TIME SERIES ANALYSIS ON CRYPTOCURRENCY: AN ARIMA MODEL APPROACH

Forecasting Bitcoin and Ethereum Prices

[1]Tinni Chaudhuri, [2]Pragya Dhar

[1]Assistant Professor, [2]Student Mstat
[1]Department of Statistics, AIASK
[1]Amity University, Kolkata, India

Abstract: A Box-Jenkins ARIMA (Auto Regressive Integrated Moving Average) model is used in here to predict the price of Bitcoin for future weeks. ARIMA (6,1,3) model is found to be the best ARIMA model for this current study of the Bitcoin prices. The efforts here are focussed on predicting weekly prices for the next 10 weeks and checking whether ARIMA works accurately with weekly data. The forecast results are validated for the first 6 weeks using the available data from online sources.

Keywords: ARIMA, Bitcoin, Ethereum, forecast, time series, modelling

I. INTRODUCTION

Cryptocurrency is a form of encrypted currency which does not exist in the actual world in a material form. Because of its encryption, it becomes impossible to track or counterfeit any transaction. It is a decentralised method of currency transaction. Some examples of famous cryptocurrencies include Bitcoin, Ethereum, Litecoin, Ripple and Stellar among many others.

Bitcoin is the first blockchain created by Satoshi Nakamoto in 2008. Its value has exponentially risen since then. It is used in more than 50 countries of the world. Current price of 1 bitcoin stands at $53,480.83 and its trading volume till March 24, 2021 is $62,341,375,133.95.

Very recently in March 2021, Indian Government made sure that no transactions in any form of cryptocurrency was carried out. But slowly the outlook toward cryptocurrency is changing in the country. Discussion and questions about the cryptocurrency have been brough up in the parliament even in the month of March 2021. Although currently any transaction of cryptocurrency is banned in India, there is scope for regulated transaction allowance of bitcoin in the future.

A time series analysis will provide the predicted price of bitcoin and Ethereum in near future. Although this will give a forecast for a short period in the future, this work can be made useful by varying the dates of the data.

An ARIMA (p,d,q) model will be fit to the stationary series obtained from the price and the forecast will be based on the predicted values of this series. In this work an effort is made to forecast price for the next 10 weeks. The said model was introduced first by Box and Jenkins in 1960. The open-source analytical software 'R' including packages 'tseries', 'forecast', ggplot2 etc. have been used in this study.

II. DATA DESCRIPTION

For this study, the data is obtained on weekly price of Bitcoin and Ethereum respectively, since January 5, 2020 as obtained from online source Investing.com till February 28, 2021.

The original data as obtained from the mentioned site contains other columns such as Opening Price, Closing Price, Highest Price, Volume, Change % and so on. Since this project is only concerned with the Price column, an extract has been given here. It can be observed that the price changes and the percentage of change varies week by week. Over time the price has usually increased to an incredible peak even though the increment is not monotonous.

III. METHODOLOGY

1. *Stationary process*: Any time series X_t is stationary if for any $(t_1, t_2, \ldots, t_k)$ and h>0,

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_k}}(\mathrm{x}_{t_1}, \mathrm{x}_{t_2}, \ldots, \mathrm{x}_{t_k}) = F_{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h}}(\mathrm{x}_{t_1}, \mathrm{x}_{t_2}, \ldots, \mathrm{x}_{t_k})$$

That is,

- $E(X_t)$ is independent of t.
- Covariance of $(\mathrm{X}_{t_1}, \mathrm{X}_{t_2}, \ldots, \mathrm{X}_{t_k})$ is independent of t

To test the stationarity of any series Augmented Dickey-Fuller Test is performed. If the raw series of Price fails the test, the data is changed into a stationary one using different methods.

Another method of visually checking the stationarity of the Price column, line diagram is plotted. If any trend is observed in the data, it will be non-stationary.

2. *Augmented Dickey Fuller Test*: This is an augmented version of the original Dickey Fuller tets. It tests the existence of unit root in the time series sample. The presence of unit root indicates that the data is non-stationary and needs to be made stationary. The value of the test statistic is negative. The lower the value, the stronger the rejection of the null becomes.

Suppose the model has both constant and trend:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda_t + \nu_t$$

Hypothesis: $H_0 : \gamma = 0$

Test statistic: $DF = \dfrac{\hat{\gamma}}{SE(\hat{\gamma})}$

3. *Box-Jenkins Method*: There are 3 steps in this method – Model identification and model selection, Parameter estimation, and Statistical model checking.

ARIMA: This is the most commonly used time series model for forecasting. For given non-seasonal, non-stationary time series Xt having trend, the ARIMA model is written as

$$\phi(B)(1-B)^{\delta} X_t = \psi(B)e_t$$

where

δ : Number of times we need to differentiate Xt in order to remove trend

B: Backward shift operator

$$\phi(B): 1-\alpha_1 B - \alpha_2 B^2 - \alpha_3 B^3 - ... - \alpha_p B^p$$

$$\psi(B): 1+\beta_1 B + \beta_2 B^2 + ... + \beta_p B^p$$

$e_t \sim$ W $N(0,1)$; WN: white noise

4. *Model identification*: Stationarity and seasonality is first checked using a trend plot of the original data. Price is plotted over time in a line graph. The graph is shown in Fig. 1.1. If any trend or seasonality is present, then that needs to be removed with the help of differences.

