dating from Jan 2017 to Sep 2022. However, for testing the performance of the ARIMA model, the dataset is divided into two parts: the ARIMA model is built on training set and later using the test set to check the accuracy of the prediction. Two models, namely ARIMA (5, 2, 1) and ARIMA (0, 2, 2) are selected and comparisons between them are made. ARIMA (5, 2, 1) is chosen with stepwise selection and approximated information criteria while ARIMA (0, 2, 2) is without stepwise selection and the information criteria is not approximated. Both pass the residual test and ARIMA (0, 2, 2) is slightly better according to AIC, AICc and BIC. Later, the predicting accuracy of the two models for different forecasting periods (5-day, 10-day, and long-term forecasting) are compared. It is not surprising that ARIMA performs better while making short term prediction. The 5-day and 10-day forecast works well while the long-term forecast is of limited practical value.

Keywords: time series, ARIMA, bitcoin price

1. Introduction

The earliest and most well-known cryptocurrency is Bitcoin, which is an encrypted digital currency being viewed as a safe investment against market volatility and inflation [1]. There are no fees or requirements for centralised identification checks when opening a bitcoin account. Overall, these regulations result in a payment system that is viewed as being more flexible, private, and less vulnerable to governmental inspection than alternative payment systems [2]. Bitcoin has earned a wide following in the media, academia, and the finance industry since its introduction in 2009 and has been increasingly capturing the public's interest without signs of slowing down [1]. There has been a sharp increase in bitcoin's value, going from about \$1 in 2010 to over \$20,000 right now. The daily transaction volume of bitcoin is also highly volatile and massive: it's 330,000 in December 2020 and rise to roughly 400,000 in early January 2021, which is greater than the transaction volume of other cryptocurrencies during the same period [3]. Unlike traditional currencies, digital currencies like bitcoin has many advantages. For example, it allows faster transactions at lower costs, making small payments possible without requiring a banking system and enabling a broader reach [4]. Like the stock market, the price of bitcoin fluctuates every day. Continuing with the analogy, predicting bitcoin's value would be advantageous for its trading as forecasting stock market value aids in the execution of more lucrative deals [5]. Predicting the price of bitcoin is consequently critical for financial investing enthusiasts. Furthermore, economists are interested in the Bitcoin price since it is a virtual currency with the potential to disrupt traditional payment methods and maybe even monetary institutions. Bitcoin is an important indicator of the whole economy, has interactions with both the traditional financial system and the actual economy [6].

In this study the ARIMA model is built based on training data. Two models are selected using different criteria: one is with stepwise selection and approximated information criteria, and another is the opposite. Two models selected are ARIMA (5, 2, 1) and ARIMA (0, 2, 2) and they are further compared according to the AIC, AICc and BIC. It seems that ARIMA (0, 2, 2) is better in this case because of the lower AIC, AICc and BIC. Moreover, the accuracy of different forecasting periods within the same model are compared and it is found that 5-day forecasting and 10-day forecasting are much more accurate than the long-term prediction. Moreover, the comparison between the two models with the same forecasting period are made and it seems that when the forecasting period changes from 5-day to 10-day, the outcome will be different. These comparison outcomes may give

insights to investors who hopes to use ARIMA to predict the short term bitcoin price. Also the steps of building ARIMA models and making predictions are detailed and investors without mathematical background could benefit from this.

The following is how the paper is organised. The underlying principle of ARIMA model is introduced in section 2. Section 2 also includes the data collection and pre-processing process. Results are presented in section 3 in detail and some discussion is made. Section 4 is the conclusion part and some limitations and potential solutions are mentioned.

