Analyzing Cryptocurrencies with Time Series

MTH517A: TIME SERIES ANALYSIS

Course Project

Instructor: Professor Amit Mitra

Submitted by:

Ashu Prakash (14146)
Kushal Kumar (14346)
Srishti Agarwal (14707)
Subham Kumar (14710)
Tushar Tiwari (14765)

Abstract:

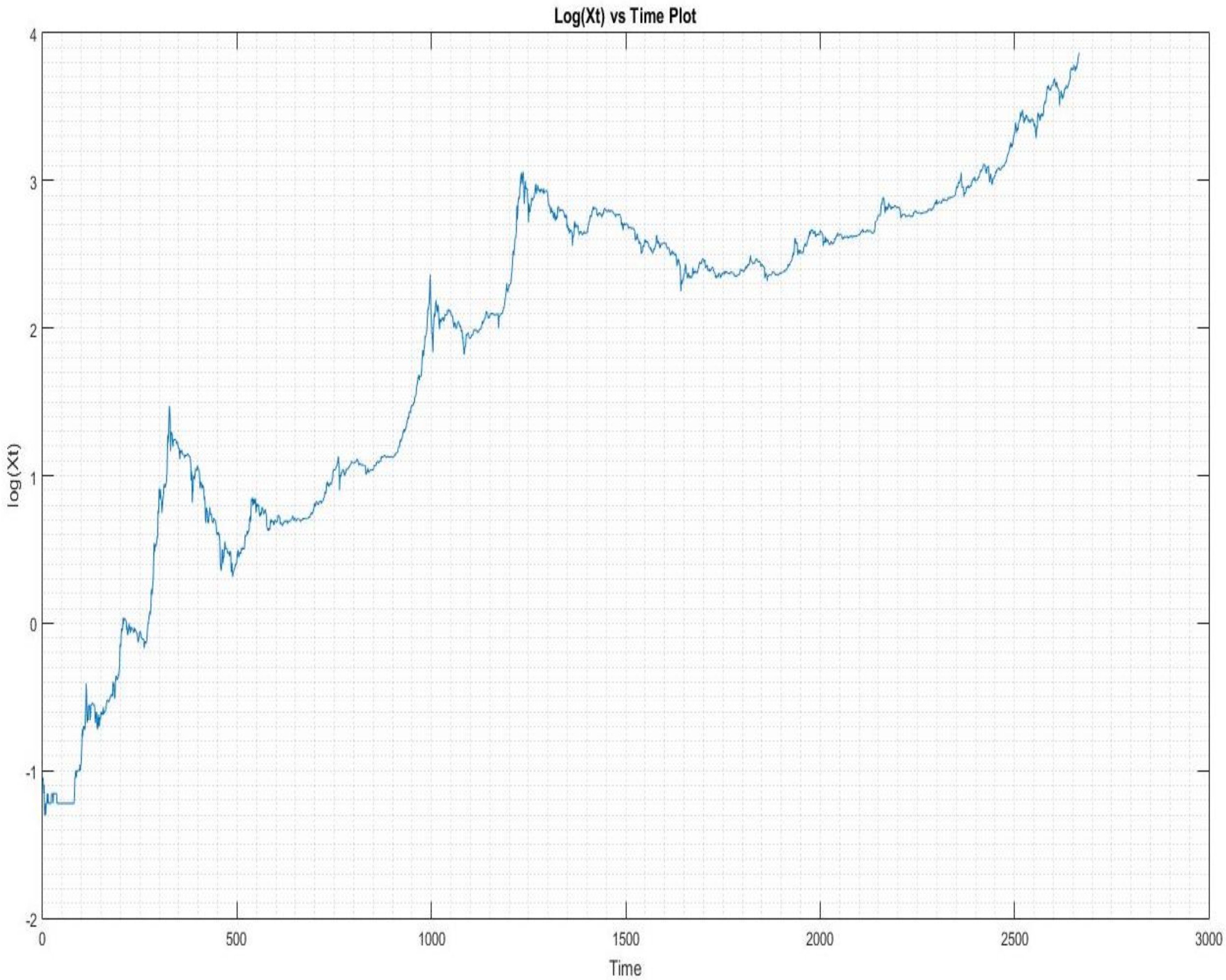
Cryptocurrencies are ruling the world market these days. Recently, the most famous cryptocurrency ‘Bitcoin’ has raised its market share to more than 200 billion USD. More and more people are investing in Bitcoins and are using it as a way of transaction. Currently, the Bitcoin prices (BTC) are sailing as high as 7000 USD. In this report, we have analysed the BTC vs USD data from 18th July 2010 to 4th Nov 2016 using different Time Series techniques. This time series method may be used to forecast Bitcoin or any other cryptocurrency prices in the future.

Introduction:

In this project, we have tried to analyse the prices of Bitcoin in a 6+ years window. We are working on the logarithmic data of BTC vs USD. The trend and seasonality of the logarithmic data are analysed, estimated and are removed. Our fitted model is an S-ARIMA model, in which we remove trend and seasonality making it an effective ARMA model. We estimate the order of this ARMA model and its coefficients such that it best fits our data.