Fig. 1.1: Line plot of Price of Bitcoin (in $ USD)

1. The line chart for Ether prices is given in Fig. 2.1.

Fig. 2.1: Line chart for Ethereum price (in $ USD)

The plot clearly shows that trend is present in the data although there seems to be no seasonality. Therefore this also needs to be differenced in order to get a stationary data.

5. *Differencing to achieve stationarity*: First order difference (d=1) is then applied on the Price column for both the data sets. This will give a differenced series, denoted by FOD. The first value for this series will be null. A line chart for this FOD is then plotted to check whether it is stationary.

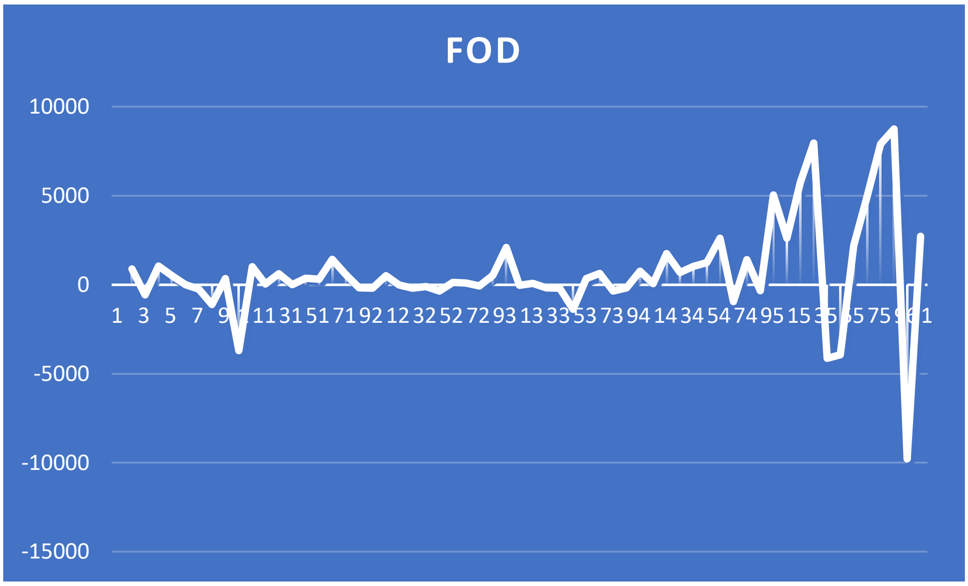

Fig. 1.2: Line Plot for differenced series of Bitcoin

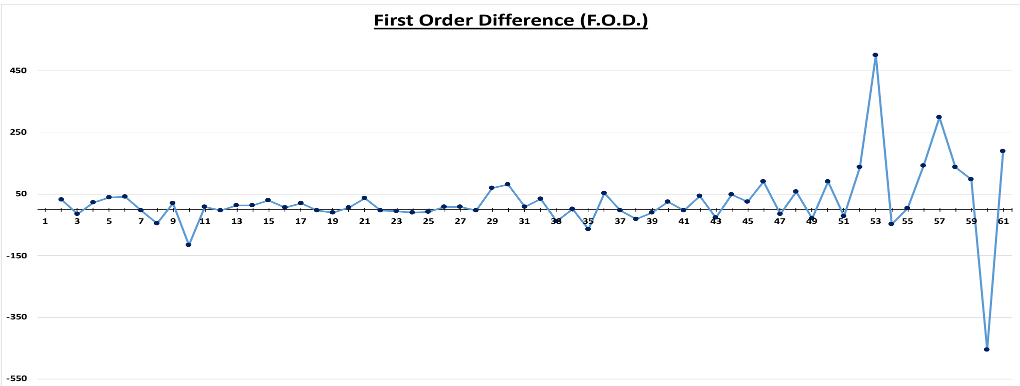

Fig. 1.2: Line Plot for differenced series of Ethereum

Although the both the plots look stationary, this can be confirmed using ADF test as explained before.

The test is carried out in the R software.

Bitcoin:

The obtained test statistic after performing the test in R is -6.5635 for lag order 3 and the p-value turned out to be 0.01. Therefore, at 5% level of significance, the p-value gets rejected. There is no unit root present in FOD which implied that this series is stationary. Hence parameter d=1 is fixed.

Ethereum:

The R output for Ethereum gives the test statistic as -3.701 with lag order 3 and p-value 0.03229. So, here also the null hypothesis gets rejected at 5% significance level. This states that the differenced series of Ethereum price is stationary and further work can be done using this series. Parameter d=1.

6. *Identifying the AR and MA parameters*: The AR parameter is denoted by p, obtained from Partial Auto-correlation function (PACF) plot, while, the MA parameter q is obtained from the auto-correlation (ACF) plot. An AR(1) process should have an exponentially decreasing function. For AR(p) process, it becomes 0 at lag p+1 and higher, so departure from 0 is observed in PACF plot. The outliers for the process are obtained by checking for the lag points in which the functional value lies outside the control bands. These control bands are 95% confidence interval of the function itself.

The same thing is done for the ACF plot. MA(q) becomes 0 at lag q+1 and more.
The ACF and PACF plots for FOD of Bitcoin Price is given in Fig. 1.3.