2. Methodology

2.1 ARIMA Model

2.1.1 Related work

ARIMA model is a widely-used statistical technique for time series forecasting. Poongodi M et al employed the ARIMA model without styles and seasons to estimate Bitcoin closure rate in 2021 and achieved 49% accuracy, which is a satisfactory outcome [7]. In 2020, Jinan Fiaidhi et al. developed an analogous project in which they employed the ARIMA approach to choose models with the lowest Mean Squared Error (MSE) [8]. According to the findings, the ARIMA Model outperformed the Neural Network Model and the results are significant when seasonality is abolished. In 2014, a study about the efficiency of time series analysis in predicting stock value in India was conducted by Mondal, Shit and Goshami [9]. They worked with data from the past twenty-three months and the ARIMA model was able to predict stock market movements with more than 85% accuracy. Although ARIMA is a forecasting approach which disregards the independent variables during predicting, this makes it appropriate for connected statistical data and requires some assumptions, such as autocorrelation and seasonal patterns [10]. The advantage of ARIMA is its independence and efficiency while processing time series financial data as well as the benefit of providing accurate short-term forecasts. In addition, it is capable of handling seasonal data variances and projecting historical data that is technically challenging to grasp [11]. In this paper, ARIMA model is chosen to make short-term predictions in Bitcoin price because of its simplicity and its effectiveness.

2.1.2 ARIMA model

ARIMA, whose full name is Autoregressive Integrated Moving Average, was created in 1970 by George Box and Gwilyn Jenkins, is a time series forecasting model which is commonly used in the finance field to predict future financial market moves. The ARIMA model was created by combining the Moving Average (MA) and Auto-Regressive (AR) models, both of which predict future values using lagged data. The ARIMA model under the specific researching scenario with order p, d, q can be denoted by the following formula:

$$BP_t = c + \alpha_1 BP_{t-1} + \dots + \alpha_p BP_{t-p} + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where BP_t represents the adjusting closing price of bitcoin when the time is at t . On the right of the equation, the "predictors" comprise both lagged BP_t values and lagged error. The terms containing the lagged values is the AR model component, which is frequently used to analyse time-varying processes since they are based on the idea that forecasting the variable using linear combinations of its past values. However, simple use of AR model may result in inaccurate predictions because they are exclusively based on historical data. The predicting power of the model can be enhanced if MA component is added. The MA model employs past forecast mistakes rather than past values forecast errors. Combining AR model with order p and MA model with order q can derive an ARIMA(p, d, q) model. The lag order is known as p, reflecting the number of observations with a lag incorporated in this model. As for d, it is the number of times of differencing needed to get a stationary time series. The order of moving average is denoted by q, representing the size of the moving average window.

2.2 Data

The dataset is collected from the Yahoo Finance website¹, containing 6 parameters describing bitcoin price on daily basis: opening price, the highest price, the lowest price, closing price, adjusted closing price and the trading volume. The time range selected is dated from Jan 2017 to Sep 2022 and the total number of observations is 2084.

Additionally, the entire dataset is separated into the training set and the test set in order to evaluate the predictive ability of the ARIMA model. The training dataset includes daily bitcoin price from the start of 2017 to the beginning of 2021. And the test set includes all the remaining data: from the beginning of 2021 to Sep 2022.

Three actions are taken to process the data: columns elimination, stationary testing and first differencing. Columns removal is used to choose or eliminate unnecessary attributes. In this research, only the adjusted closing price column is selected and will be transformed as time series data to be dealt with. Other columns are not utilised and will have no impact on the outcome. Secondly, because the ARIMA model requires stationary time series data, the Augmented Dickey-fuller test is used to assess whether or not the data is stationary, which is a kind of factual test and is good for complex time series data. The p-value can be checked to test the stationarity of the dataset. The data has a unit root and is non-stationary if the p-value is greater than 0.05, vice versa. The original dataset is tested and the p-value is seen as 0.99, indicating that the null hypothesis cannot be rejected, and the data is non-stationary. Therefore, data transformation is needed to obtain a stationary times series data because of the requirement of the ARIMA model. First differencing is developed in this case, and the same test is performed. After first differencing, the time series can pass the residual test and the data is stationary.

