2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018

Bitcoin Price Forecasting Using Time Series Analysis

Shaily Roy, Samiha Nanjiba, Amitabha Chakrabarty
Department of Computer Science and Engineering, BRAC University
66, Mohakhali, Dhaka, Bangladesh
{shailyroy37, snanjiba}@gmail.com, amitabha@bracu.ac.bd

***Abstract*—Over the past few years, Bitcoin has been a topic of interest of many, from academic researchers to trade investors. Bitcoin is the first as well as the most popular cryptocurrency till date. Since its launch in 2009, it has become widely popular amongst various kinds of people for its trading system without the need of a third party and also due to high volatility of Bitcoin price. In this paper, we propose a suitable model that can predict the market price of Bitcoin best by applying a few statistical analysis. Our work is done on four year's bitcoin data from 2013 to 2017 based on time series approaches especially autoregressive integrated moving average (ARIMA) model and the work finally could acquire an accuracy of 90% for deciding volatility in weighted costs of bitcoin in the short run.**

Keywords—bitcoin; stationarity; trend; close price; time series analysis; ARIMA

I. Introduction

Bitcoin is a cryptocurrency which is developed as a decentralized, peer-to-peer payment system to encourage safer online transactions without the need of a third party [1]. The transactions, issuing and storing is independent of any central bank or server as it is based on a public distributive network known as blockchain [2]. It utilizes cryptographic techniques and depends on an open-source algorithm which checks decentralized exchanges and controls the formation of new Bitcoin. Regardless of the substantial vacillations of Bitcoin prices (particularly during 2013 and late 2017) and the massive growth in the capitalization of the related market, the condemnations about illicit uses and social concerns, it has still managed to draw the attention of many investors, such as China who is buying Bitcoin, seeing this as an opportunity of investments [3], as well as researchers in the scientific community to study and understand the market in order to predict the worth of Bitcoin. Most importantly, for the huge popularity of bitcoin, the end of the year 2017 has been the time when the price has increased most noticeably which was worth to 1600 US dollar for 1 bitcoin [4]. Therefore, the analysis of financial data for predicting the future bitcoin price has always been an important field of research with a direct and indirect effect on world economy.

In this paper, we have used time series analysis in order to finding out the pattern of bitcoin price movement and forecasting the closing price of the next few days as well as analyzing the performance of the time series models.

II. Related Work

A large number of research have been conducted on the fluctuation of Bitcoin price using Time Series Analysis. Stenqvist and Lönnö have predicted Bitcoin price using twitter sentiment data [5]. They have worked on Valence Aware Dictionary and sEntiment Reasoner (VADER) which is a combined lexicon based approach to measure the individual tweet sentiment intensity and then grouped the scores into time-series. They have proposed a prediction model and evaluated aggregating tweet sentiments over a 30 minute period along with 4 shifts forward with a constant change in sentiment resulted a 79% of accuracy. Matta et al. has researched on interrelation between Bitcoin price and social activities like tweets and Google Trend and discovered a weak correlation. However he has worked with 60 days' data to complete the research. Kim predicted financial price movement using Linear SVM and Neural Network [7]. The result of his experiment interprets that the value of upper bound and the kernel parameter has a sensitive role in performance of Support Vector Machine's prediction. The best prediction performance of SVM according to this study is 64.75%. Mondal, Shit and Goshami have conducted a study on effectiveness of time series analysis basically ARIMA model for predicting stock market price in India. [8] They have worked with past twenty-three months data. Their models can predict the stock market movement with an accuracy of above 85%.

Our paper is mainly inspired by the work of [9]. They have detected the determinants of Bitcoin price fluctuation using time series analysis and created relation between fundamental economic and technical factors with Bitcoin price volatility. They have mentioned that Bitcoin is kind of exchange medium which should be valued based on its supply and demand curve interaction. They have explained it with Fisher's equation, MV=PT, where M is value of money, V is velocity of money, P is price level and T is the size of underlying economy. The supply of Bitcoin is given by M=P^BB (where P^B is the cost of Bitcoin and B is the supply of Bitcoin available for use), in this way inferring P^B = PT/VB. In that case, the balance cost of Bitcoin ought to be emphatically identified with the general value level (P) and the measurement of the Bitcoin economy (T) however contrarily identified with the aggregate load of Bitcoin available for use (B). The mean accuracy of their work is almost 89.6%.

