University of Sri Jayewardenepura
Department of Statistics

# Bitcoin Price Prediction and Forecast Models for Daily, Weekly and Monthly

**Group Number: 7**

**Authors:**

| | |
|---|---|
| H.M.T.D. Bandara | AS2019312 |
| K.T. Dahanayake | AS2019324 |
| A.G.E.D. De Silva | AS2019329 |
| W.H. Devindi | AS2019337 |
| S.R.R. Gomas | AS2019359 |
| W.A.D.R. Jayarathne | AS2019387 |
| M.M.S.B. Munasinghe | AS2019455 |

# Bitcoin Price Prediction and Forecast Models for Daily, Weekly and Monthly

Presented to: Dr. Neluka Devpura

Department of Statistics

Prepared by:

Group 7

STA 351 2.0

Research Methodology

Third Year Group Research

December 9th, 2022

# ACKNOWLEDGEMENT

We would like to express our sincere gratitude to senior lecturer Dr. Neluka Devpura for the guidance and supervision. It would never have been possible for us to take the research to the expected standards without her continuous support and engagement.

We would also like to thank the senior lecturer Dr. Hasanthi Pathberiya for giving us additional knowledge on time series analysis and helping us to sort out the problems when using the Eviews software.

Finally, our thanks and appreciation also go to all the team members in developing the project and all those who encouraged and supported us in completing this research project.

# Contents

# 1. Abstract

In the modern world Cryptocurrency is the trending virtual money transaction method based on Blockchain technology. There are several numbers of cryptocurrencies are currently used and among them Bitcoin, Ethereum, Tether, BNB, USD Coin, XRP are more talked about. Over the past few years, Bitcoin is the most popular and expensive cryptocurrency in economics and finance. This dissertation aims to discuss the prediction of the Bitcoin price daily, weekly, and monthly. Studies regarding this area are based on Machine Learning techniques, traditional time series models in forecasting and double exponential smoothing methods. According to recent research, Machine Learning and Deep Learning techniques give the most accurate results and conclusions.

In our study we use the Auto Regressive Integrated Moving Average (ARIMA) which is a traditional time series method to obtain the best model. For this, historical data was collected from 17$^{th}$ September 2014 to 31$^{st}$ of December 2019 on Yahoo Finance website and made predictions for 31$^{st}$ of January 2023. EViews was used as the statistical tool to obtain the above best models. Using this tool ARIMA (14,1,14), ARIMA (26,1,26), ARIMA (0,1,8) models were selected as the best models to forecast Bitcoin prices for daily, weekly, and monthly respectively by considering the lowest AIC and the highest number of significant coefficients.

ARIMA does not capture high fluctuating behaviors in time series, unlike Machine Learning and Deep Learning. The final goal of this study is to improve your knowledge regarding Bitcoin and give a best prediction to variation of Bitcoin price for cryptocurrency users.

Keywords: Cryptocurrency, Virtual, Blockchain, Bitcoin, Ethereum, Tether, BNB, USD Coin, XRP, Double Exponential Smoothing Method, Machine Learning, Deep Learning, Time Series Analysis, ARIMA, EViews

## 2. Introduction

Cryptocurrency is a digital currency, which is utilized in the modern world. At present, the demand for various cryptocurrencies is rapidly increasing due to secure transactions sided with cryptography. Users can exchange value digitally without the intervention of a third party. It is based on the theory of solving encryption algorithms to generate unique hash values with a finite number of possibilities. Users can exchange those values as if they were exchanging physical currency when combined with a network of computers that verifies transactions [1]. There are no governmental, central authority, or financial institutions because of the decentralized network system. All cryptocurrencies are based on Blockchain technology, which is complex and stores data in a way that makes it resistant to unauthorized access or illegal modification. Blockchain mainly differs from typical databases as it stores data in units called blocks each connected with cryptography.

There are different types of cryptocurrencies including Bitcoin, Ethereum, Tether, BNB, USD Coin, XRP etc. Bitcoin is the most popular and well-known among them. It was first introduced by an unknown individual or group using the name Satoshi Nakamoto in 2009 [2]. Because of its inherent power as a digital currency, Bitcoin has been seen to increase in its status over time. Until the maximum number of Bitcoins is reached, Bitcoin will be produced with diminishing returns every 4 years [3]. According to Satoshi Nakamoto, the maximum number of Bitcoins that can be produced is 21 million [2]. The main purpose of our study is to provide insights on how Bitcoin prices will fluctuate in the future. This is a time based data set because of that we use time series analysis to forecast. Time series forecasting occurs when you make scientific predictions based on historical time stamped data.

In this study, we derive the best model for forecasting Bitcoin prices using daily, weekly, and monthly data. For analysis, we use the ARIMA model and select the best type of ARIMA model from the daily, weekly, and monthly to estimate future values. ARIMA is a statistical analysis model that makes use of time series data to perform forecasts or to better comprehend the current data set. According to Feng and Palomar [4], stationarity represents a time series time-invariant behavior and is considerably simpler to model, estimate, and analyze. It is an important property for time series analysis. Since market prices are inherently non-stationary, it is necessary to ensure stationarity by differencing the time series before the forecast can be performed [5]. The ARIMA model is simple but nonetheless powerful and it aims to describe autocorrelations in time series data [5][6]. The stationarity of a time series is a crucial assumption in ARIMA modeling. The ARIMA model is straightforward but effective.

Fundamentally, the future value of a variable is determined by combining its past observations (time lags) and past error terms in a linear fashion. Because, they show a tendency in values to relate to earlier versions of themselves. Lags are a particularly helpful tool in time series analysis. A representation of the ARIMA model is the ARIMA (p,d,q). ARIMA model is the combination of the Autoregressive (AR) model and the Moving Average (MA) model with the stationarity.

ARIMA (p,d,q) means,

p is the number of lags of observations.

q is the number of lags of error term.

d is the number of times that the series was difference to obtain stationarity.

Model equation of ARIMA is,

$$Y_t' = C + \alpha_1 Y_{t-1}' + \alpha_2 Y_{t-2}' + \cdots + \alpha_p Y_{t-p}' + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots + \beta_q Z_{t-q}$$

Where the $Y_t'$ is the differenced series. It may have been differenced more than once. For instance, ARIMA(p,d,q) for $Y_t'$ is similar as ARMA(p,q) for $Y_t'$. Here, $Y_t' = \Delta^d Y_t'$ ( $d^{th}$ differenced time series).

Nowadays Bitcoin is the most popular and the largest cryptocurrency in the world. As a result of the global pandemic COVID-19, Bitcoin users quickly increased. Therefore, forecasting the price of Bitcoin is essential for everyone who uses it and who likes to transact with Bitcoin in future. This is especially important because of the popularity of Bitcoin and the volatility of its price.

## 2.1 Objectives

This study mainly focuses to predict and forecast Bitcoin prices obtained daily, weekly and monthly using time series ARIMA. To achieve this, we carried out the research work over the following which has simplified our aim.

1. Identify the best-fitted model to forecast daily, weekly, and monthly Bitcoin prices
2. Identify suitable statistical tools and techniques to build forecast models
3. Select the most feasible approaches to obtain best-fitted model
4. Evaluate whether the developed models can make accurate predictions
5. Provide a basis for managerial decisions in finance according to statistical time series analysis

## 3. Literature Review

Bitcoin has become a major attraction for investors across the globe. Thus, many scientific and financial studies have been conducted to predict the future price of Bitcoin. Such predictions of a volatile cryptocurrency can provide insights for the investors, to plan their investments in a way that reduces possible losses. Many studies which have focused on predicting future prices of cryptocurrency mainly focus on modeling techniques such as ARIMA, ML, or DL Algorithms. Most frequently, studies have analytically compared ARIMA against ML approaches for the accuracy of forecasts. The selection of models and ML techniques differs due to having various aspects and objectives considered in existing studies.