3. Results

3.1 Descriptive Statistics

The following plot shows the daily adjusted closing Bitcoin price in USD against time. According to Figure 1, after the 2017 Bitcoin boom, the adjusted closing price rapidly grew nearly to 20000 USD in 2018. Then it fluctuated in the next three years and the adjusted closing price at the end of 2020 was nearly the same as that of 2018. However, when it came to 2021, bitcoin coin price has a boom: the price tripled and arrived its peak at approximately 60000 USD. Although a rapid decrease followed, the price flourished again and had high volatility in the next two years. The rapid fluctuations of the price suggested a highly volatile market price of bitcoin. According to the graph, the bitcoin price generally rises to a higher level from the end of 2020 till 2022 before another decreasing trend occurs. The data plotted against the individual seasons (a year) in which the data were observed is shown as following in Figure 2.

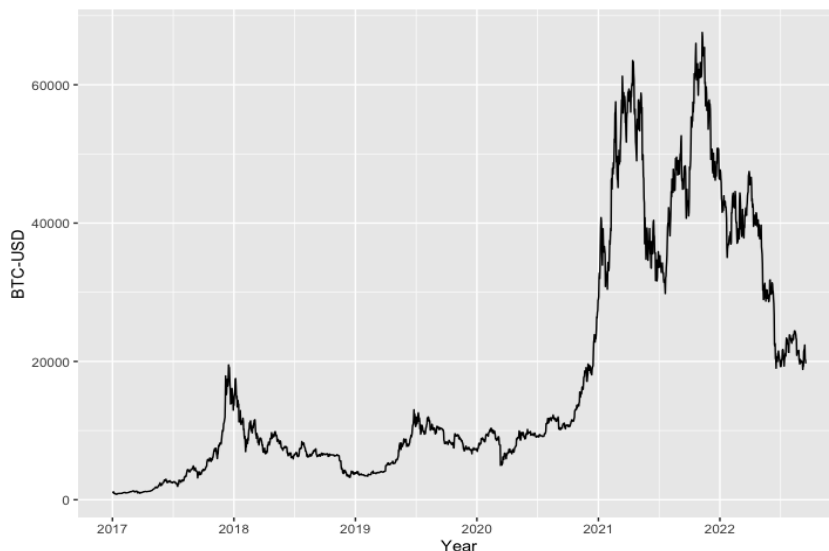


Fig. 1 Bitcoin Price in USD

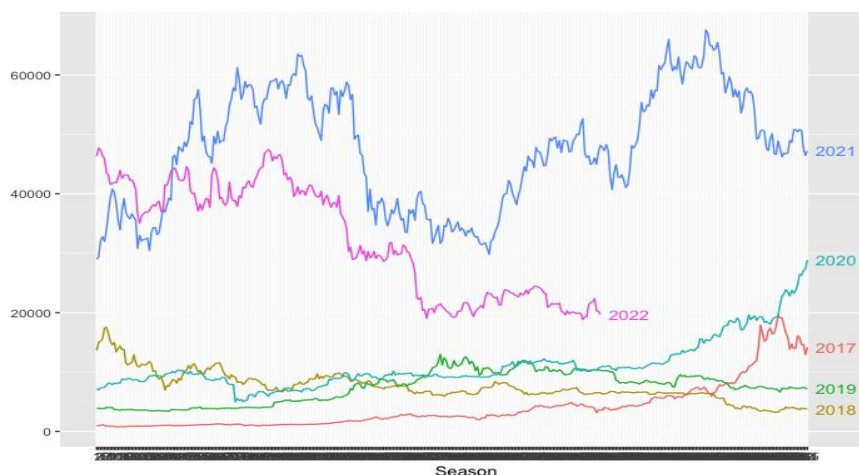


Fig. 2 Seasonal plot: Adjusted Closing price of Bitcoin

The plot makes it possible to perceive the underlying seasonal pattern more clearly and makes it simple to see any significant deviations from the seasonal trend. From the seasonal plot, there is no obvious seasonal pattern. Thus, only non-seasonal ARIMA model will be used.