Data:

Xt: Historical Time Series of Bitcoin prices in USD from 18/07/2010 to 04/11/2016.
Frequency: Daily (No. of data points: 2667)
Define: Yt = log(Xt)
Following is the graph of Yt vs time:



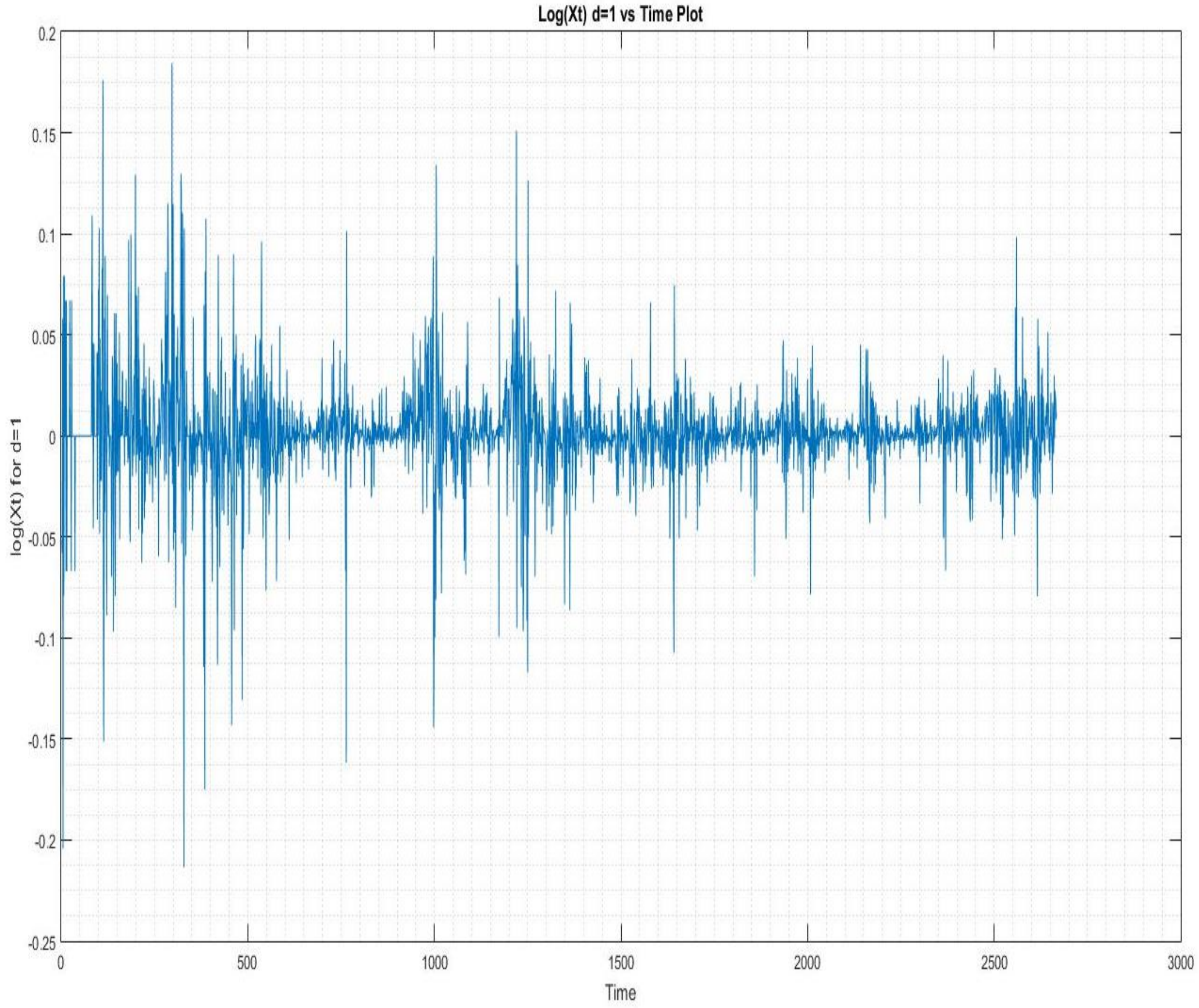
Estimation and Elimination of trend:

We have used the Relative Ordering Test for testing the presence of trend component. The results of Relative Ordering Test are: (Here d is the order of differencing)

For d=0, Computed value of Q/E(Q) = 0.23 (this shows rising trend)
For d=1, Computed value of Q/E(Q) = 0.9901
For d=2, Computed value of Q/E(Q) = 0.9971