Using the Brown's Double Exponential Smoothing method, Liantoni and Agusti [7] in 2020, predicted Bitcoin price on 1st of January 2020 based on the model which gave the lowest Mean Absolute Percentage Error (MAPE). As results suggested they have concluded that 2.89% of MAPE was generated at alpha equals to 0.9.

According to a study conducted in 2019, Joao Fillipe and Batista Mendes [8] carried out a comparison between ARIMA and LSTM forecasts. They compared those Bitcoin price predicting models in terms of accuracy. Mainly they have taken RMSE, MAE into consideration when comparing. It was concluded that using LSTM, one can reduce these model evaluation measures up to some significant percentages.

In the study of Vidyulantha et al. [9], they have obtained the dataset of Bitcoin prices in USD from July 2015 to February 2020 from CoinDesk. Both ARIMA and LR models were used under their methodology. They have concluded that ARIMA models perform better than LR. However, they have further emphasized that Deep Learning Frameworks such as LSTM, RNN, and CNN are more effective in predicting cryptocurrency prices than ML and Time Series Models.

Jinan et al. [10] have fitted ARIMA (2,1,0), ARIMA (0,1,18), and ARIMA (8,1,0) respectively for 5-month Bitcoin prices in 2019 for the training dataset. The main focus of this research was to find ARIMA-composed model parameters minimizing the prediction's Mean Squared Error. The best out of the three models, which is ARIMA (8,1,0) was then tested against Bitcoin prices in 2020. The study especially gives a note that the work can be additionally improved with CNN or LSTM mixed ARIMA.

The study of Olufunke et al. [11] has conducted predictions on Bitcoin prices using ARIMA. Closing Bitcoin prices (in US Dollars) from January 2016 to May 2021 have been obtained via the CoinDesk platform. After that, they have performed stationarity inspection, applying differencing and modeling tasks in order. The modeling component is inclusive of steps: model identification, parameter estimation, diagnostics, and forecasting. ARIMA (6,1,12) was found to be the best fit for this purpose. Accuracy was measured over Mean Absolute Percentage Error, however, dropped by 0.35% and 0.1% for 2-week and 1-month periods respectively. This study has not considered external factors such as Tweets/comments posted, governmental intervention, and media influence.

Most recently, Bitcoin price was predicted by Ranjith et al. [12] combining LSTM and Sentiment Analysis on Bitcoin-related Tweets and Reddit posts. They have concluded that modifying the proposed ML-driven systematic approach with a small improvement, provides the capability in predicting any cryptocurrency.

Another study in 2022, conducted by Jacques et al. [13] considered the Time Series Analysis of Cryptocurrency prices to find the best model between the LSTM approach and ARIMA in terms of price forecasting. LSTM is a machine learning method that is used for time series data processing, prediction, and classification. According to the study, the LSTM method is the most accurate model for predicting future values. Nonetheless, a limitation of this study is that it has not explored in detail the architecture for introducing such hyperparameters.

# 4. Methodology

Secondary data analysis involves a researcher using the information that someone else has gathered for his or her purposes. We used secondary data for this study. The dataset has been taken from the finance.yahoo.com website and it has recorded all the Bitcoin data daily, weekly, and monthly. Yahoo Finance is a media property that is part of the Yahoo! Network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. In 2017 Yahoo Finance added a feature to look at the news surrounding cryptocurrency. It lists over 9,000 unique coins including Bitcoin.

We selected historical data of Bitcoin price from 17$^{th}$ Sep 2014 to 31$^{st}$ Dec 2019. It has 7 features namely Date, Open, High, Low, Close, Adj close and Volume of the Bitcoin data. The High and the Low refer to the maximum and the minimum Bitcoin values. The Open and the Close are the periods where the consideration of Bitcoin values began and ended.

Here we used the Adjusted close which denotes **Adj close** for the analysis. Adjusted close is the closing value after adjustments for all applicable splits and dividend distributions. Since the adjusted close begins where the close ends.

After January of 2020, we can see some fluctuations and after April 2021, we can see a huge fluctuation according to the data set. The World Health Organization (WHO) declared the outbreak a public health emergency of international concern on 30$^{th}$ January 2020 and a pandemic on 11$^{th}$ March 2020. Hence to remove the impact of the COVID-19, we only considered the data set till the 31$^{st}$ of December 2019. And also in the daily data set we retrieve only weekdays for the analysis because trading activities of Bitcoin is getting down on weekends.

When using a data set, data cleaning is one of the most important steps in quality decision-making. Removing duplicates or irrelevant observations, filtering unwanted outliers, and handling missing data are some of the methods to clean data. In this data set we do not use filtering outliers because in general it is not giving any significant effect to improve the accuracy of the forecast in time series analysis. If there is missing value in the dataset like holidays, we replace it using the previous Bitcoin price value.

The statistical analysis part was performed using the EViews software. EViews is used for general statistical and econometric analysis, such as cross-section and time series estimation and forecasting.

We divided the data set into two parts as training set and the testing set. Nearly 80% of the data included in the training set and the other 20% included in the testing set. We make our analysis part on the training set and predict the values to the testing set and always take our level of significance as 0.05 throughout the analysis.

We find the best ARIMA model for Bitcoin price prediction with the help of the Box-Jenkins method. It takes several steps to identify the best model.

The Box-Jenkins method consists of three main parts which are identification, estimation, and diagnostic checking.

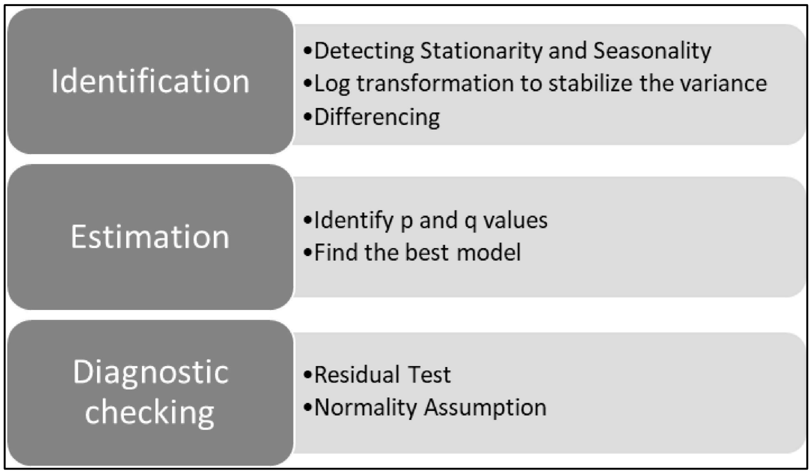



*Figure 1: The Box-Jenkins method*

1. **Identification**

**Detecting Stationarity and Seasonality**

Stationarity can be detected from the Correlogram as well as the Augmented Dicky-Fuller test.

The Augmented Dickey-Fuller test, tests the null hypothesis that there is a unit root obtained in a time series sample. The alternative hypothesis differs from the equation that we use to analyze the time series.

The Augmented Dickey-Fuller statistic, used in the test, is a negative number. More negative, it means there is a high probability of rejecting the null hypothesis that there is a unit root at some level of confidence and the series is not stationary.

Seasonality can usually be observed from a correlogram. If there is any seasonality, we use seasonal differencing to remove it. But in our data set we could not observe any seasonal variation.