3.2 Results from ARIMA Model

The first part is building ARIMA using the training set dating from 2017 to 2021 to make predictions and the second part is comparing the accuracy of the 2 selected ARIMA models. Two kinds of auto selection functions are used to fit the best ARIMA model according to their AIC, AICc or BIC value, the first one is with stepwise selection and approximated information criteria for model selection, the second one is without stepwise selection and the information criteria is not approximated. And the two models selected are ARIMA (5, 2, 1) and ARIMA (0, 2, 2) respectively. The residuals of the fitted ARIMA model is examined through the Ljung-Box test to see if the residuals contain autocorrelation. From the output, the p-value of Ljung-Box test for ARIMA (5, 2, 1) is 0.996 and for ARIMA (0, 2, 2) the p value is 0.981. Thus the null hypothesis is not rejected and the residuals from these two ARIMA models are white noise. Also, from the ACF plot of ARIMA (5, 2, 1) and ARIMA (0, 2, 2), most values are lying within the critical value limits. Therefore, the same conclusion of the residuals from the two ARIMA models are white noise can be obtained from the graphs.

Since both ARIMA models pass the residual test, the final model will be chosen based on their AIC, AICc and BIC values. The following table shows the AIC, AICc and BIC value for these two models. As the table shown below, the values of three criteria of ARIMA (0, 2, 2) are all lower than ARIMA(5, 2, 1). In this case ARIMA (0, 2, 2) will be preferred based on the three information criteria.

Table 1. AIC, AICc and BIC values for ARIMA (5, 2, 1) and ARIMA (0, 2, 2)

Model	AIC	AICc	BIC
ARIMA (5, 2, 1)	21,873.0	21,873.1	21,910.1
ARIMA (0, 2, 2)	21,871.7	21,871.7	21,887.6

And Figure 3 and Figure 4 are the plots of long run forecasts from ARIMA (5, 2, 1) and ARIMA(0, 2, 2). From the plots, the ARIMA model might work relatively well for short term predictions. Short term main trends can be captured by the ARIMA model but in the long run the accuracy will low and the predicting power is poor.

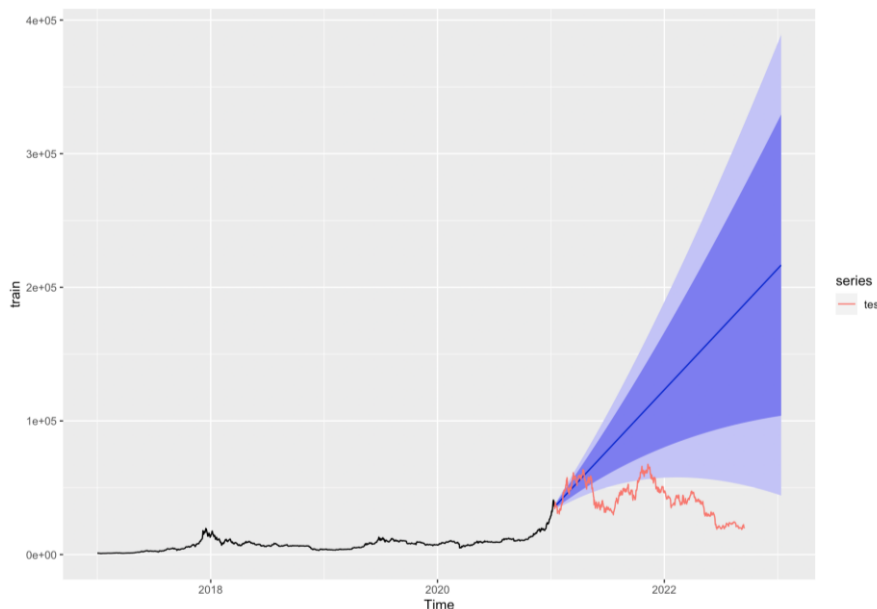


Fig. 3 Forecasts from ARIMA (5, 2, 1)

