AN ANALYSIS OF BITCOINS

Price Analysis, Volatility & Forecasting

School of Information Studies - Syracuse University

Financial Analytics (FIN 654)

Submitted By:

Aishwarya Nagaraj

Divisha Khandelwal

Sapan Badjatiya

Table of Contents

1. Introduction:

There are different cryptocurrencies like Bitcoin, Ethereum, Litecoin and so on. Bitcoin is a digital currency with market cap of more than $100 billion. It works without a central repository/single administrator. Peer-to-peer transactions take place between users directly, using cryptography. These transactions are verified by network nodes and recorded in a public distributed ledger called a blockchain. Bitcoin was invented by an unknown person or group of people under the name Satoshi Nakamoto and released as open-source software in 2009. Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services.

Blockchain was originally developed for the sole purpose of tracking Bitcoin transactions. For the last couple of years, it has unexpectedly evolved into something much more. Many developers are exploring new ways to use blockchain technology. Blockchain will probably actually have a much bigger impact on our world than Bitcoin itself because it can be used for many everyday purposes that will impact people that don't even use Bitcoin. This was the motivation for analysis on this dataset.

2. Overview:

In this project, we will analyze the prices for different cryptocurrencies. We will also analyze the correlation between Market Returns and Bitcoin Returns. We will be investigating several types of volatility models and measures. We will study the volatility models and will estimate future volatilities and correlations.

3. Data Description:

First dataset:

•Date: 2013 Apr 29th– 2017 Sep 29th formatted as Year-month-day

•MarketReturn: Overall return of market portfolio

•SP500_Price_Index: Weighted average market capitalization

•Bitcoin_Price: Daily price of Bitcoin

•BitcoinReturn: Daily return of Bitcoin

•Bitcoin_VIX: Volatility of Bitcoin price

Second dataset:

•Date: 2017 Aug 3– 2017 Nov 28th formatted as Year-month-day

•Ether_Price: Daily price of Ethereum

•Ether_Change: Percentage change of daily price

•Bitcoin_price: Daily price of Bitcoin

•Bitcoin_Change: Percentage change of daily price

•Litecoin_Price: Daily price of Litecoin

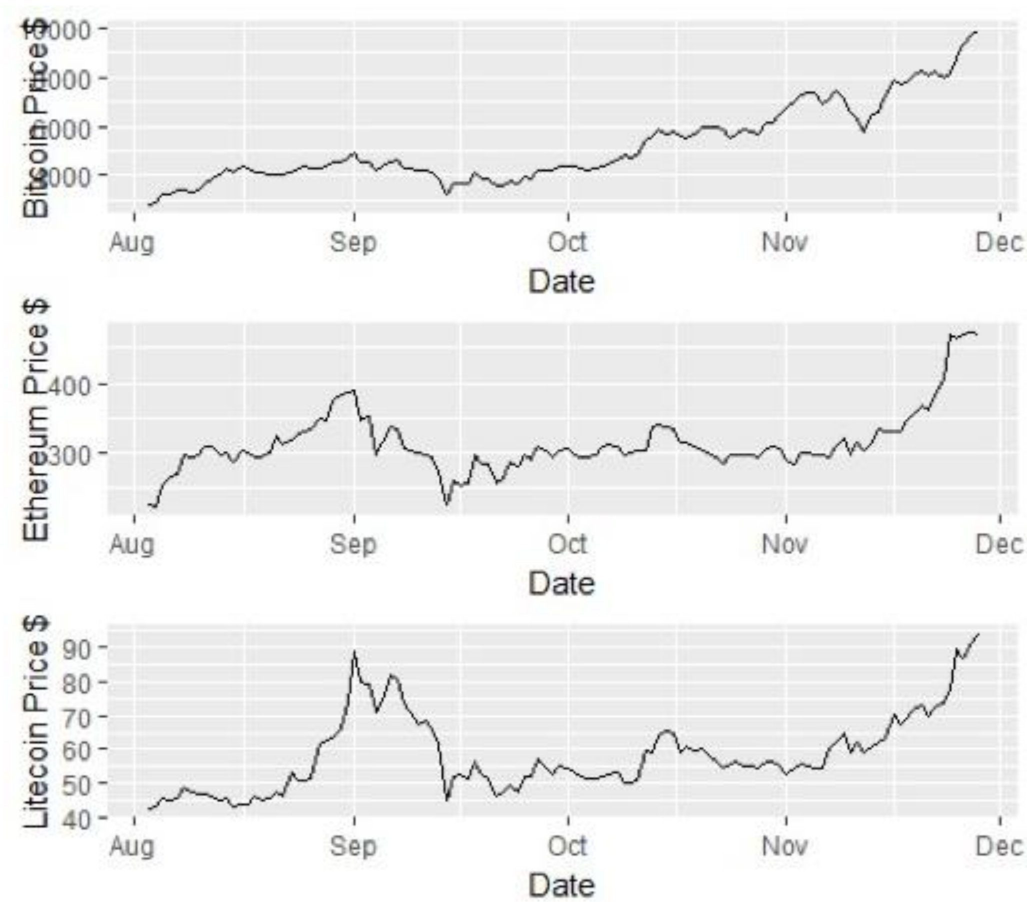
•Litecoin_Change: Percentage change of daily price