EViews allows all these options to choose from.

**Applying log transformation to stabilize the variance**

In order to stabilize the variance, we apply a power transformation. Here we used log transformation because it is effective at removing exponential variance.

**Differencing to remove the trend**

We develop an ARIMA model when the variables are stationary. If the series is not stationary, we use differencing methods to convert it stationary.

2. **Estimation**

**Identify *p* and *q* values**

After the dataset becomes stationary, the next step is to identify the order of the autoregressive - AR(p) and moving average - MA(q) terms. Observing ACF and PACF graphs we can suggest some p and q values.

**Best model Identification**

Now we have a set of possible models. The best models are selected by minimizing Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC). Number of significant parameters are also taken into account.

3. **Diagnostic checking**

**Residual Test**

It is essential to check whether the residuals are white noise or not. If the residuals are not white noise, it means that they are not independent and the residuals show a certain relationship among them. Since we cannot get correct assumptions for time series, then we use Q-test statistics to test this situation. If the p-value is greater than 0.05, the residuals indicate white noise. Otherwise, we reject that ARIMA model.

**Normality Assumption**

The histogram can be used to show the frequency distribution of the residuals. The histogram is obtained by dividing the range of the series of data into equal-sized classes or bins and then for each bin, the number of observations from the data set is counted. After that we use the Jarque-Bera test to determine whether sample data have normal distribution or not. The test statistic for the Jarque-Bera test is always positive and, if it is too far from zero, it means that the sample data are not normally distributed.

But the problem is that most of the time series that are based on financial data violate the normality assumption.

However, the best model is selected when the residuals are independent and normally distributed. If the normality assumption is violated, we only consider the residual test and get the best model for predicting future values.

Once we have completed the Box Jenkins analysis, we forecast the values using the best model we chose. We took forecast values up to 31$^{st}$ of January 2023.

## 5. Analysis of Data

### 5.1 Reasons behind the Sudden Increase in Bitcoin Price

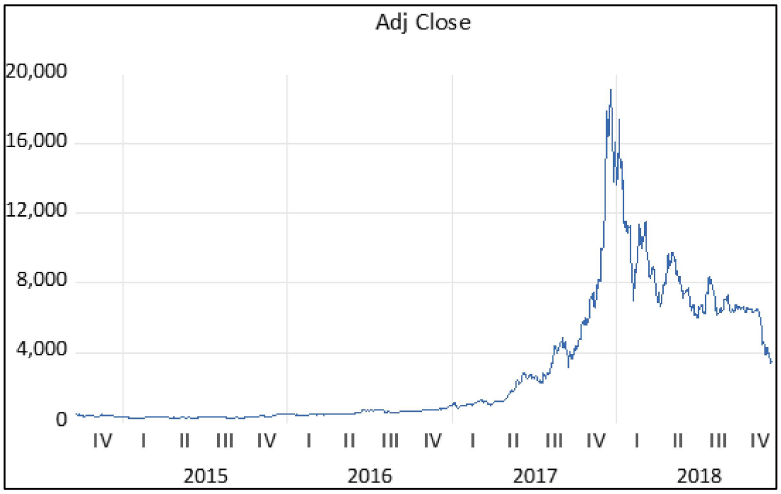



*Figure 17: Time series plot for daily Bitcoin value*

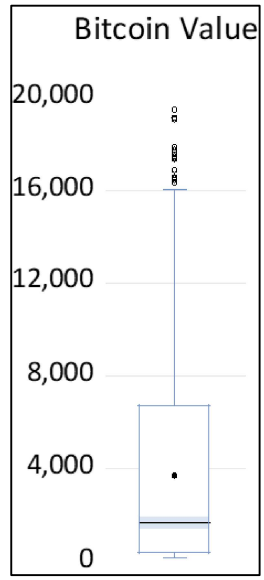



*Figure 18: Box-plot for daily Bitcoin value*

By observing the box-plot for the daily data set we can see there are some outliers. Those outliers were dated in the period of 6th December 2017 to 15th January 2018. Reasons for this can be discussed under several areas.

Scarcity is considered as the main reason for this. At the beginning Sathoshi Nakamoto embedded a limit of 21 million coins in the original Bitcoin code. That means only 21 million Bitcoins can ever exist. As the network approaches the maximum float, Nakamoto intended the coins to become more difficult to obtain. It was estimated that it would take 120 years to mine the remaining 10% of Bitcoins needed to get to the 21 million limit. Since there is a belief that there will be fewer opportunities to obtain Bitcoin, or that it will reside in the hands of a selected few are often cited as a reason for the demand for Bitcoins. Because of these reasons Bitcoin's popularity has been growing among average folks and institutional investors.

## 5.2 Model Building and Selection

In order to build and select the best model, analysis was done as follows using training sets.

### 1. Interpretations for daily data (Training set)


*Figure 19: Time series plot for daily training set*

This figure presents the time series plot for daily training set. Daily adjusted close price for training set slightly increase until the 4th quarter of year 2017. Afterwards, it has shown a sudden increase of nearly 2.5 times. From the 1st quarter of 2018 onwards, the plot depicts a downward trend. To confirm the non-stationarity of the data set we use Augmented Dickey Fuller test.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.889558 | 0.3374 |
| Test critical values: | 1% level | -3.436138 | |
| | 5% level | -2.863984 | |
| | 10% level | -2.568122 | |
| *MacKinnon (1996) one-sided p-values. | | | |

*Figure 20: Results from Augmented Dickey Fuller test on daily training set*

Null hypothesis: Series is not stationary

It was found that this original series is not stationary as p-value is equals to 0.3374 which is greater than 0.05. We do not reject null hypothesis at 5% level of significance. Therefore, results suggest that time series is not stationary.

Since the original series is not stationary, it means mean or variance or both are not constant over the time. Therefore, we first apply log transformation to make variance constant.


*Figure 21: Time series plot for log transformed series on daily training set*

This above figure presents the series after applying log transformation to stabilize the variance. In this figure the volatility was roughly eliminated. To find out the stationarity of log transformed series, Augmented Dickey Fuller test was applied.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -0.472760 | 0.8938 |
| Test critical values: | 1% level | -3.436062 | |
| | 5% level | -2.863950 | |
| | 10% level | -2.568104 | |
| *MacKinnon (1996) one-sided p-values. | | | |

*Figure 22: Results from Augmented Dickey Fuller test for log transformed series on daily training set*

Null hypothesis: Series is not stationary

The null hypothesis is not rejected at 5% level of significance. The Augmented Dickey Fuller test provides sufficient evidence for non-stationarity of log transformed series.

Since the log transformed series is also not stationary, the 1st order differencing is applied to make mean constant.