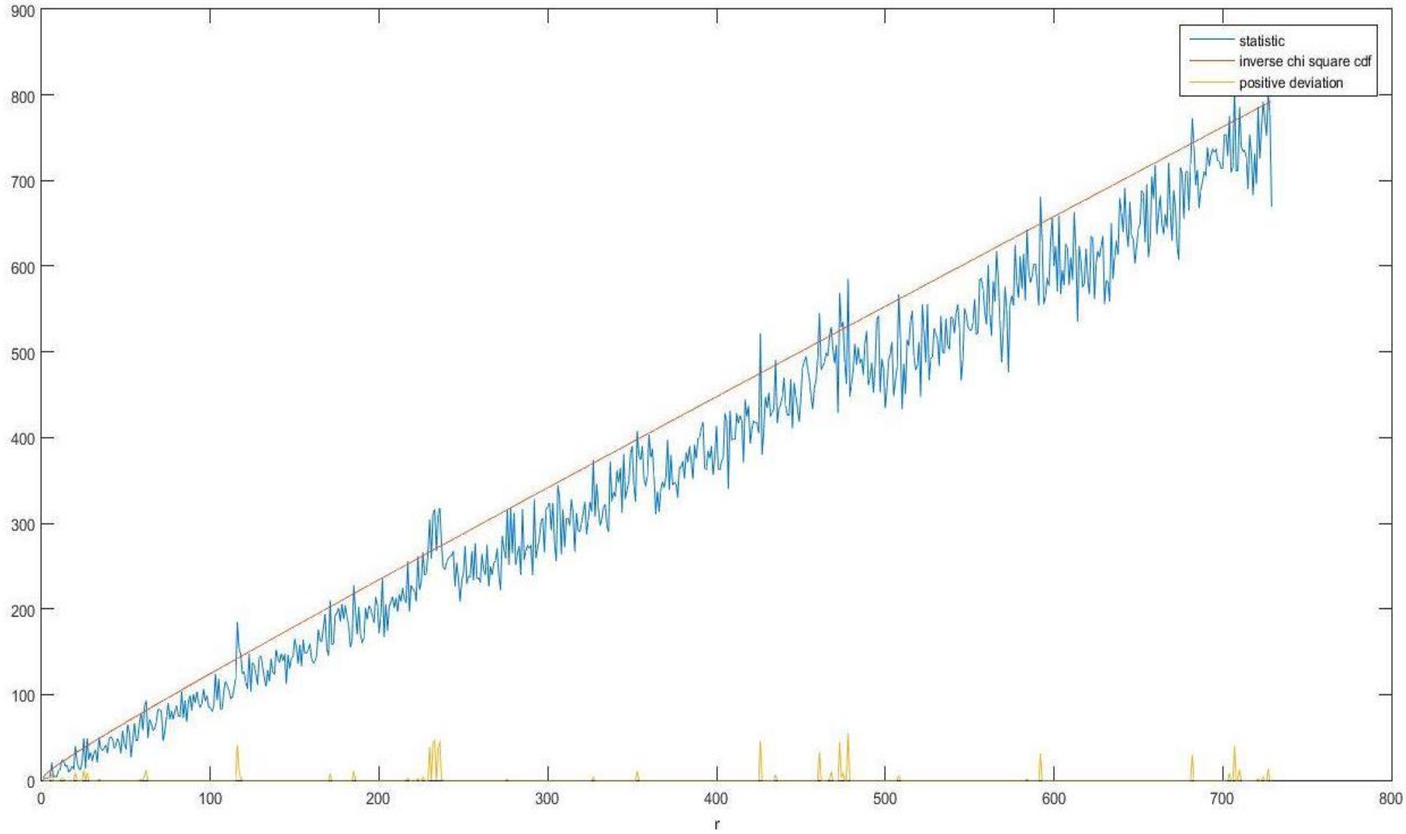
Hence, the trend gets eliminated for d=1, since at d=1, Q from the data equals the expected value of Q, indicating the acceptance of null hypothesis and the acceptance of no trend. Hence we conclude that the trend is linear.

We have used method of differencing (order of differencing, d=1) for elimination of trend.
Hence, Y(t) – Y(t-1) = (1-B)Y(t) = ∨Y(t)
Following is the graph of Yt vs time after the elimination of trend:



Estimation and Elimination of Seasonality:

We have used Friedmann's test for testing the existence of seasonality in the time series. We plot a graph with the value of test statistic and the value of chi square (with r-1 degrees of freedom on the y axis and different possible values of seasonality (d) on x axis.