978-1-5386-9242-4/18/$31.00 ©2018 IEEE

III. METHODOLOGY

In order to fit ARIMA models to any time series data, the most important condition is that the dataset has to be consistent. In this paper, we have mainly focused on to create a consistent time series dataset and then predict the future Bitcoin closing price according to the nature of previous data. We collected the dataset from coindesk [10] which contains daily market capital, volume in transactions, opening and closing price of bitcoin in USD from July 2013 to August 2017.

A. Data Visualization

Fig. 1. Closing Price of Bitcoin

According to Fig 1, the year 2017 has been the year of Bitcoin boom. It is reported that the price in USD has almost doubled to about 950 USD at the beginning of the year 2016 and later. While towards the mid of 2017, the price was eventually tripled reaching its peak to about 2700 USD. However stating of 2017, the price was gradually being started to fall as shown with the orange line in the graph. This up and down of closing price proves the volatility of the market price of Bitcoin and can be visualized by the scattered graph of Figure 1.

B. Data Stationarization

To check the stationarity, Augmented Dickey-fuller test has been used based on the nature of our dataset. The Augmented Dickey Fuller test is a sort of factual test called a unit root test and is good for larger and complicated time series data. The instinct behind a unit root test is that it has decided how unequivocally a time series was characterized by a pattern. There are number of unit root tests and ADF (Augmented Dickey-Fuller Test) is a standout amongst the most generally utilized [11]. Dickey-Fuller test includes an intercept and time trend.

- **p value > 0.05:** Accepts the Null Hypothesis (H0), the data has a unit root and is non-stationary.
- **p value < = 0.05:** Rejects the Null Hypothesis (H0), the data is stationary. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

In the beginning, we have tried to measure the p value from Dickey fuller test which has resulted in 0.999060. Therefore, we have assumed that the data we have used is not stationary as value of p is greater than 0.05. In order to making the data stationary, we have used Logarithmic (LOG) transformation and differencing method. Log-transformed data follows a normal or near normal `distribution. It is used to unskew highly skewed data. [12]

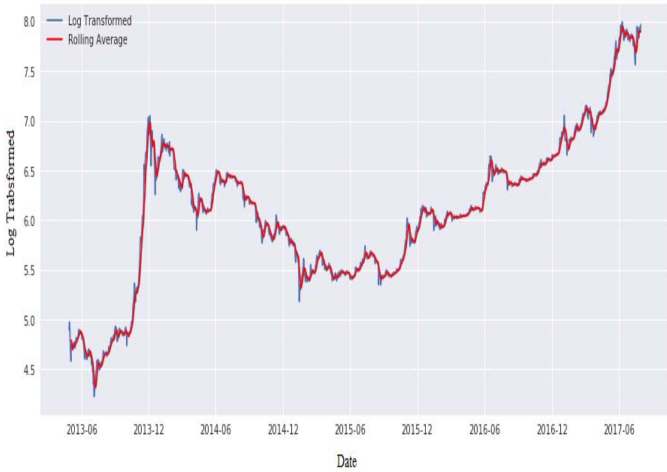

Fig. 2. Log Transformation

Fig 2. shows the graph of Log transformed price of Bitcoin which has deleted the trend from dataset. Here, graph of Bitcoin price has been upper bounded and the prices are converted to lower values. By decreasing the output values, we could decrease the difference between the weighted prices which increased consistsency of the dataset. We have been taking average of last seven days' values to measure the rolling mean. Therefore, initially there were null values in the first seven days' data. To delete the null values, we have measured rolling average of these value from the next seven days' data and formed the log transformed series. Then we have measured p value of the dataset using dickey fuller test again and have got 0.821907 which is still not a stationary value. Then we have used differencing method which is basically performed by subtracting the previous observation from the current observation. Seasonal segments in a Time series can be eliminated by utilizing differencing. Differencing is performed by subtracting the previous observation from the current observation of the dataset. After differencing the dataset, the value of p became 0.0000 which represents a stationary value. Thus our dataset has been stationarized.