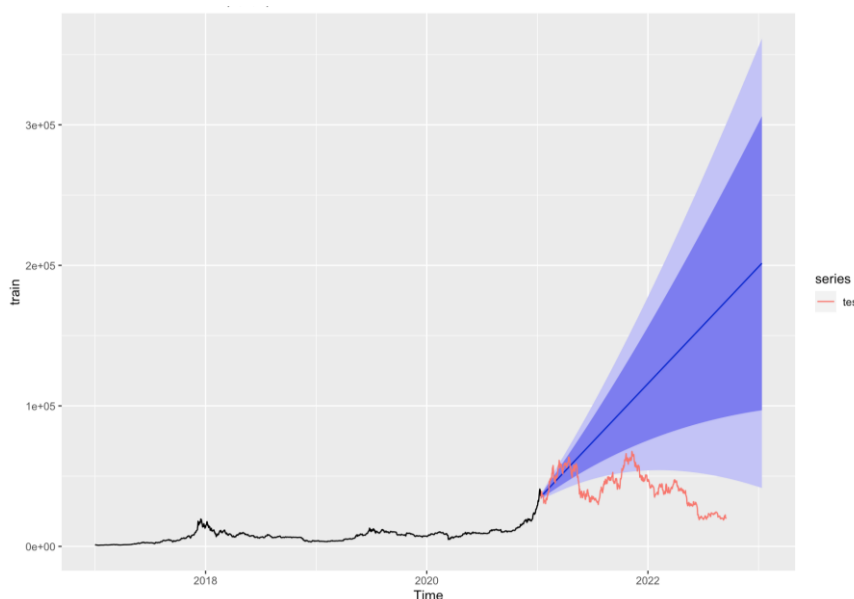


Fig. 4 Forecasts from ARIMA (0, 2, 2)

It will be hard to compare the accuracy of these two models but R can calculate it automatically. To check the predicting power of the model, training error is what should be focused rather than test error. Based on the training set, the ARIMA model is built while the test set is for testing the error of the model. The smaller the gap between the fitted values and the true values is, the better is the model. Based on the results from R, the forecasting accuracy of these two models using different forecasting periods are in the following table. It should be noted that all error in Table 1 is for test error because training error does not have influence on the predicting accuracy.

Table 2. Accuracy of ARIMA (5, 2, 1) and ARIMA (0, 2, 2) for different forecasting periods

Model	Forecasting period (h)	RMSE	MAE	MAPE	MASE
ARIMA (5, 2, 1)	Containing all test set	87,384.877	71,108.253	229.860	14.641
ARIMA (0, 2, 2)	Containing all test set	80,074.487	64,941.640	210.678	13.371
ARIMA (5, 2, 1)	5	1,610.242	1,318.954	3.563	0.272
ARIMA (0, 2, 2)	5	1,781.926	1,453.225	3.905	0.299
ARIMA (5, 2, 1)	10	2,710.253	1,927.969	5.646	0.397
ARIMA (0, 2, 2)	10	2,572.588	1,790.195	5.227	0.369

As shown in Table 2, if the whole period for the training set is considered, the RMSE and MAE will be much higher than short period. If only 5 future values are forecasted, the error will drop dramatically: RMSE for ARIMA (5, 2, 1) decrease from 87384.877 to 1610.242 and MAE drops from 71108.253 to 1318.954. Also, the MAPE and MASE for the 5-day forecasting period retains at a much lower level. If h is extended to 10, it can be observed that the errors of ARIMA (5, 2, 1) are nearly doubled compared to that of 5-day forecasting period. It is similar to analyse the condition of ARIMA (0, 2, 2). However, for a fixed forecasting period, when comparing ARIMA (5, 2, 1) and ARIMA (0, 2, 2), it can be seen that ARIMA (5, 2, 1) is better than ARIMA (0, 2, 2) when the forecasting period is 5-day, but when the test error of ARIMA (0, 2, 2) is lower. To conclude, ARIMA model perform better while making short run predictions than in the long run.

3.3 Discussion

M. Wirawan, T. Widiyaningtyas and M. M. Hasan used ARIMA model to make short term predictions of bitcoin price and evaluate the prediction accuracy using MAPE [12]. They found that ARIMA (4, 1, 4) generates high level accuracy while making one to seven days forecast ahead and as the periods of predicting increase, the accuracy level also drops. This result is similar with the outcome in this paper, which also concludes that shorter time span could give rise to higher predicting power because the 5-day and 10-day forecast are more accurate than long term prediction.