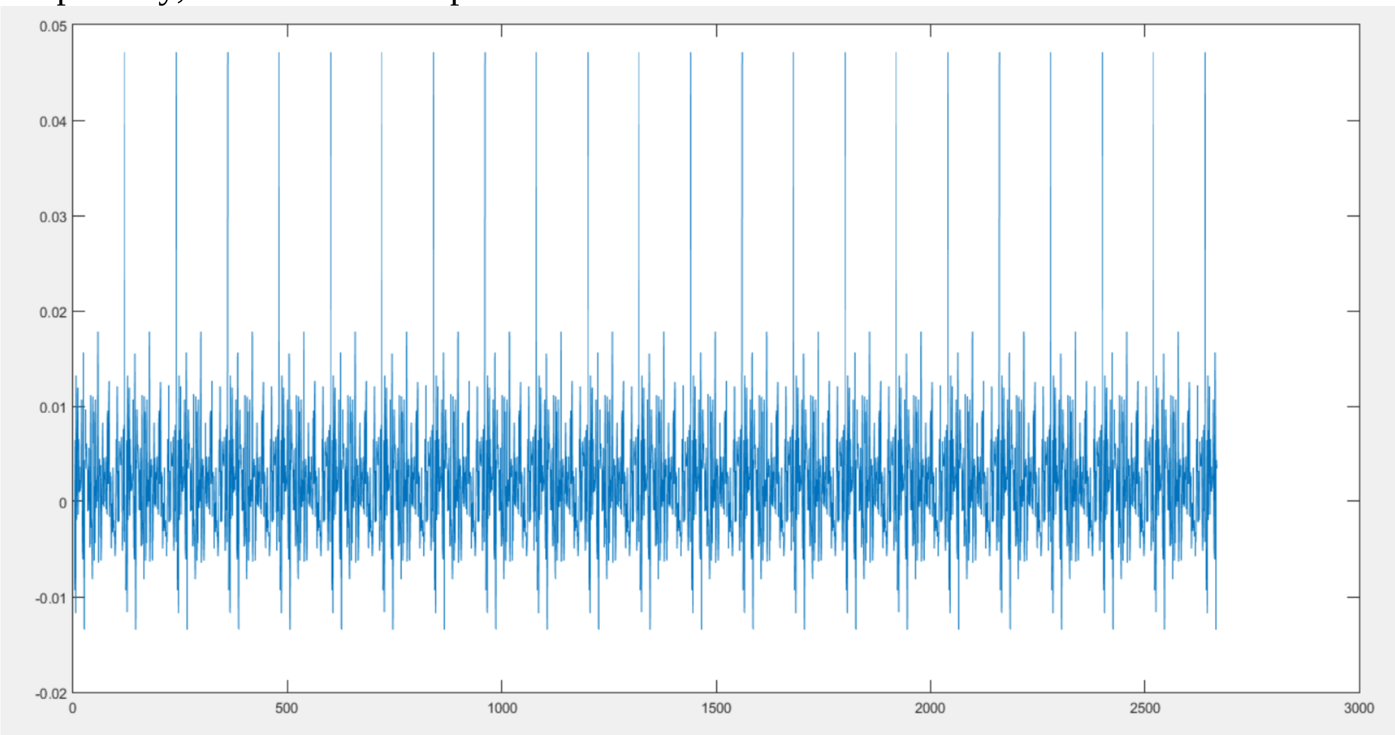


We decide the period of seasonality as the one at which the test statistic value is greater than the chi square value. At this value of d and at all the multiples of d, if the value of test statistic is greater than chi square, then we confirm d to be the period of seasonality for this time series. For this data, the period of seasonality comes out to be equal to 120 days.

The results of Friedmann's test are:

Test Statistic	Chi Square	d
146.5674	71.1298	120
277.1376	224.1594	240
405.2435	354.0829	360

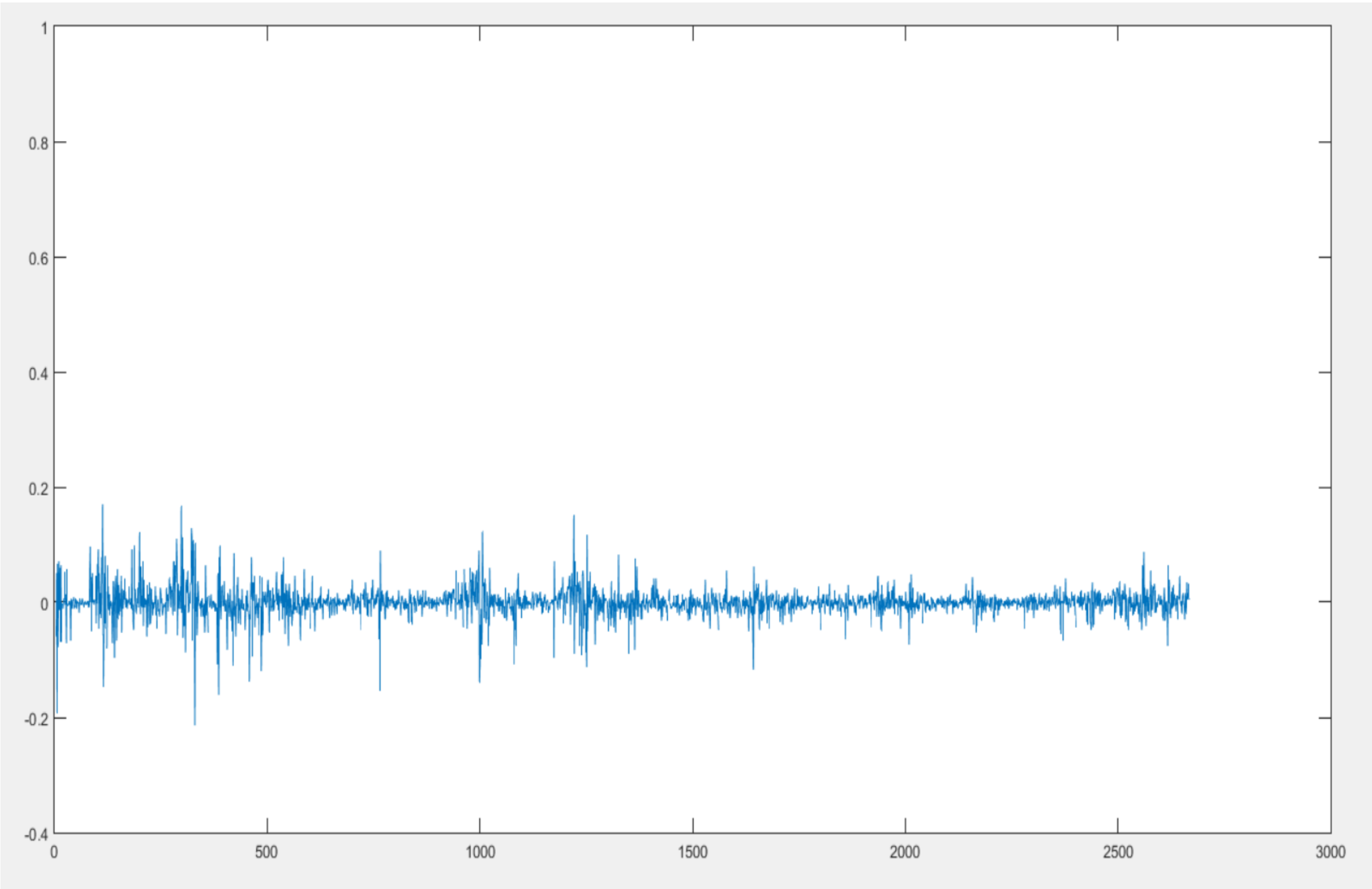
Graphically, the seasonal component is:



We have estimated the seasonal component as follows:
Since the seasonality is of order d=120 days, we take average of X1, X121, X241… and conclude it to be the seasonal component corresponding to X1, X121, X241.... Similarly we take average of X2, X122, X242…. and conclude it to be the seasonal component corresponding to X2, X122, X242…..

In the Friedmann’s test, if the value of test statistic becomes equal to the value of Chi Square, then we say that seasonality is not present. Hence after removal of seasonality, we again apply Friedmann’s test for existence of further seasonality and the results obtained prove that no further seasonality is present since test statistic values were close to chi square values.