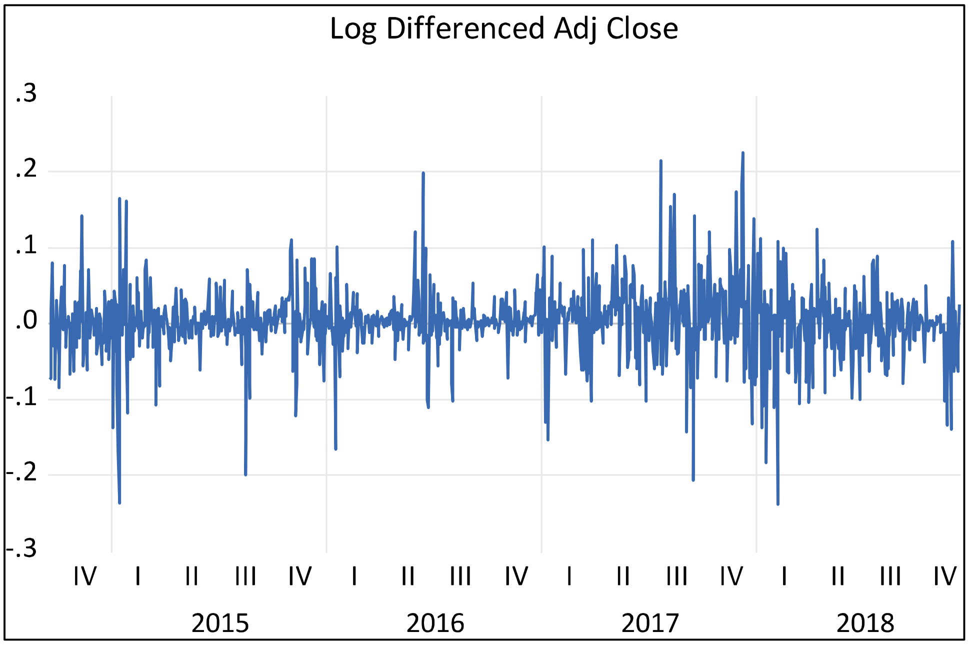



*Figure 38: Time series plot after applying 1st order differencing on log transformed series for daily training set*

The above figure is time series plotted for 1st order differenced log transformation of adjusted close price for training set. The mean has been stabilized after applying the 1st order differencing. To confirm the stationarity of 1st order differencing on log transformed series the same test was applied.

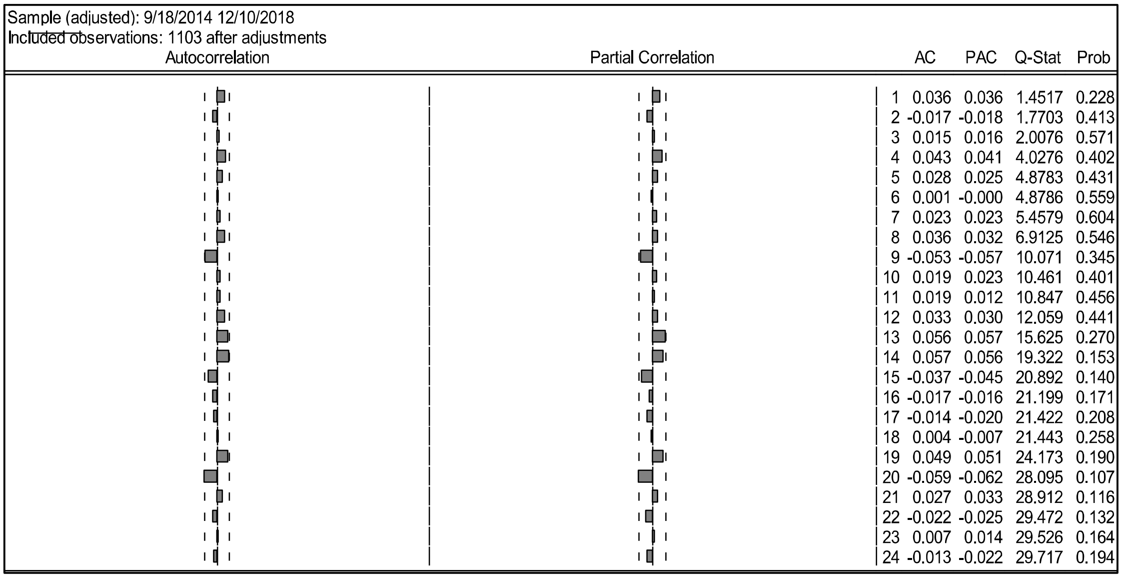
| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -32.02376 | 0.0000 |
| Test critical values: | 1% level | -3.436067 | |
| | 5% level | -2.863953 | |
| | 10% level | -2.568106 | |
| *MacKinnon (1996) one-sided p-values. | | | |

*Figure 39:Results from Augmented Dickey Fuller test on 1st order differenced log transformed series for daily training set*

Null hypothesis: Series is not stationary

Augmented Dickey Fuller test confirms further that, p-value equals to zero and is less than considered 0.05. Therefore, we rejected the null hypothesis at 5% level of significance. Thus, it provides sufficient evidence that differenced log transformed series is stationary.

In order to build ARIMA models we observed correlogram of 1st order differencing on log transformed series.



Sample (adjusted): 9/18/2014 12/10/2018
Included observations: 1103 after adjustments

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.036 | 0.036 | 1.4517 | 0.228 |
| | | 2 | -0.017 | -0.018 | 1.7703 | 0.413 |
| | | 3 | 0.015 | 0.016 | 2.0076 | 0.571 |
| | | 4 | 0.043 | 0.041 | 4.0276 | 0.402 |
| | | 5 | 0.028 | 0.025 | 4.8783 | 0.431 |
| | | 6 | 0.001 | -0.000 | 4.8786 | 0.559 |
| | | 7 | 0.023 | 0.023 | 5.4579 | 0.604 |
| | | 8 | 0.036 | 0.032 | 6.9125 | 0.546 |
| | | 9 | -0.053 | -0.057 | 10.071 | 0.345 |
| | | 10 | 0.019 | 0.023 | 10.461 | 0.401 |
| | | 11 | 0.019 | 0.012 | 10.847 | 0.456 |
| | | 12 | 0.033 | 0.030 | 12.059 | 0.441 |
| | | 13 | 0.056 | 0.057 | 15.625 | 0.270 |
| | | 14 | 0.057 | 0.056 | 19.322 | 0.153 |
| | | 15 | -0.037 | -0.045 | 20.892 | 0.140 |
| | | 16 | -0.017 | -0.016 | 21.199 | 0.171 |
| | | 17 | -0.014 | -0.020 | 21.422 | 0.208 |
| | | 18 | 0.004 | -0.007 | 21.443 | 0.258 |
| | | 19 | 0.049 | 0.051 | 24.173 | 0.190 |
| | | 20 | -0.059 | -0.062 | 28.095 | 0.107 |
| | | 21 | 0.027 | 0.033 | 28.912 | 0.116 |
| | | 22 | -0.022 | -0.025 | 29.472 | 0.132 |
| | | 23 | 0.007 | 0.014 | 29.526 | 0.164 |
| | | 24 | -0.013 | -0.022 | 29.717 | 0.194 |

*Figure 40:Correlogram after applying 1st order differencing on log transformed series for daily training set*

Correlogram annotated above, represents that both ACF and PACF have similar behavioral patterns except few mild changes.

Observing Figure 4.2.1.7 we generated few ARIMA models. Also, we checked the validity of assumptions with Q-test statistic and Jarque-Bera test. Out of the models adhere to the assumptions, we selected the best according to the lowest AIC value. Hence ARIMA (14,1,14) is the best model to forecast daily Bitcoin prices.