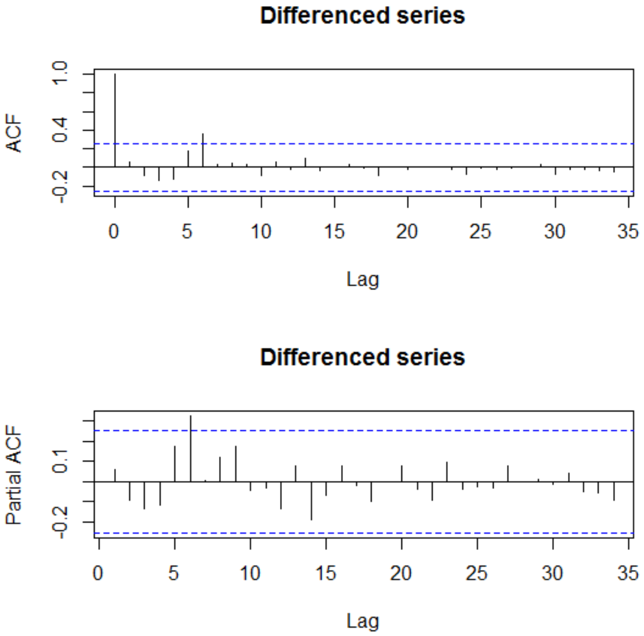

Fig. 1.3: ACF and PACF plots for Bitcoin.

It can be observed that there is significant spike at lag 6 for both the plots. The choices of p and q should therefore vary between 0,1,…,6.

The ACF and PACF plots for FOD of Ethereum Price is given in Fig. 2.3

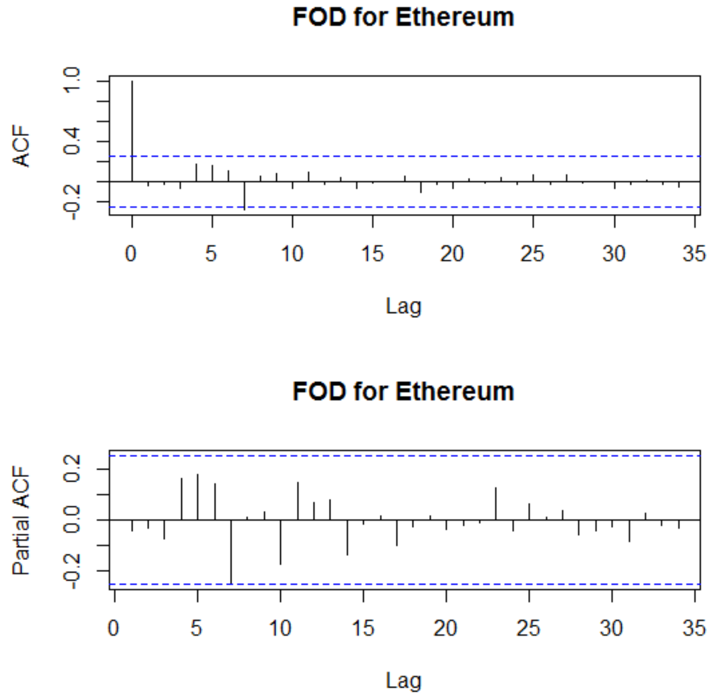

Fig. 2.3: ACF and PACF for Ethereum

Significant spikes are at lags 7 for both ACF and PACF. 7^2 combinations are then tried out in R to find the minimum AIC.

7. *Model estimation:* The most appropriate ARIMA model is the one which has the least Akaike Information Criterion (AIC) value. AIC is an estimation of the prediction error, thererfore, least value of AIC indicates a good model in which error is minimum.

Different combinations of p and q are interchanged in order to get the lowest AIC value. After trying several combinations, the lowest AIC for Bitcoin turned out to be *1110.5611*, for the combination p=3 and q=6. So, the appropriate model for bitcoin is therefore ARIMA(3,1,6).

The lowest AIC for Ethereum came as *725.9639*, for the combination p=4 and q=6 and the appropriate model for Ethereum is, therefore, ARIMA(4,1,6).

8. *Model diagnostics:* A number of diagnostic checks can be done on the model to validate it.

Portmanteau Test: This test checks whether the fit is good.
H0: Fit is good
H1: Fit is not good

Test statistic:

$$Q_p = n\sum_{h=1}^{h^*} \hat{\varphi}_r^2(h) \sim \chi^2_{h^*-(p+q)};\ \text{assymptotically}$$

The null hypothesis is rejected if $Q_p > \chi^2_{\alpha,h^*-(p+q)}$, α: level of significance.

$\hat{\varphi}_r(h)$: Sample ACF of lag h of residual series (r) and h* is chosen such that $h \le h^* \forall h$.

Ljung Box Test: This test is quite similar to the previous test. This test also deals with the residuals.

Test Statistic:

$$Q^*_p = (n+2)n\sum_{h=1}^{h^*} \hat{\varphi}_r^2(h)/(n-h) \sim \chi^2_{h^*-(p+q)};\ \text{assymptotically}$$

Decision Rule: Reject H0 if $Q^*_p > \chi^2_{\alpha,h^*-(p+q)}$

Bitcoin:
The obtained test statistic after performing the test in R is 0.78921 with degrees of freedom 4 and the p-value turned out to be 0.9399. Therefore, at 5% level of significance, we fail to reject the null hypothesis. So the fit is good for Bitcoin.

Ethereum:

The obtained test statistic after performing the test in R is 0.53435 with degrees of freedom 4 and the p-value turned out to be 0.9701. Therefore, at 5% level of significance, we fail to reject the null hypothesis. So the fit is good for Ethereum as well.