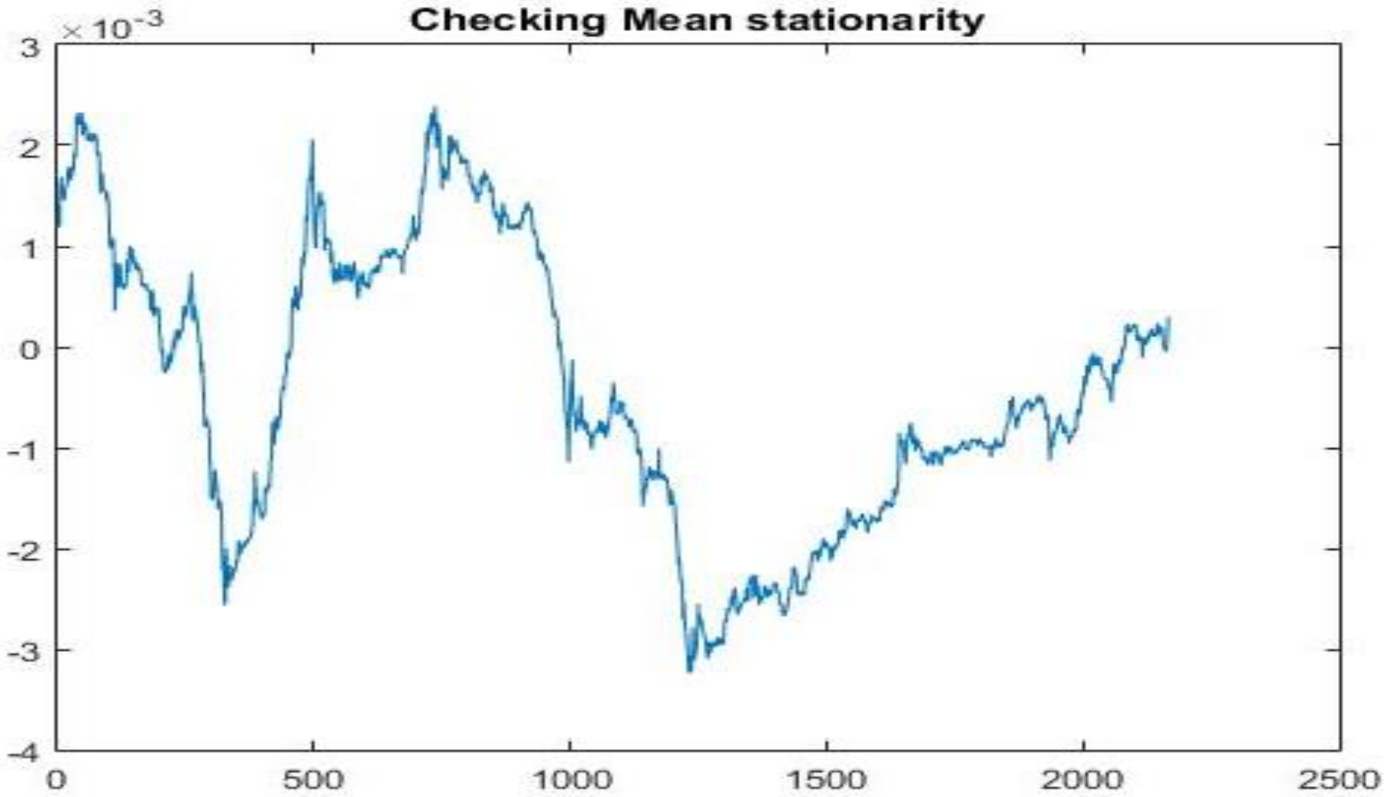
After estimation of this seasonal component corresponding to all X’s, we eliminate it from the data to make it completely random.
After trend and seasonality elimination, the data looks like:



Check for stationarity:

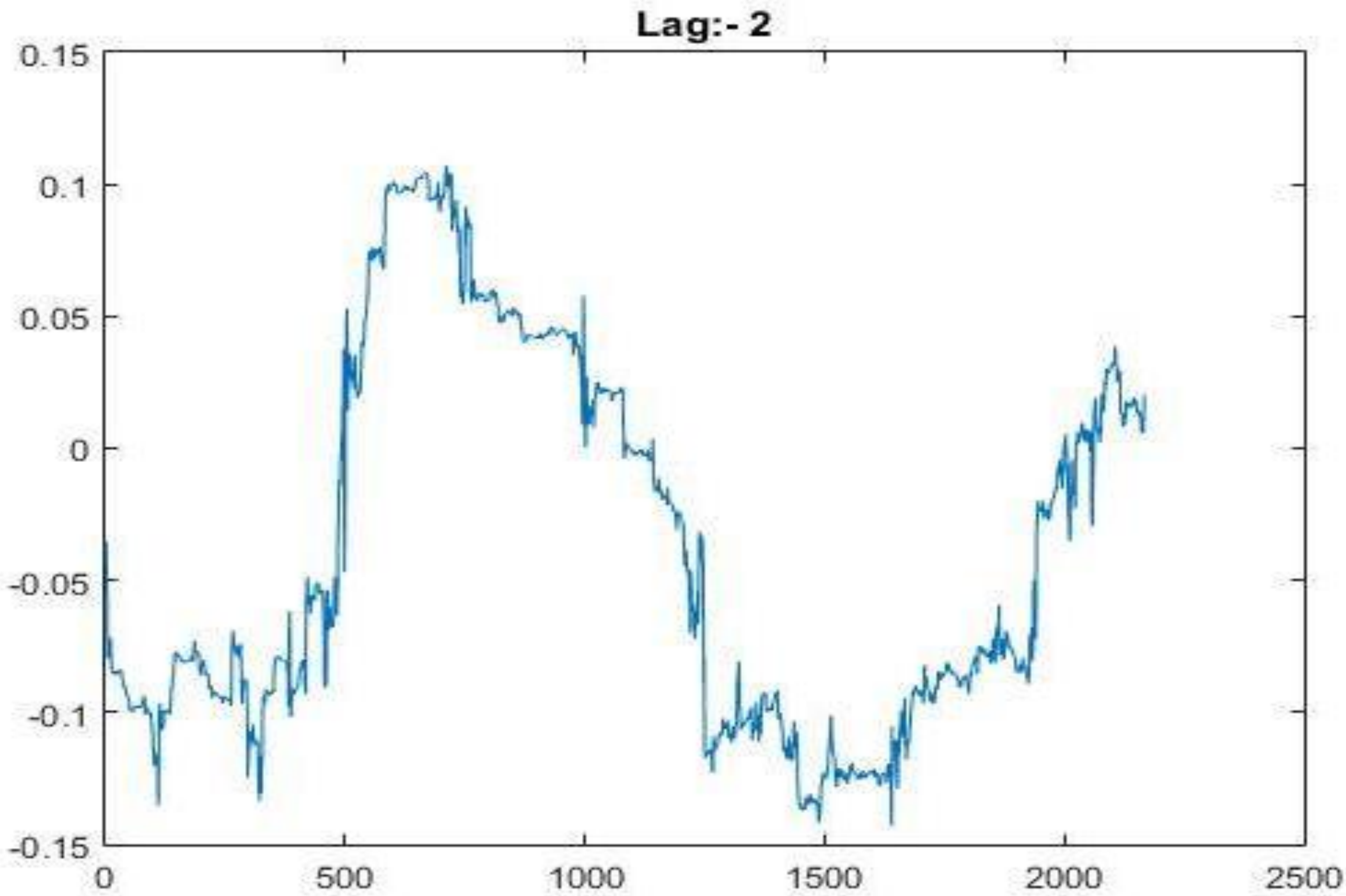
We check the above obtained random data for existence of mean and covariance stationarity. For mean stationarity, we take different intervals of same length and calculate the mean for all those intervals of same length.

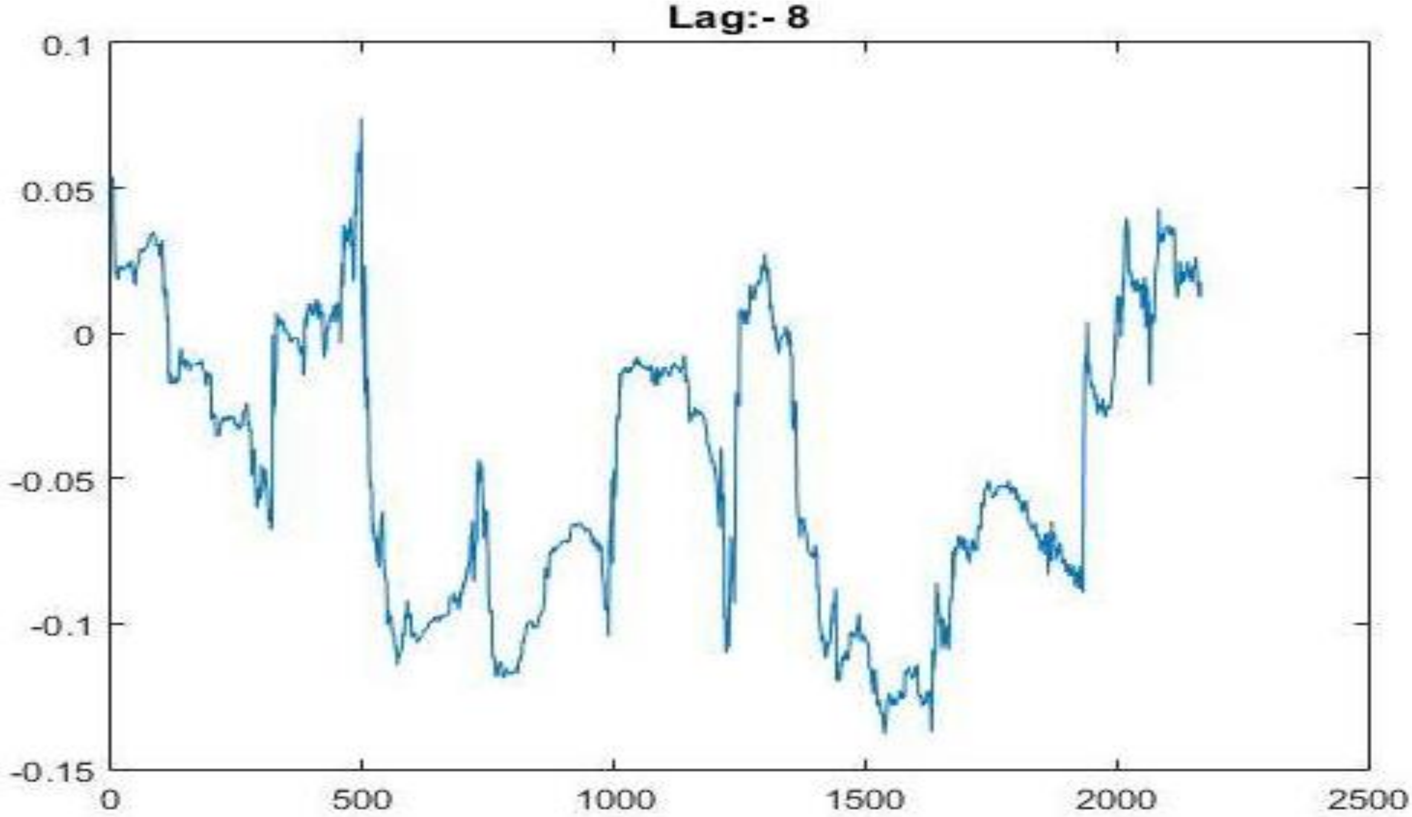
The graph below shows that this random process generated in mean stationary, since the variation in mean is of the order 10^-3 which can be neglected.