C. Autocorrelation

Autocorrelation is a measurement of the inter connection inside a time series. It is a method for estimating and clarifying interior relationship between perceptions in a time series analysis [13]. According to the concept of auto-correlation, if the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one. This function is formed by autocorrelation coefficient:

$$R_h = C_h / C_0 \text{.................................[i]}$$

Where C_h is the auto covariance function, C_0 is the variance function and R_h is between minus 1 to plus 1.

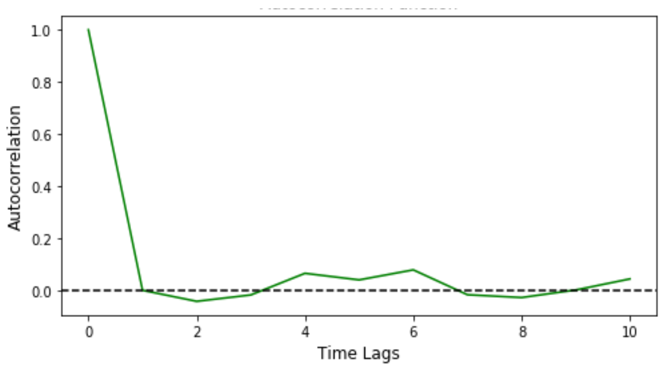

Fig. 3. Autocorrelation

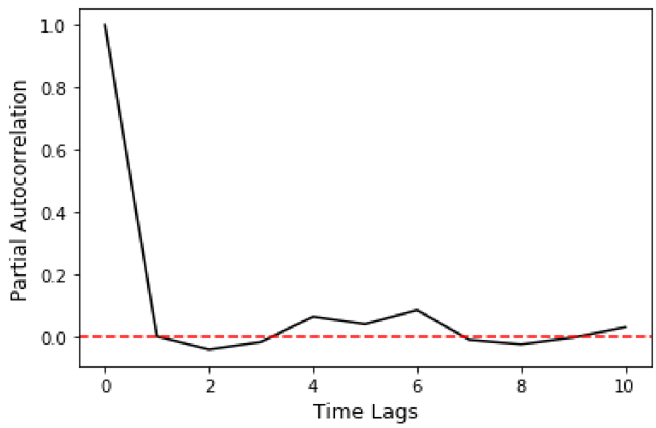

Fig. 4. Partial Autocorrelation

Autocorrelation function (ACF) helps to determine the order of moving average (MA) model in the dataset. Starting from 0, the lag after which the ACF stops crossing the significance bound (red dashed line), is the order of the MA model. If the ACF does not cross the significance bound in the first lag, but does so in case of later lags, then we assume that order of MA is 0. From the plot Figure 3, we can see that the ACF at lag 1 has crossed the significance bound. Therefore, we can assume that the series has an MA (1) process in it. Partial autocorrelation function (PACF) is used to identify order of AR in a time series model. In the Figure 4, red dotted line is the significance bound for AR model. If the PACF cross the significance line in the second lag, the order of AR will be 1. According to Fig 4, the PACF crosses the significant bound after the 1st lag. Therefore, the order of AR is 1 here.

In our paper, autocorrelation has been used for checking randomness in the data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. Calculating the order of parameters, autocorrelation helps to determine the optimal solution for a particular dataset.

D. Model Selection

We have applied time series models Autoregressive integrated moving average model (ARIMA), Autoregressive model (AR) and Moving average model (MA) in our processed dataset and plotted the resultant graph. Based on the accuracy of the models, we have chosen ARIMA to predict Bitcoin price as our data fitted well in it.