9. *Results and Discussion:* The coefficients of the ARIMA(p,d,q) model and their p values are obtained. These p values indicate the significance of the respective coefficients.

The table with the obtained values of Bitcoin is given below:

Table 1(a): Coefficients tests of ARIMA(3,1,6)

Coefficients	**Estimates**	**SE**	**z statistic**	**p-value**
ar1	-0.089606	0.118791	-0.7543	0.4506583
ar2	0.263303	0.140238	1.8775	0.0604445
ar3	0.599537	0.17389	3.4478	0.0005652
ma1	0.184199	0.49982	0.3685	0.7124781
ma2	-0.29786	0.50294	-0.5922	0.5536916
ma3	-0.905449	0.275766	-3.2834	0.0010257
ma4	-0.301932	0.348926	-0.8653	0.386864
ma5	0.180228	0.569625	0.3164	0.7517011
ma6	0.996798	0.527951	1.8881	0.0590188

The table with the obtained values of Ethereum is given below:

Table 1(b): Coefficients tests of ARIMA(4,1,6)

Coefficients	*Estimates*	*SE*	*z statistic*	*p-value*
ar1	-0.6386	0.1071	-5.9628	2.48E-09
ar2	-1.15951	0.13489	-8.5958	< 2.2e-16
ar3	-0.63215	0.12424	-5.0883	3.61E-07
ar4	-0.66751	0.13023	-5.1255	2.97E-07
ma1	0.74881	0.14166	5.286	1.25E-07
ma2	1.34101	0.15399	8.7087	< 2.2e-16
ma3	0.92153	0.18268	5.0445	4.55E-07
ma4	1.34046	0.28679	4.6739	2.96E-06
ma5	0.74818	0.24438	3.0616	0.002202
ma6	0.99955	0.20562	4.861	1.17E-06

The residuals of these models can be easily obtained using an extraction code. These residuals need to follow normal distribution. Moreover, the residual values need to be inside the control bands of ACF and PACF plots of the residuals, for each of the models to be a good one.

Ideally no point should lie outside the plots. However, a couple of points outside can be overlooked, if other diagnostic measures validate the model.

The plots for Bitcoin are given below in Fig. 1.4.

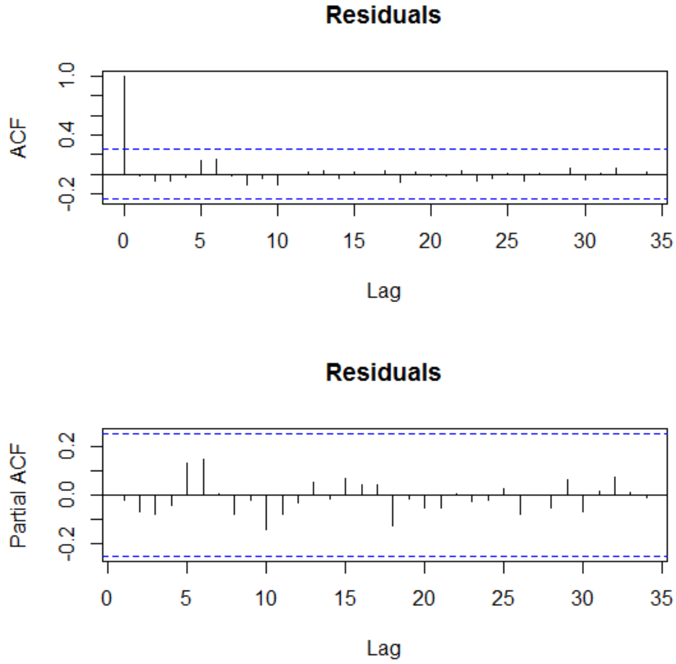

Fig.1.4: ACF and PACF plots of residuals of Bitcoin

The plots for Ethereum are given below in Fig. 2.4.

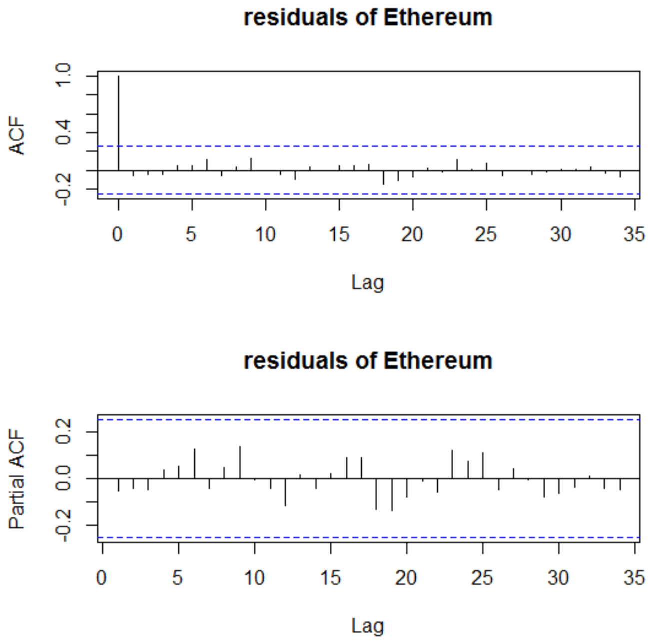

Fig.2.4: ACF and PACF plots of residuals of Ethereum

In both the figures 1.4 and 2.4, all the residuals are well within the control bands. Therefore this also indicates that the models are good fit.