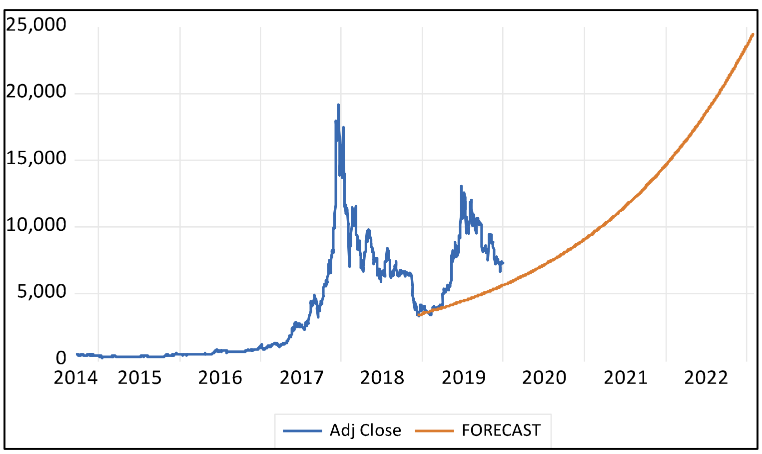



*Figure 41:Adjusted close and forecasted Bitcoin prices for daily data*

| Date | Forecasted Values |
|---|---|
| 1/02/2023 | 23556.61 |
| 1/03/2023 | 23599.83 |
| 1/04/2023 | 23643.12 |
| 1/05/2023 | 23686.50 |
| 1/06/2023 | 23729.95 |
| 1/09/2023 | 23773.48 |
| 1/10/2023 | 23817.09 |
| 1/11/2023 | 23860.79 |
| 1/12/2023 | 23904.56 |
| 1/13/2023 | 23948.41 |
| 1/16/2023 | 23992.34 |
| 1/17/2023 | 24036.36 |
| 1/18/2023 | 24080.45 |
| 1/19/2023 | 24124.63 |
| 1/20/2023 | 24168.89 |
| 1/23/2023 | 24213.22 |
| 1/24/2023 | 24257.64 |
| 1/25/2023 | 24302.14 |
| 1/26/2023 | 24346.72 |
| 1/27/2023 | 24391.39 |
| 1/30/2023 | 24436.13 |
| 1/31/2023 | 24480.96 |

*Figure 42: Daily forecasted Bitcoin prices from 1st to 31st of January 2023*

In Figure 11, We obtained forecasted values from the beginning of testing set up to 31st of January 2023.

## 2. Interpretations for weekly data (Training set)

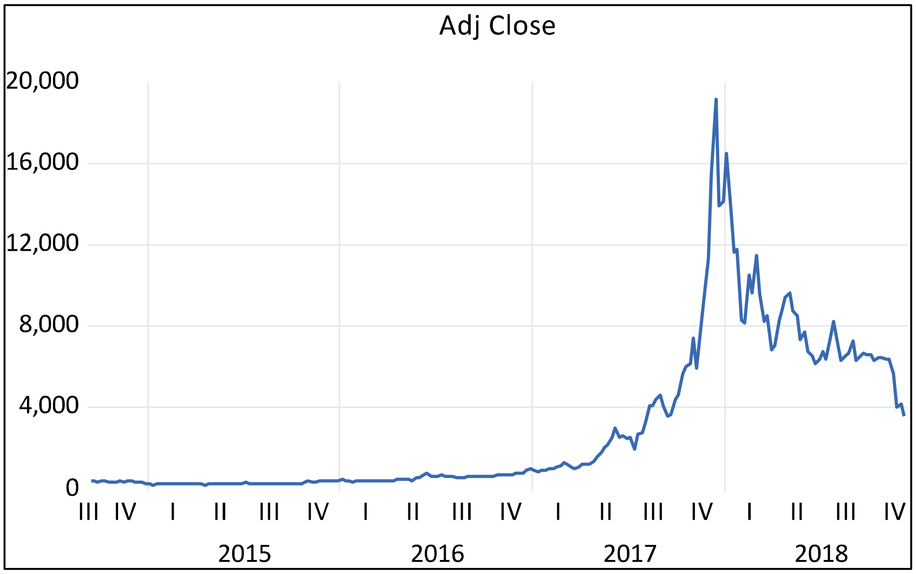



*Figure 43: Time series plot for weekly training set*

This figure presents the time series plot for weekly training set. It shows less fluctuation than daily data. To confirm the non-stationarity of the data set we use Augmented Dickey Fuller test.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.614550 | 0.4734 |
| Test critical values: | 1% level | -3.460453 | |
| | 5% level | -2.874679 | |
| | 10% level | -2.573850 | |
| *MacKinnon (1996) one-sided p-values. | | | |

*Figure 44: Results from Augmented Dickey Fuller test on weekly training set*

Null hypothesis: Series is not stationary

The null hypothesis is not rejected at 5% level of significance. From the obtained output of Augmented Dickey Fuller test for the original series, we can conclude the absence of stationarity.

As the original series is not stationary, it means mean or variance or both are not constant over the time. Therefore, we first apply log transformation to make variance constant.


*Figure 60: Time series plot for log transformed series on weekly training set*

Previous time series plot suggested that we need to proceed with stabilizing the variance of the adjusted close price value. The above figure illustrates the series after applying log transformation to eliminate the instability of the variance.
To find out the stationarity of log transformed series, Augmented Dickey Fuller test was applied.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -0.623913 | 0.8614 |
| Test critical values: | 1% level | -3.459898 | |
| | 5% level | -2.874435 | |
| | 10% level | -2.573719 | |

*MacKinnon (1996) one-sided p-values.

*Figure 76:Results from Augmented Dickey Fuller test on log transformed series for weekly training set*

Null hypothesis: Series is not stationary

The above figure represents output of Augmented Dickey Fuller test, applied on log transformed series and p-value is greater than 0.05. The null hypothesis is not rejected at 5% level of significance which proves the log transformed series is still not stationary.

Since the log transformed series is also not stationary, the 1st order differencing is applied to make mean constant.

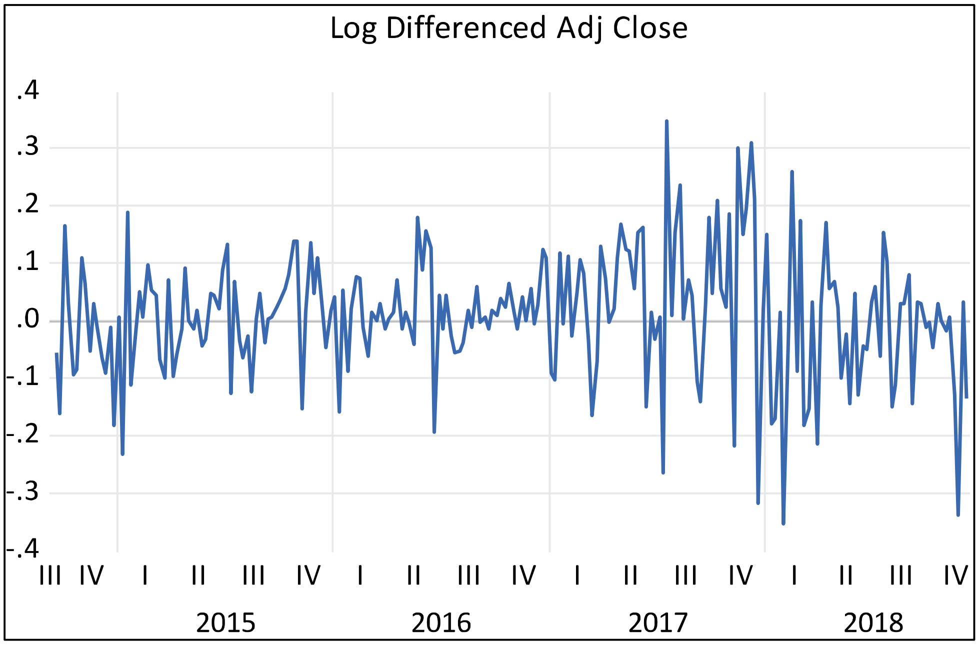



*Figure 92:Time series plot after applying 1st order differencing on log transformed series for weekly training set*

Plotted time series after applying 1st order differencing on log transformed series is depicted in the figure above. The trend has been eliminated after applying the 1st order differencing.