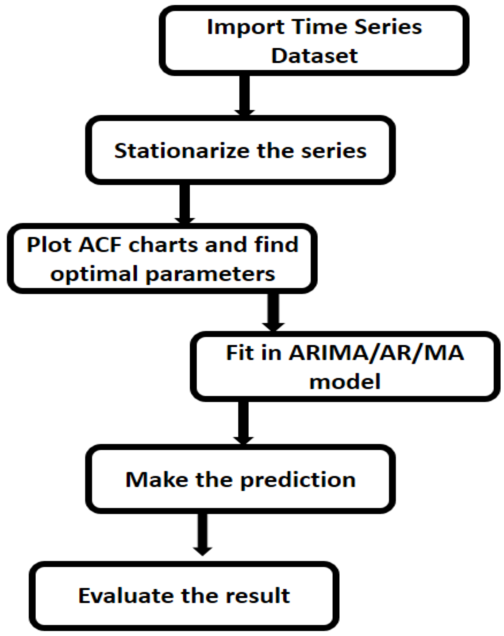

Fig. 5. Proposed Model of Time Series Analysis

Fig. 5 indicates the data flow of our time series model. Historical data is collected and stationarized. According to autocorrelation and partial auto-correlation graphs, randomness due to time lags is determined and the dataset has been fit in ARIMA/AR/MA model with all available features. Then our model has correlated the day wise closing price with other features such as market capital and volume in transactions which are in the dataset and has found out the pattern of forecasting method that fit more precisely. Then the prediction models make a prediction of next consecutive 10days' bitcoin price and user evaluates the result with the actual price of bitcoin that has been previously stored in CSV file. After calculating the accuracy user is able to find the best model for price prediction of Bitcoin for the given dataset.

IV. EVALUATION

Evaluation enables us to test the model against the information that has never been utilized for the training. We have tried to use several different models and compare their results in this paper. These results were obtained using the following hardware: 4-core CPU, 16 GB RAM and by fitting each model ten times with different random states. For calculating the accuracy we used normalized root mean squared error (RMSE) method.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(y_0 - y_t)^2}{T}} \quad \text{[ii]}$$

where, T is total number of predicted data, y_0 is original value of bitcoin price and y_t indicates the predicted value [14]. We have normalized the RMSE value and calculated the percentage error from it. We have chosen time series analysis for two technical characteristics of financial data [15].

1. **Price Moves in Trends**

The price movements follow a pattern and when such pattern has been established, financial data most likely follows the same direction to increase.

2. **History Tends to Repeat Itself**

The history of value movement in the previous tends to recap itself in the present as participants tend to respond in a steady manner for similar incentives constantly.

We have analyzed the price of Bitcoin with respect to the US Dollar using some of the popular time-series models Autoregressive integrated moving average model (ARIMA), Autoregressive model (AR) and Moving Average model (MA) and then forecasted the Bitcoin price in USD for the next consecutive 10 days. As we have dataset till 10th august and bitcoin price has been quite unstable at that time, we have chosen to predict the 10days's bitcoin price and have kept the remained unstable data to fit our model for better prediction with an appropriate accuracy.