Residuals are plotted using the normal probability plots to check the normality assumption of the residuals.
Figures 5(a) and 5(b) give the qqplots of the residuals of fitted ARIMA model for Bitcoin and Ethereum respectively.

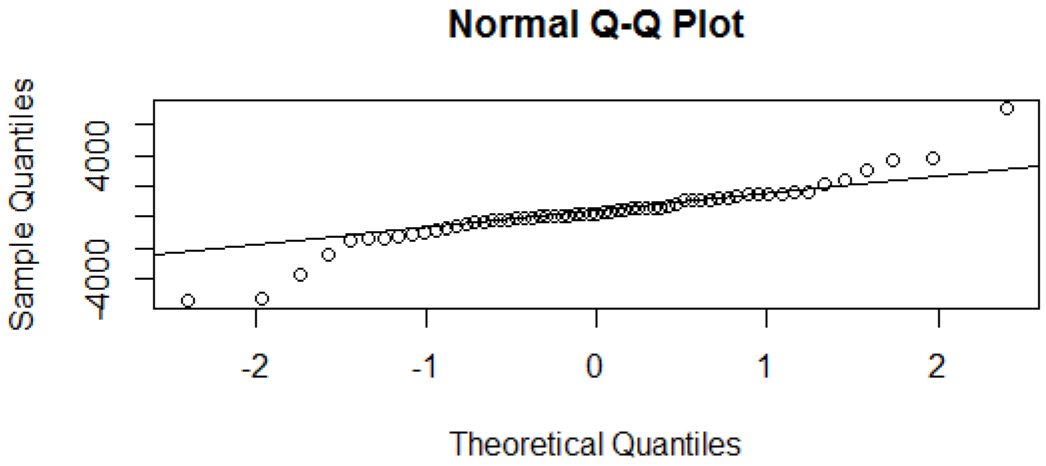

Fig. 5(a): QQLine for Bitcoin

The qqline plot shows that most of the points lie on the straight line, which indicates that normal assumptions hold true.

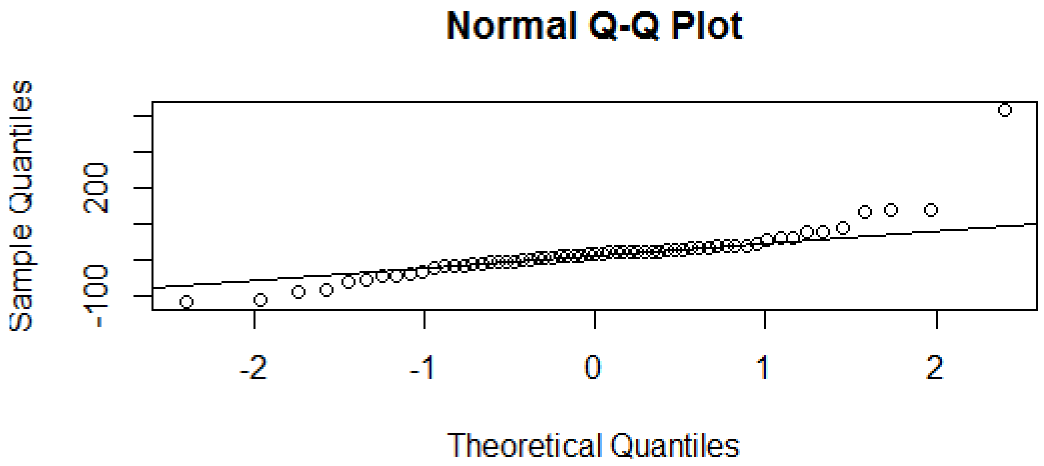

Fig. 5(b): QQLine for Ethereum

This plot shows that almost all the points lie on the qqline. This model is an even better fit and normality assumption holds true for this as well.

The Ljung Box test should give p values which are greater than the significance level at all lag points.
In this case, the p values for different lags are plotted for both Bitcoin and Ethereum.