To confirm the stationarity of 1st order differencing on log transformed series the same test was applied.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -14.56257 | 0.0000 |
| Test critical values: | 1% level | -3.460035 | |
| | 5% level | -2.874495 | |
| | 10% level | -2.573751 | |

*MacKinnon (1996) one-sided p-values.

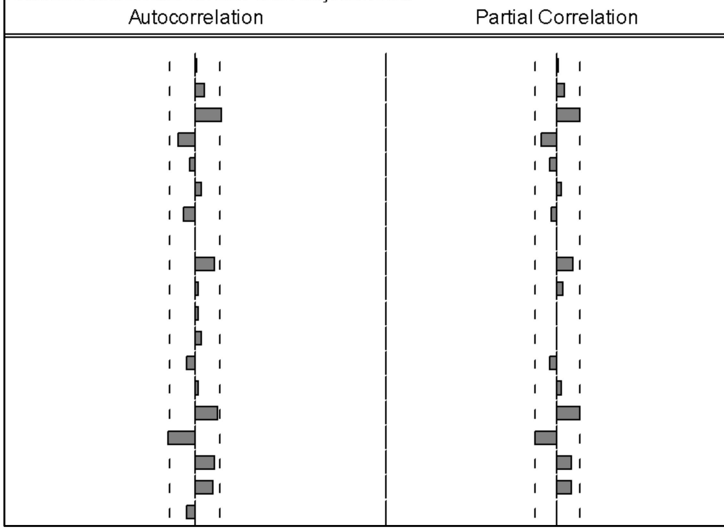
*Figure 108:Results from Augmented Dickey Fuller test on 1st order differenced log transformed series for weekly training set*

Null hypothesis: Series is not stationary

Using the Augmented Dickey Fuller test on differenced log transformed series, output implies that p-value equals zero and is less than 0.05. Thus, we rejected the null hypothesis at 5% level of significance. The stationarity is present here throughout the differenced log transformed series.

In order to build ARIMA models we observed correlogram of 1st order differencing on log transformed series.

Sample (adjusted): 9/22/2014 12/10/2018
Included observations: 221 after adjustments




| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | 0.012 | 0.012 | 0.0325 | 0.857 |
| 2 | 0.054 | 0.054 | 0.6898 | 0.708 |
| 3 | 0.141 | 0.140 | 5.1910 | 0.158 |
| 4 | -0.090 | -0.097 | 7.0242 | 0.135 |
| 5 | -0.027 | -0.041 | 7.1874 | 0.207 |
| 6 | 0.038 | 0.030 | 7.5155 | 0.276 |
| 7 | -0.059 | -0.031 | 8.3174 | 0.305 |
| 8 | -0.003 | -0.006 | 8.3199 | 0.403 |
| 9 | 0.112 | 0.106 | 11.232 | 0.260 |
| 10 | 0.023 | 0.040 | 11.360 | 0.330 |
| 11 | 0.022 | 0.003 | 11.477 | 0.404 |
| 12 | 0.035 | -0.004 | 11.762 | 0.465 |
| 13 | -0.048 | -0.039 | 12.309 | 0.503 |
| 14 | 0.022 | 0.029 | 12.428 | 0.572 |
| 15 | 0.127 | 0.132 | 16.288 | 0.363 |
| 16 | -0.141 | -0.132 | 21.073 | 0.176 |
| 17 | 0.110 | 0.095 | 23.973 | 0.120 |
| 18 | 0.104 | 0.083 | 26.612 | 0.087 |
| 19 | -0.041 | -0.007 | 27.023 | 0.104 |

*Figure 124: Correlogram after applying 1st order differencing on log transformed series for weekly training set*

Correlogram suggests ACF and PACF are having approximately similar patterns.

By observing above figure, we generated few ARIMA models and then we checked the validity of assumptions using Q-test statistic and Jarque-Bera test. We selected the best model based on its low AIC value, from those that meet the assumptions. Hence ARIMA (26,1,26) is the best model to forecast weekly Bitcoin prices.

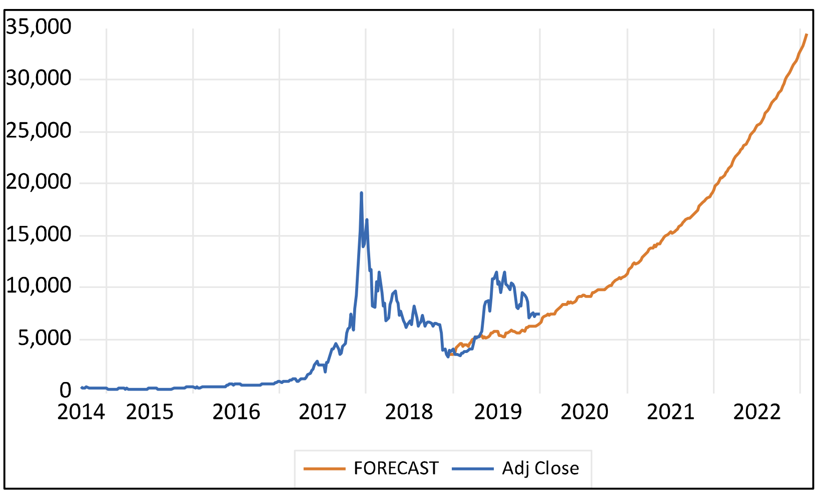



*Figure 172:Adjusted close and forecasted Bitcoin prices for weekly data*

| Date | Forecasted Bitcoin prices |
|---|---|
| 1/02/2023 | 32788.61 |
| 1/09/2023 | 33286.24 |
| 1/16/2023 | 33671.20 |
| 1/23/2023 | 34067.86 |
| 1/30/2023 | 34418.61 |

*Figure 140: :Weekly forecasted Bitcoin prices from 1st week to last week of January 2023*

In Figure 20, We obtained forecasted values from the beginning of testing set up to last week of January 2023.

### 3. Interpretations for monthly data (Training set)

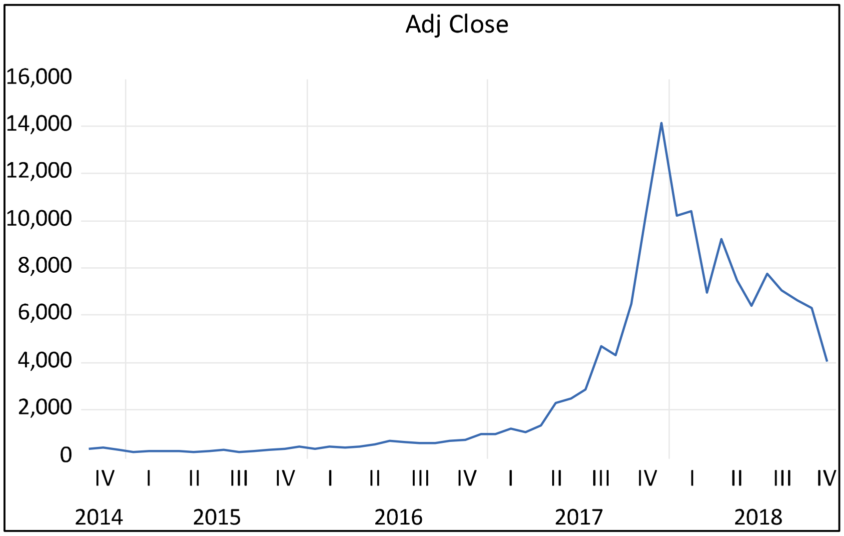



*Figure 188:Time series plot for monthly training set*

According the figure, adjusted close prices for training set approximately similar to time series of those recorded daily and monthly. However, the curve shows least fluctuation compare to the original time series for daily and weekly adjusted close prices.
To confirm the non-stationarity of the data set we use Augmented Dickey Fuller test.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.339553 | 0.6039 |
| Test critical values: | 1% level | -3.571310 | |
| | 5% level | -2.922449 | |
| | 10% level | -2.599224 | |

*MacKinnon (1996) one-sided p-values.