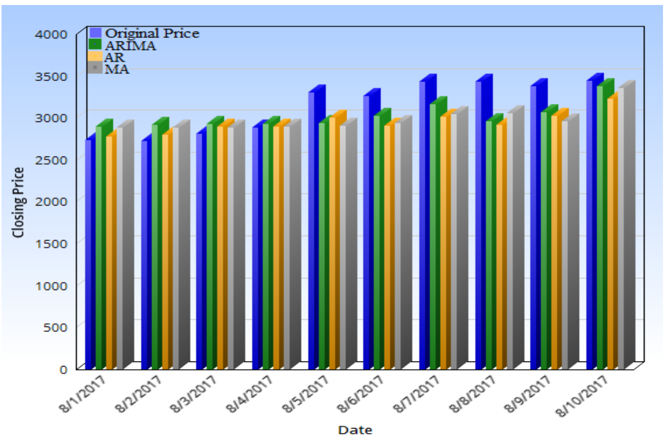

Fig. 6. Result on next 10 days bitcoin data of ARIMA, AR and MA

Fig. 6 shows the results of time series models where we have plotted both actual bitcoin price and also predicted bitcoin price. After analyzing the result we have observed that, in case of MA, many a times we are provided with a dataset, in which the prices of bitcoin increased or decreased sharply some time periods ago. Hence, mean of all the previous data is used so that it can forecast the price of the next day somewhere similar to the average of all the past days. We have wanted to use prices of the initial period which will highly affect the forecast for the next period. Therefore as an improvement over simple average, we have taken the average of the prices for last few time periods in case of MA model. For applying auto-regressive model (AR), it is needed to compare value with the previous values from a same time series data with a regression model. As AR model follows almost same theory of MA which is comparing with previous prices to predict the future bitcoin price and additionally uses regression to fit the data in the model, it has given a little bit better accuracy than MA model. Lastly, ARIMA model is basically combination of MA and AR model. This actually uses the result of the auto-correlation charts where it finds out the order or AR and MA and then fits the data integrating MA and AR observations. Therefore it results in better accuracy than the others.

The price difference between the original value and the predicted values can be visualized from the dotted line graph of Fig 7. From the graph, it can be noticed that the bitcoin has been increased quite fast compared to the previous data from 5th August. Therefore, our models could not follow the pattern which have created a bit difference between the original and the predicted prices. Behind the inconsistency of the dataset, popularity of bitcoin is the most important factor. As bitcoin price was increasing gradually from the mid of 2017, investors found it as an opportunity to invest more on it and this interest caused highest number of transactions of bitcoin in the August 2017. This highly increasing number of transactions was the reason behind the faster price growing of bitcoin. Initially our models have predicted the price observing the mean of previous data which was less than the actual price and after a certain period of observation our model could identify the increasing pattern from 9th August 2017 as shown with the red, orange and ash colored lines.

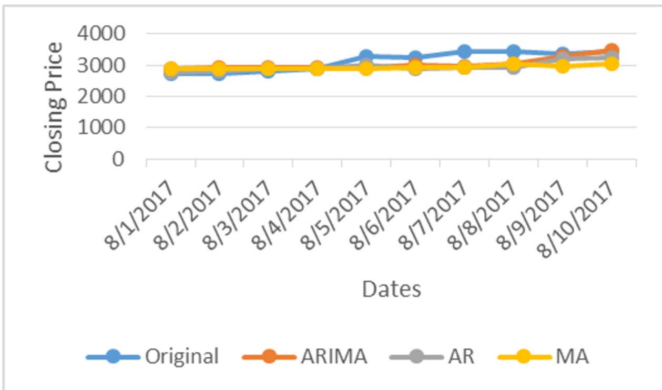

Fig. 7. Original Bitcoin price vs predicted bitcoin price of ARIMA, AR and MA

Table I shows the corresponding percentage of accuracy that we had obtained from our prediction models. Using the ARIMA model, the Bitcoin price is predicted with an accuracy of 90.31%, whereas with AR and MA, the accuracy was 89.24% and 87.58% respectively. We can see that the best result was obtained by using the Autoregressive integrated moving average model (ARIMA).

TABLE I. COMPARISON OF ARIMA, AR AND MA

Prediction Model	Accuracy
ARIMA	90.31%
AR	89.25%
MA	87.58%

Fig 8 represents the comparison bar chart of prediction algorithms where green, red and yellow indicate ARIMA, AR and MA respectively. Though the bar charts seems to be same, there is differences between the accuracy of these 3 prediction models which can be noticed in the values.

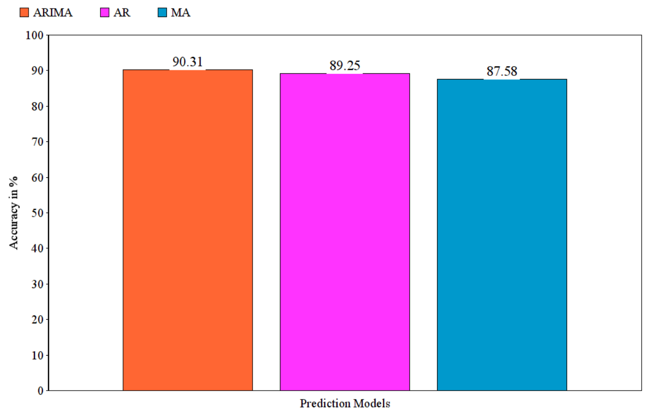

Fig. 8. Comparison barcharts between prediction Time Series Algorithms

V. CONCLUSION AND FUTURE WORK

In our research, we have contributed in financial market area by enabling the investors to figure out how to dissect Bitcoin information and furthermore to utilize that learning to predict the future bitcoin price movement. This study represents consecutive next 10 days' bitcoin price forecasting method using different statistical time series analysis such as AR, MA and ARIMA model. After the analysis, finally we have found that MA model has the lowest accuracy of 87.58% and performance of ARIMA was better with an exactness of 90.31% which is also better than the work of Georgoula and Pournarkis [9] that we have followed.