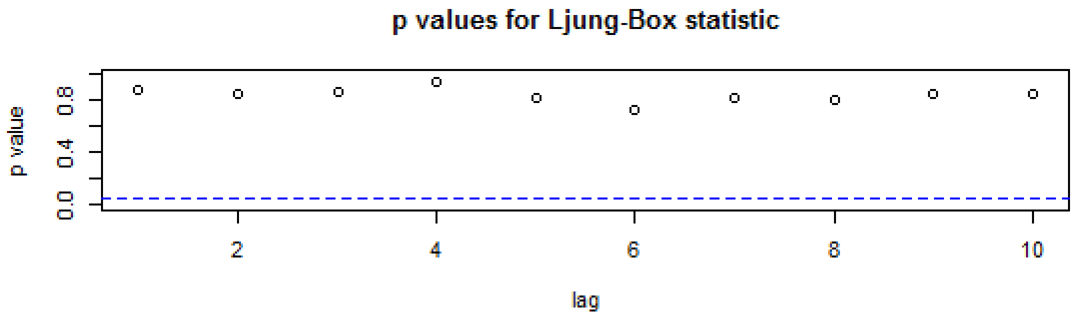

Fig. 6(a): p-values of Ljung Box Test for Bitcoin

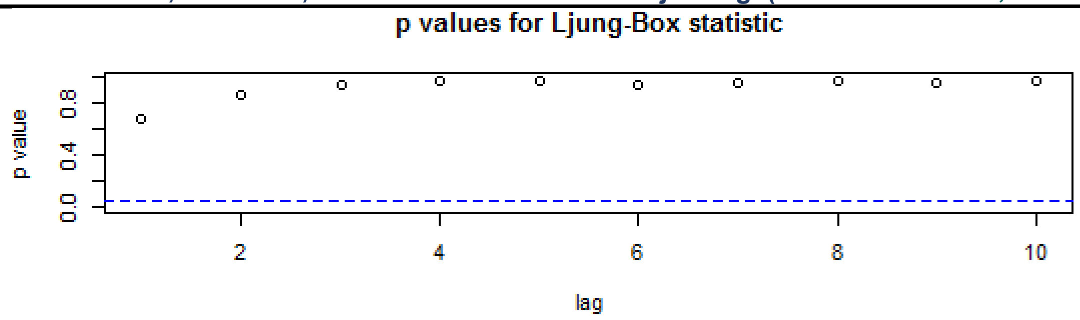

Fig. 6(b): p-values of Ljung Box Test for Ethereum

In both the figures 6(a) and 6(b) the p values are well above the significance level marked in blue line.

ARIMA fitted model with parameter estimation and diagnostic values:

The table for the parameters for Bitcoin is given in Table 1.3.

Table 1.3: Parameters of Bitcoin

Coefficients	Estimates	log-likelihood	AIC	BIC
ar1	-0.089606	-545.28	1110.56	1131.505
ar2	0.263303			
ar3	0.599537			
ma1	0.184199			
ma2	-0.29786			
ma3	-0.905449			
ma4	-0.301932			
ma5	0.180228			
ma6	0.996798			

The table for the parameters for Ethereum is given in Table 2.3

Table 2.3: Parameters of Ethereum

Coefficients	Estimates	log-likelihood	AIC	BIC
ar1	-0.6386	-351.98	725.964	749.002
ar2	-1.15951			
ar3	-0.63215			
ar4	-0.66751			
ma1	0.74881			
ma2	1.34101			
ma3	0.92153			
ma4	1.34046			
ma5	0.74818			
ma6	0.99955			

10. *Forescating/Prediction:*

The predicted values of Bitcoin price for the next 10 weeks is given in Table 4 along with their 80% and 95% confidence intervals.

Table 1.4: Forecast Bitcoin Price for the next 10 weeks

Date	**Point Forecast**	**LCL 80%**	**UCL 80%**	**LCL 95%**	**UCL 95%**
Mar 07, 2021	52342.56	49813.62	54871.5	48474.88	56210.24
Mar 14, 2021	54269.23	50535.72	58002.73	48559.32	59979.13
Mar 21, 2021	59299.04	54714.95	63883.13	52288.28	66309.8
Mar 28,2021	61782.8	56761.53	66804.07	54103.43	69462.16
Apr 04, 2021	58951.82	53710.88	64192.76	50936.49	66967.15
Apr 11, 2021	63237.61	57707.08	68768.14	54779.4	71695.82
Apr 18, 2021	63597.28	57030.39	70164.16	53554.09	73640.46
Apr 25, 2021	62996.23	55724.56	70267.9	51875.17	74117.3
May 2, 2021	65714.28	57517.08	73911.47	53177.75	78250.81
May 9, 2021	65528.1	56090.32	74965.89	51094.26	79961.95

The forecast plot is obtained using auto.plot() function in R.

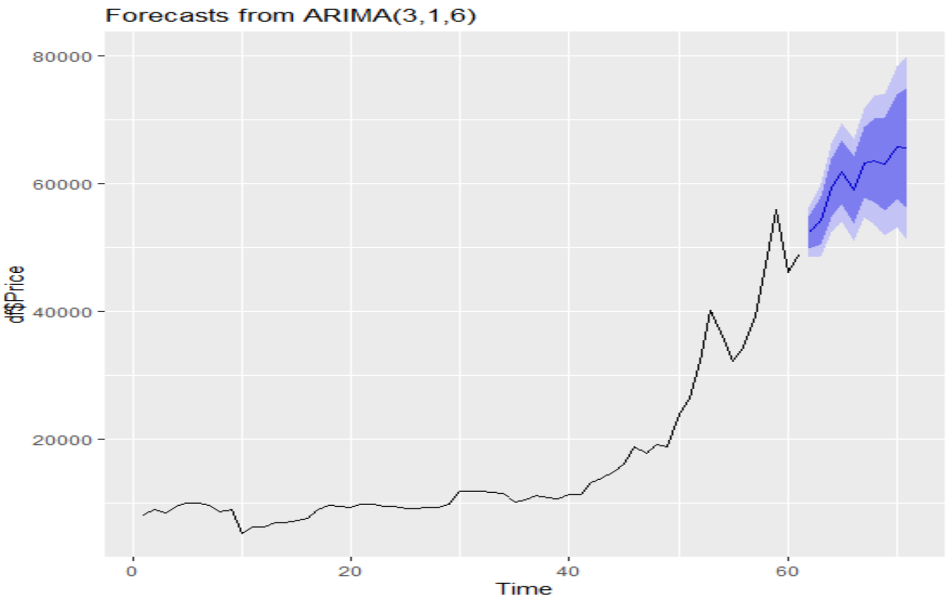

Fig. 1.7: Forecast of Bitcoin Price (in $ USD)

The shaded region in dark blue gives the 80% confidence region and the shaded region in lighter blue gives the 95% confidence region.