*Figure 189: Results from Augmented Dickey Fuller test on monthly training set*

Null hypothesis: Series is not stationary

According to the above figure, the p-value is greater than 0.05. The null hypothesis is not rejected at 5% level of significance. Thus, Augmented Dickey Fuller test on monthly training data confirms that it is not stationary.

Since the original series is not stationary, it means mean or variance or both are not constant over the time. Therefore, we first apply log transformation to make variance constant.

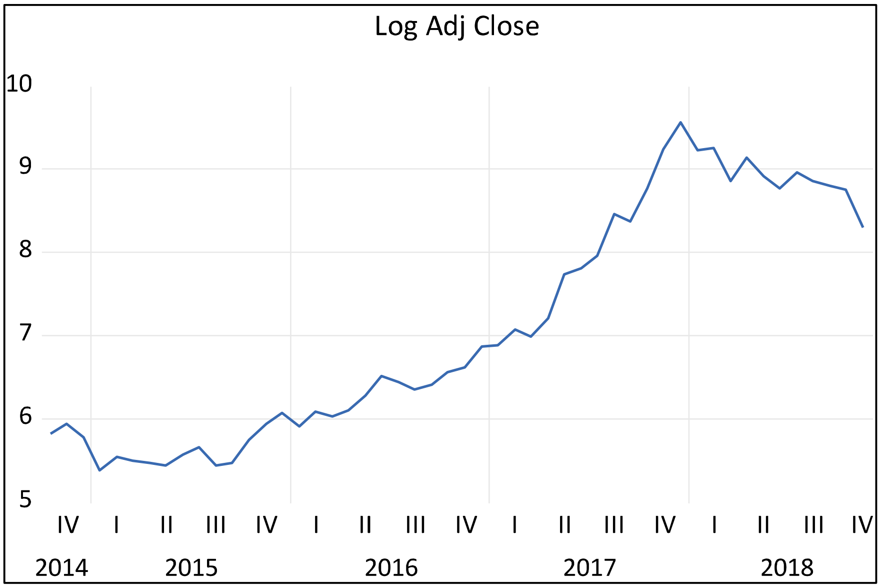



*Figure 205:Time series plot for log transformed series on monthly training set*

Above figure shows that, for log transformed series of training set of monthly prices, the behavior is approximately as same as those in daily and weekly cases. Instability of variance was roughly eliminated, within the log transformed series for monthly training set.
To find out the stationarity of log transformed series, Augmented Dickey Fuller test was applied.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -0.586285 | 0.8641 |
| Test critical values: | 1% level | -3.571310 | |
| | 5% level | -2.922449 | |
| | 10% level | -2.599224 | |

*MacKinnon (1996) one-sided p-values.

*Figure 221:: Results from Augmented Dickey Fuller test on log transformed series for monthly training set*

Null hypothesis: Series is not stationary

From the obtained output it is clear that, p-value is greater than considered 0.05. The null hypothesis is not rejected at 5% level of significance. We then, concluded that log transformed series is non-stationary.

Since the log transformed series is also not stationary, the 1st order differencing is applied to make mean constant.

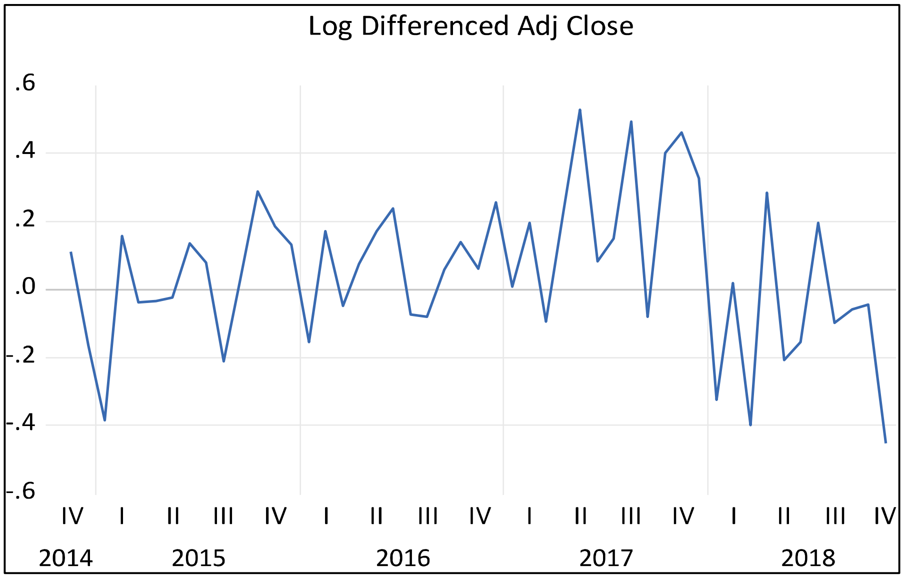



*Figure 237:Time series plot after applying 1st order differencing on log transformed series for monthly training set*

The above figure is time series plotted for 1st order differenced log transformation of adjusted close price for monthly training set. The trend has been removed after applying the 1st order differencing.

In order to confirm the stationarity of 1st order differencing on log transformed series the same test was applied.

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -6.072065 | 0.0000 |
| Test critical values: | 1% level | -3.574446 | |
| | 5% level | -2.923780 | |
| | 10% level | -2.599925 | |
| *MacKinnon (1996) one-sided p-values. | | | |

*Figure 253:Results from Augmented Dickey Fuller test on 1st order differenced log transformed series for monthly training set*

Null hypothesis: Series is not stationary

The output suggests that, the p-value equals to zero and is less than 0.05. The null hypothesis is rejected at 5% level of significance. Thus, we have enough evidence to conclude the stationarity after applying 1st order differencing of the log transformed series.

In order to build ARIMA models we observed correlogram of 1st order differencing on log transformed series.

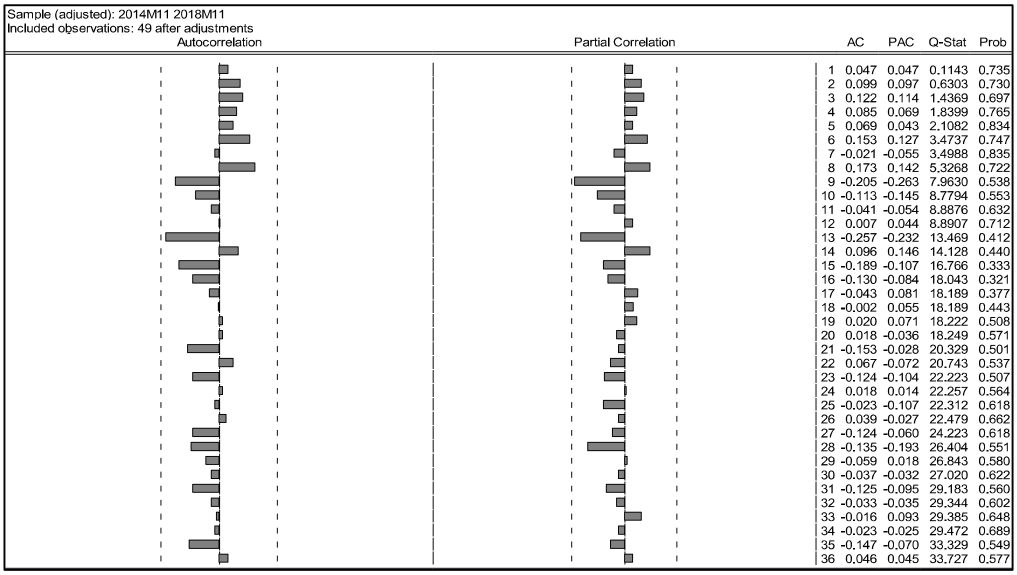



*Figure 269:Correlogram after applying 1st order differencing on log transformed series for monthly training set*

Both ACF and PACF are having approximately similar lag patterns within confidence band.