In this paper, 10 consecutive days' bitcoin prices are predicted. In the future, we will try to predict the bitcoin price of random date using the same models. We intend to collect live data and process them in real time to give outputs to the users. Also, we plan to use more parameters to predict the price of bitcoin more precisely and incorporate the time series model with some machine learning algorithms. Furthermore, if we had minutely data in our hands, we could have been able to fit models more efficiently to the data and thus we could have gotten better predictions.

ACKNOWLEDGMENT

We will like to thank Dr. Mahbub Alam Majumdar, Professor, BRAC University for helping us to work with time series analysis.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] "Bitcoin Mining," *Steemit.com* [Online]. Available at: https://steemit.com/Bitcoin/@tturpin/Bitcoin-mining [Accessed: Jul. 19, 2018]

[3] S. Chan, J. Chu, S. Nadarajah, and J. Osterrieder, "A Statistical Analysis of Cryptocurrencies," Journal of Risk and Financial Management, vol. 10, no. 2, pp. 12, May 2017 [Online]. Available at: http://dx.doi.org/10.3390/jrfm10020012

[4] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, C. H. Kim, "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies," PLoS ONE, vol. 11, no. 2, August 2016. [Online] Available at: https://doi.org/10.1371/journal.pone.0161197 [Accessed: July 19, 2005].

[5] E. Stenqvist, J. Lönnö, "Predicting Bitcoin price fluctuation with Twitter sentiment analysis," unpublished.

[6] M. Matta, I. Lunesu, M. Marchesi, "Bitcoin spread prediction using social and web search media," *Proceedings of DeCAT*, 2015.

[7] K. Kim, "Financial time series forecasting using support vector machines", *Neurocomputing* , vol. 55, no. 12, pp. 307319, 2003.

[8] 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), 29, April 2014.

[9] I. Georgoula, D. Pournarakis, C. Bilanakos, N. D. Sotiropoulos, and G. M. Giaglis, "Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices," 2015 [Online] Available at: https://www.researchgate.net/publication/290428532_Using_Time-Series_and_Sentiment_Analysis_to_Detect_the_Determinants_of_Bitcoin_Prices [Accessed: July 19, 2018].

[10] M. Shen, W. Zhao and N. De, "CoinDesk - Leader in blockchain news," *CoinDesk.com*, 2018 [Online]. Available at: https://www.coindesk.com/ [Accessed: 13 Jul. 2018].

[11] E. Kang, "Time Series: Check Stationarity," *Medium*, August 26, 2017 [Online] Available at: https://medium.com/@kangeugine/time-series-check-stationarity-1bee9085da05 [Accessed: July 14, 2018].

[12] Changyong FENG, X. "Log-transformation and its implications for data analysis." *PubMed Central (PMC),* 2018 [onine] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120293/ [Accessed 29 Aug. 2018].

[13] "Autocorrelation Plot," Itl.nist.gov, vol. 1.3, no. 3.1, 2018 [Online] Available at: https://www.itl.nist.gov/div898/handbook/eda/section3/ autocopl.htm [Accessed 21 Jul. 2018].

[14] GIS Geography. "How to Calculate Root Mean Square Error (RMSE) in Excel" *GIS Geography,* 2018 [online] Available at: https://gisgeography.com/root-mean-square-error-rmse-gis/ [Accessed 29 Aug. 2018].

[15] Skwarek, S. "Technical Analysis: Three Premises." *CMT Association,* 2018 [online] Available at: https://cmtassociation.org/kb/technical-analysis-three-premises/ [Accessed 28 Aug. 2018].