4. Conclusion

As the most popular cryptocurrency, bitcoin has drawn interest from economists, financiers, and even computer scientists. Its significant price volatility and changes make forecasting difficult but also appealing. In this research work, the using of traditional ARIMA model to make predictions of bitcoin price is investigated. The dataset is divided to two sections which are used for building the model and testing the model respectively. Two ARIMA models are selected to make predictions in both short span and long term. Comparisons between two models is through the AIC, AICc and BIC, it seems the ARIMA (0, 2, 2) is better. As for the comparisons of the accuracy, ARIMA (5, 2, 1) performs better when the forecasting period is 5-day. When the forecasting period increases to 10-day, the errors of ARIMA (0, 2, 2) is lower. Different time spans prediction accuracy is also compared among 5-day, 10-day and long-term predictions, and it is found that predictions for shorter time span is more accurate because the error estimate such as RMSE, MAE, MAPE, MASE are smaller.

To conclude, this work may benefit investors without statistical background and beginners who want to make predictions using the traditional ARIMA model. However, ARIMA model does have

many limitations because it has much poorer performance if long-term forecasting is needed. Also, the ARIMA is the basic time series model, which means the model can still be improved if more complicated methods like machine learning can be integrated into the ARIMA model. In the future, the project might be expanded and improved by combining ARIMA with a mixed machine learning technique to anticipate the bitcoin price more accurately.

References

- [1] S. K. U, S. P. Aanandhi, S. P. Akhilaa, V. Vardarajan and M. Sathiyarayanan, "Cryptocurrency Price Prediction using Time Series Forecasting (ARIMA)," 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021, pp. 598-602, doi: 10.1109/ISRITI54043.2021.9702842.
- [2] Böhme, Rainer, et al. "Bitcoin: Economics, Technology, and Governance." *The Journal of Economic Perspectives*, vol. 29, no. 2, 2015, pp. 213–38. JSTOR, <http://www.jstor.org/stable/24292130>. Accessed 3 Sep. 2022.
- [3] Chan, Stephen, et al. "A Statistical Analysis of Cryptocurrencies." *Journal of Risk and Financial Management*, vol. 10, no. 2, May 2017, p. 12. Crossref, <https://doi.org/10.3390/jrfm10020012>.
- [4] Sangeetha, R.G., Hemanth, C., Arora, N., Singh, K., Mulye, V. (2022). Bitcoin Value Prediction. In: , et al. *Hybrid Intelligent Systems. HIS 2021. Lecture Notes in Networks and Systems*, vol 420. Springer, Cham. https://doi.org/10.1007/978-3-030-96305-7_14
- [5] Pavel Ciaian, Miroslava Rajcaniova & d'Artis Kancs (2016) The economics of BitCoin price formation, *Applied Economics*, 48:19, 1799-1815, DOI: 10.1080/00036846.2015.1109038
- [6] Bjerg, O. (2016). How is Bitcoin Money? *Theory, Culture & Society*, 33(1), 53–72. <https://doi.org/10.1177/0263276415619015>
- [7] M Poongodi and Naveen Chilamkurti Vijayakumar, "Bitcoin Price Prediction using ARIMA Model", *International Journal of Web Based Communities*, vol. x, no. x, pp. 1-10, 2021.
- [8] SM Raju and Ali Mohammad Tarif, Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis, June 2020.
- [9] P. Mondal, L. Shit and S. Goswami, "Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices", *International Journal of Computer Science Engineering and Applications (IJCSEA)*, vol. 29, April 2014.
- [10] A. Qonita, A. G. Pertiwi and T. Widiyaningtyas, "Prediction of Rupiah Against Us Dollar by Using ARIMA", *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 4, pp. 746-750, September 2017.
- [11] I. M. Wirawan, T. Widiyaningtyas and M. M. Hasan, "Short Term Prediction on Bitcoin Price Using ARIMA Method," 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), 2019, pp. 260-265, doi: 10.1109/ISEMANTIC.2019.8884257.
- [12] S. Roy, S. Nanjiba and A. Chakrabarty, "Bitcoin Price Forecasting Using Time Series Analysis," 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018, pp. 1-5, doi: 10.1109/ICCITECHN.2018.8631923.