The forecast of Ethereum price for the next 10 weeks is given in Table 2.4 along with their 80% and 95% confidence intervals.

Table 2.4: Forecast Ethereum Price for the next 10 weeks

Date	Point Forecast	LCL 80%	UCL 80%	LCL 95%	UCL 95%
Mar 07, 2021	2037.106	1937.742	2136.47	1885.142	2189.07
Mar 14, 2021	2028.52	1880.932	2176.107	1802.804	2254.236
Mar 21, 2021	1761.712	1573.467	1949.957	1473.815	2049.609
Mar 28,2021	1644.671	1418.749	1870.593	1299.153	1990.189
Apr04,2021	1709.319	1430.463	1988.176	1282.845	2135.793
Apr 11, 2021	2032.792	1697.161	2368.423	1519.488	2546.096
Apr 18, 2021	2003.345	1608.784	2397.906	1399.916	2606.774
Apr 25, 2021	1684.337	1266.499	2102.175	1045.31	2323.364
May 2, 2021	1674.564	1240.466	2108.661	1010.669	2338.459
May 9, 2021	1853.394	1382.349	2324.438	1132.993	2573.794

Graphical representation:

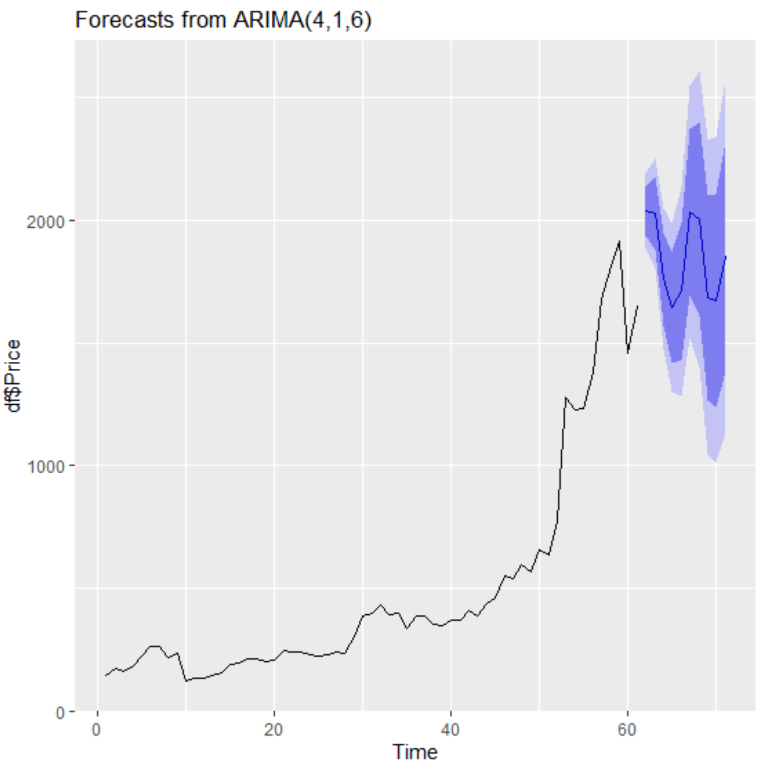

Fig. 2.7: Forecast of Ethereum Price (in $ USD)

11. *Validation :* The actual values as obtained till April 25, 2021 are available for both types of crypto-currencies. The predicted values can be validated by comparing with the actual values. If the actual values lie within the confidence interval of each of the weeks then this model is said to be appropriate.

Table 1.5: Validation for Bitcoin

Date	**Actual Price**	**Comment**
Mar 07, 2021	61195.3	outside control limits
Mar 14, 2021	58093.4	within 95% CL
Mar 21, 2021	55862.9	within 80% CL
Mar 28, 2021	57059.9	within 80% CL
Apr 04, 2021	59748.4	within 80% CL
Apr 11, 2021	60041.9	within 80% CL
Apr 18, 2021	50088.9	outside control limits
Apr 25, 2021	56504	within 80% CL

Table 2.5: Validation for Ethereum

Date	**Actual price**	**Comment**
Mar 07, 2021	1921.13	within 95% CL
Mar 14, 2021	1804.6	within 95% CL
Mar 21, 2021	1713.01	within 80% CL
Mar 28,2021	2008.59	slightly outside limits
Apr 04, 2021	2133.79	within 95% CL
Apr 11, 2021	2318.33	within 80% CL
Apr 18, 2021	2215.93	within 80% CL
Apr 25, 2021	2779.55	outside limits

Comparisons

Line charts including the Actual vs. Fitted Price for each type gives an idea about the accuracy of our prediction.