As above figure illustrates, several ARIMA models were generated and then we performed the validity check for assumptions using Q-test statistic and Jarque-Bera test. Then we determined the best out of them considering which gives the lowest AIC value and the highest number of significant parameters. Hence, ARIMA (0,1,8) is the best model to forecast monthly Bitcoin prices.

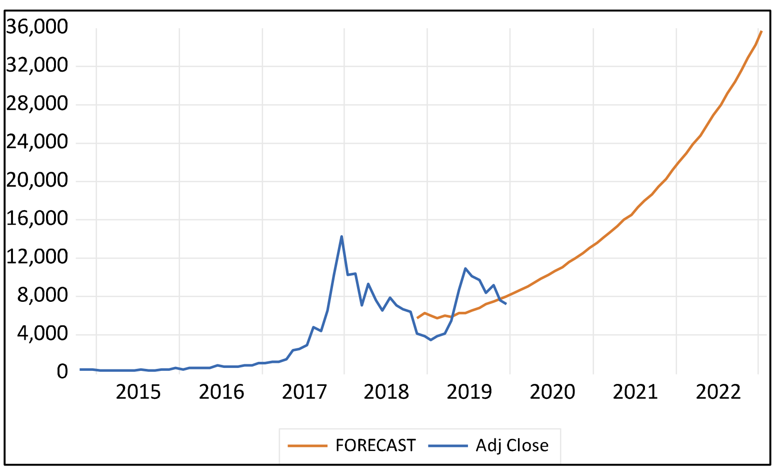



*Figure 301:Adjusted close and forecasted Bitcoin prices for monthly data*

| Date | Forecasted Bitcoin prices |
|---|---|
| 2023M01 | 35659.83 |

*Figure 285: Monthly forecasted Bitcoin price for January 2023*

In Figure 29, We obtained forecasted values from the beginning of testing set up to January 2023.

## 6. Conclusion and Discussion

We proceeded our study with historical data from Yahoo Finance. Obtained data are Bitcoin Prices collected daily, weekly and monthly. Based on lower AIC values ARIMA (14,1,14) and ARIMA (26,1,26) were found to be the best models to forecast daily and weekly Bitcoin prices respectively. For monthly recorded data, ARIMA (0,1,8) is the best model based on lower AIC value with more significant parameters.

We have encountered a few obstacles during this study. They are namely time constraint and scope constraint. It has been underlined that when forecasting cryptocurrencies, ML approaches in particular for Bitcoin exhibited the best accuracy. ML or DL techniques were introduced to stand out among other forecasting methods. Consequently, LSTM which falls under prominent DL algorithms can easily address high volatility in financial data. Time constraint however, has limited the study duration. And, it was not very convenient to make use of LSTM for this purpose. Additionally, given that the conductors must have extensive understanding of ML and DL, the intricacy of such an algorithm could have broadened the study's focus. However, we proceeded with ARIMA, making the scope simple.

A comparison of accuracy between ARIMA and ML or DL approaches would have enhanced this work. One of our study's omissions is this. There were some aspects that had a negative impact on model creation and selection in particular, even though we employed ARIMA. The period 2020 onwards is considered a COVID 19 period. During this period, we observed high volatility of Bitcoin prices. Unlike ML or DL approaches, ARIMA is less capable to capture highly fluctuating behaviors especially in financial time series and perform forecasts/predictions accordingly. That period had to be excluded and prices up to 31st December 2019 was the considered period. Having that mentioned, we then focused on sorting out problems experienced in model building. In some ways, adapting study conductors to utilize EViews made the process simple; nonetheless, this unexpectedly required more work and time.

.

## List of Abbreviations

| | | |
|---|---|---|
| ACF | - | Autocorrelation Function |
| AR | - | Auto Regressive |
| ARIMA | - | Autoregressive Integrated Moving Average |
| CNN | - | Convolution Neural Network |
| COVID | - | Coronavirus Disease |
| DL | - | Deep Learning |
| LR | - | Linear Regression |
| LSTM | - | Long Short Term Memory |
| MAE | - | Mean Absolute Error |
| MAPE | - | Mean Absolute Percentage Error |
| ML | - | Machine Learning |
| PACF | - | Partial Autocorrelation Function |
| RMSE | - | Root Mean Squared Error |
| RNN | - | Recurrent Neural Network |
| USD | - | United State Dollar |
| WHO | - | World Health Organization |

# References


[1] P. D. DeVries, “An Analysis of Cryptocurrency, Bitcoin, and the Future”. *International Journal of Business Management and Commerce*, vol. 1, no. 2, 2016.

[2] G. Giudici, A. Milne, and D. Vinogradov, “Cryptocurrencies market analysis and perspectives”. *Journal of Industrial and Business Economics*, vol. 47, no. 1, 2020.

[3] R. K. Rathore and M. Shutaywi, “Real world model for Bitcoin price prediction”. *Information Processing and Management*, vol. 59, no. 4, 2022.

[4] Y. Feng and D. P. Palomar, “A signal processing perspective of financial engineering,” *Foundations and Trends® in Signal Processing*, vol. 9, no. 1-2, pp. 1–231, 2016.

[5] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 2nd ed. New York: Springer, 2002.

[6] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, “Stock price prediction using the Arima model,” *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 2014.

[7] F. Liantoni and A. Agusti, “Forecasting Bitcoin using double exponential smoothing method based on mean absolute percentage error,” *JOIV: International Journal on Informatics Visualization*, vol. 4, no. 2, p. 91, 2020.

[8] J. Filipe and B. Mendes, “FORECASTING BITCOIN PRICES ARIMA vs LSTM,” 2019.

[9] G. Vidyulatha, M. Mounika, and N. Arpitha, “Crypto Currency Prediction Model using ARIMA,” 2020. [Online]. Available: https://www.blockchain.info/

[10] Z. Ayaz, J. Fiaidhi, A. Sabahand M. Anwer Ansari, “Bitcoin Price Prediction using ARIMA Model”. TechRxiv, 09-Apr-2020 [Online]. Available: https://www.techrxiv.org/articles/preprint/Bitcoin_Price_Prediction_using_ARIMA_Model/12098067/1. [Accessed: 05-Dec-2022]

[11] O. G. Darley, A. I. O. Yussuff, and A. A. Adenowo, “Price Analysis and Forecasting for Bitcoin Using Auto Regressive Integrated Moving Average Model,” *Annals of Science and Technology*, vol. 6, no. 2, pp. 47–56, Dec. 2021, doi: 10.2478/ast-2021-0009.

[12] K. R. Reddy, K. Lenka, N. Devi, and D. v Rithika, “Bitcoin Price Prediction and Forecasting,” 2022. [Online]. Available: www.irjet.net

[13] J. P. Fleischer, G. von Laszewski, C. Theran, and Y. J. P. Bautista, “Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory,” *Algorithms*, vol. 15, no. 7, Jul. 2022, doi: 10.3390/a15070230