For covariance stationary, we select a lag and fix the length of an interval. We measure covariance between these two intervals of fixed length. Similarly we measure covariance between another two intervals of same length and for the same value of lag.

For lag = 2 and lag = 8, the covariance graphs are as follows:

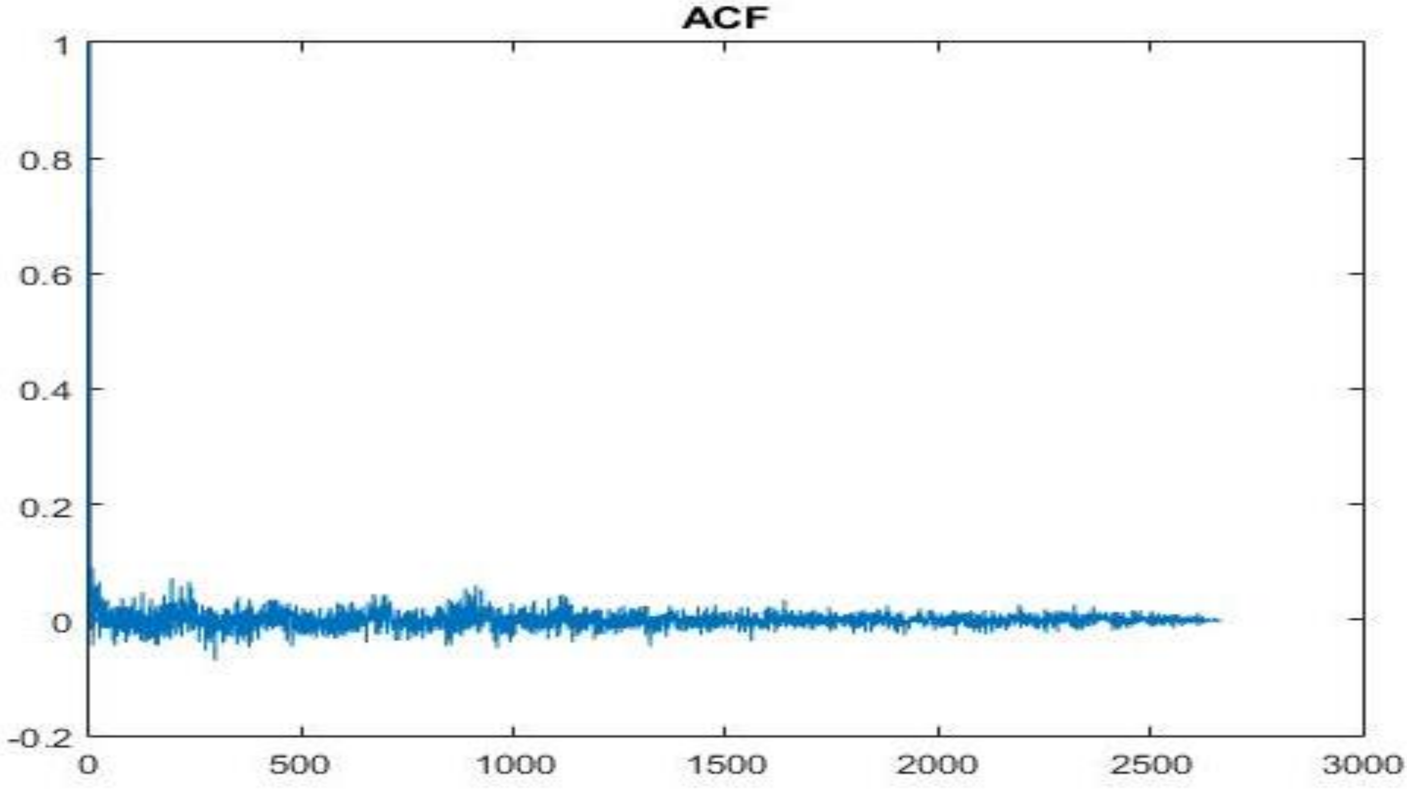




We have plotted this graph for various values of lag i.e lags from 1 to 10 and with interval of size 500. But for all the values of lags the variation in the ACF is considerable and hence cannot be neglected. As the values of covariance are not same for different same length intervals with the same lag, this proves that the random component obtained after removal of trend and seasonality is not covariance stationary and hence not weak stationary.