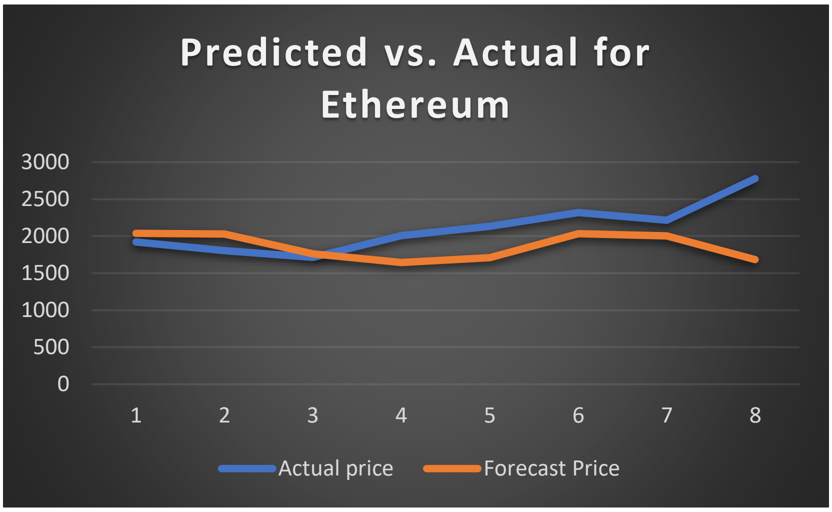

Fig. 1.8: Predicted vs. Actual Price for Bitcoin

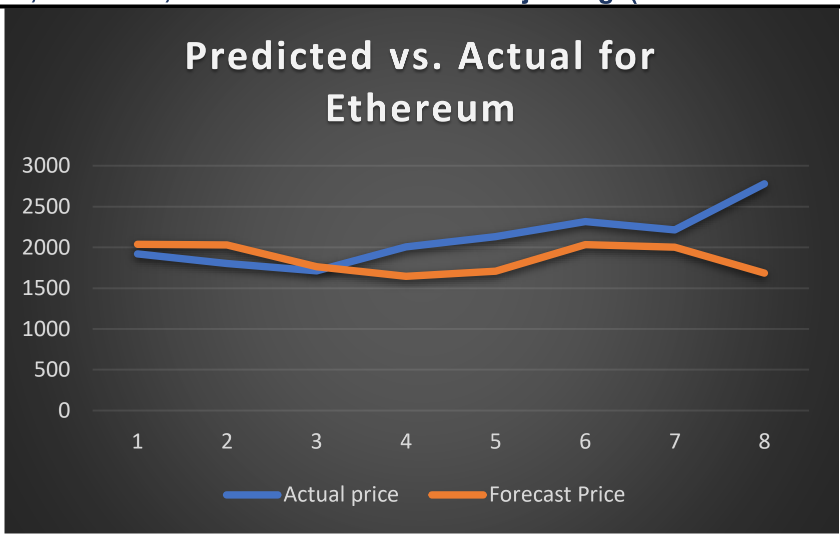

Fig. 2.8: Predicted vs. Actual Price for Ethereum

IV. CONCLUSION

There has been a hike in Bitcoin prices in the world over the weeks. Due to increasing popularity, more people are investing in cryptocurrencies. Bitcoin being the first ever blockchain introduced and made popular, it has seen its popularity increase at a great rate. Given the current market volatility in other types of currencies, cryptocurrency is a big project for the upcoming years. A similar work on Bitcoin price over the years can also be performed on available data in order to increase the forecast time limit.
For the project which has been done here, it can be clearly seen that an almost accurate prediction of the prices are obtained using the model. The price increases, but not with a constant rate. Some weeks also see a fall in price, although not by a great margin. It can be safely said that the price and popularity will continue growing as of now.

A recent dip in the Bitcoin prices have been observed in the market and Ethereum and other unpopular cryptocurrencies are getting popular. This shows that like any other market, crypto market is also volatile and long term study should be carried out to obtain proper forcasting.

A similar growth in popularity of Ethereum is also observed from our results. Although Ethereum does not seem to be as popular as Bitcoin, it has also seen a hike in its price over the weeks for the past year. The difference between Bitcoin price and next most popular crypto-currency --- Ethereum's price tells us about the margin of difference between the two. If cryptocurrencies are made legal worldwide, Ethereum might give a tough competition to Bitcoin in terms of popularity in the future.

Cryptocurrencies, like any other form of currency, have a number of threats like people can make illegal transactions, no one keep a track on the amount being transacted and hence the government and other organizations cannot make use of these transactions, which will uplift the tax system.
But cryptocurrencies have a lot of good it in. Most of the fraudulent activities can be avoided. People would have to pay an unnecessary charge for making transactions as we are expected to do in our current situations.
All that we hope is that the ban on transaction of cryptocurrency in India is uplifted and people can explore and research more on this topic.

V. ACKNOWLEDGMENT

Throughout the writing of this dissertation I have received a great deal of support and assistance.

I would first like to thank my supervisor and guide, Professor Tinni Chaudhuri, whose expertise was invaluable in formulating the research approach and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would also like to thank my tutors, Professor S. Banerjee and Dr. B. Guha, for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me. Finally, I could not have completed this dissertation without the support of my friends, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

REFERENCES

[1] Forecasting wheat production in India: An ARIMA modelling approach by **Bhola Nath, DS Dhakre and Debasis Bhattacharya.**

[2] Forecasting Area, Production and Yield of Maize in India using ARIMA Mode by **Pradeep Mishra, Jawaharlal Nehru Krishi Vishwavidyalaya and Pradip Kr. Sahu, Bidhan Chandra Krishi Viswavidyalaya.**

[3] Modeling and Forecasting of Rice Production in Some Major States of India using ARIMA by **Tinni Chaudhuri, Prashant Verma and Mukti Khetan.**

[4] ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq by **Saleh Zakaria, Nadhir Al-Ansari, Sven Knutsson and Thafer Al-Badrany.**

[5] Forecasting weekly reference crop evapotranspiration series by **S. MOHAN & N. ARUMUGAM.**

[6] Weekly Hotel Occupancy Forecasting of a Tourism Destination by **Muzi Zhang, Junyi Li Bing Pan and Gaojun Zhang.**