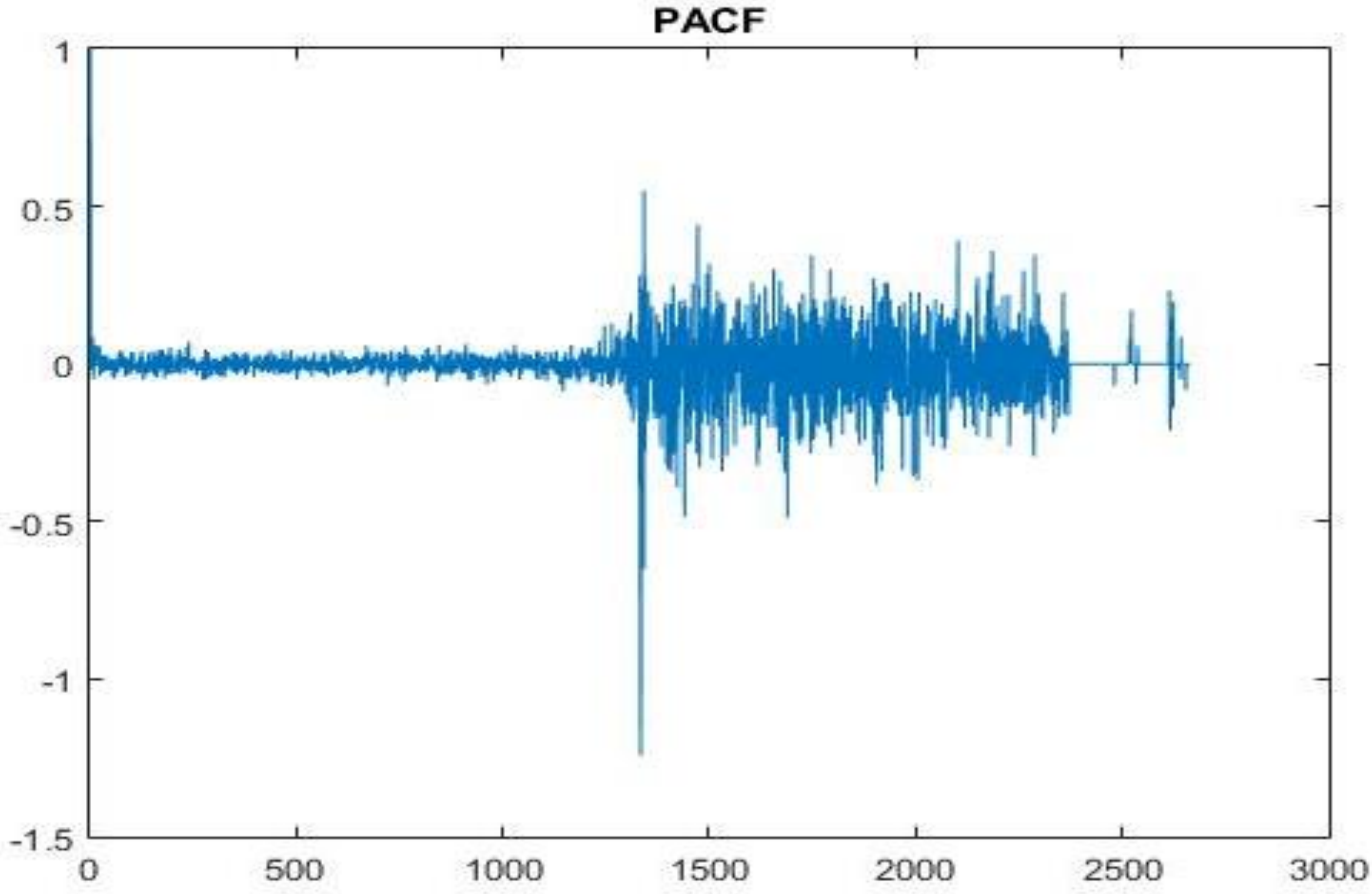
Model Selection

For selecting the model fits our data, we plot the Auto-Correlation function for the data.



Since ACF tails off as lags increase, we can conclude that ARMA model can be fitted to this data.

The plot of PACF is:



We observe that PACF also tails off as lags increase, hence an ARMA(p,q) model can be fitted.

Order Estimation

For estimating the order of ARMA process i.e., for estimating the values of p and q, we have used Bayesian Interpretation Criterion. The range of p for estimation is kept to be [1,20] and the range of q for estimation is kept to be [1,20]. For all possible different combinations of p and q selected from this range we calculate the values of BIC. The combination of p and q for which the value of BIC is minimum is the order of ARMA.
The estimated order of ARMA process is (7,6) (p=7, q=6). The process is ARMA(7,6).

Parameter Estimation

For estimation of parameters (coefficients) of this ARMA(7,6) model, we use Conditional Maximum Likelihood Estimator.

Hence the forecasted value of X at some future time t is:

Xt = -0.744X(t-1) + 0.0165X(t-2) + 0.1366X(t-3) + 0.044X(t-4) + 0.421X(t-5) + 0.6946X(t-6) + 0.026X(t-7) + **ϵ**t + 0.792**ϵ**(t-1) - 0.0107**ϵ**(t-2) -0.167**ϵ**(t-3) -0.054**ϵ**(t-4) - -0.362**ϵ**(t-5) - 0.59**ϵ**(t-6)
where **ϵ**t ~ WN(0, 0.00058)

We solved the polynomial of AR part of the ARMA process to find roots of that polynomial, which came out to be greater than 1, hence we prove that the process indeed is non stationary.

Conclusion

From this project, we have learnt how to handle real life data *esp;* financial data and that too the most famous cryptocurrency ‘Bitcoin’ which is too unpredictable. The raw BTC vs USD data was so aleatory that we decided to use the logarithm of it. It solved our purpose, by having some trend and seasonality. We also learnt to estimate and eliminate trend and seasonality from a real data. We have used classical ARIMA to find that the log of the BTC vs USD prices follow an ARIMA(7,0,6) which is same as ARMA(7,6). Thus, we have an estimate of the variation of the data for which we may use this model to predict the future Bitcoin prices providing no market booms or crashes. Our designed model forecasts the value of Bitcoin price at some future time t.