Data Sources:
https://finance.yahoo.com/
https://www.investing.com/
https://www.cryptocompare.com/

4. Objectives:

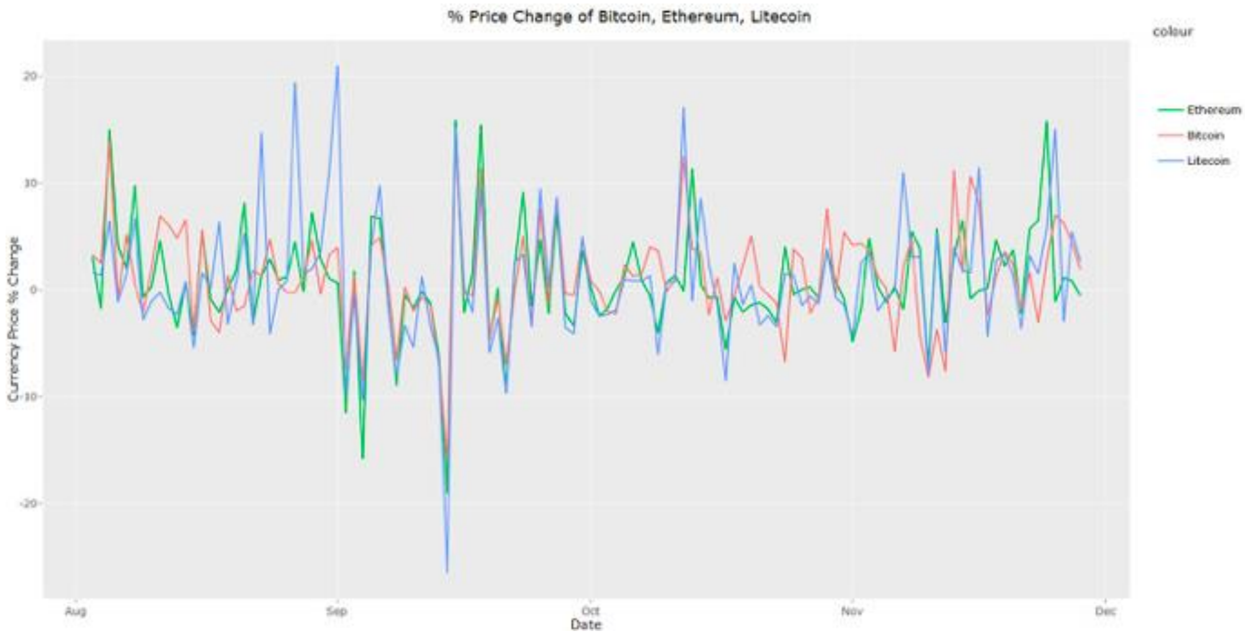
4.1. Objective 1: Price Analysis

a) Cryptocurrency price variations: Bitcoin, Ethereum and Litecoin



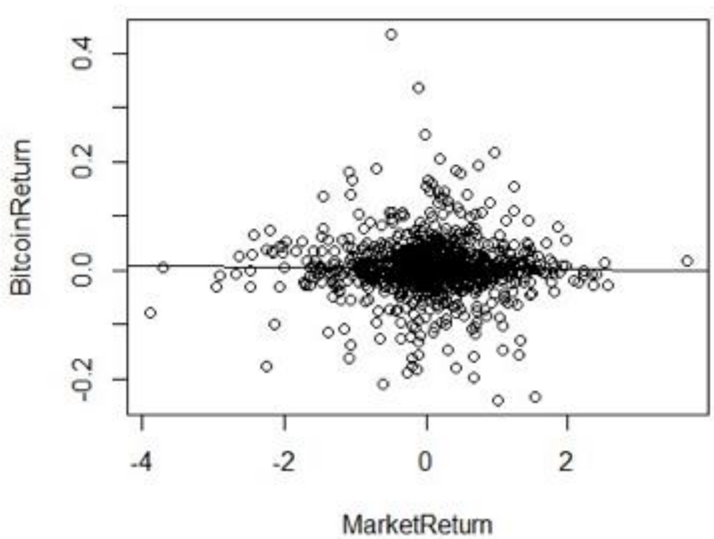
Interpretation:
The above image has three graphs which showcase the variations of Bitcoin, Ethereum and Litecoin over the months of Aug-Dec 2017.The X axis represents the Date and the Y axis represents the Bitcoin price. We can see that the three cryptocurrencies follow a similar trend for the most part. Thus, the variations in the prices are similar. To further our analysis, we plotted the percentage change in the values of the three cryptocurrencies in one plot. We confirmed that the variations between these three cryptocurrencies are very similar.

b) Percentage variation of cryptocurrencies



4.2. Objective 2: Analyzing the Correlation

Correlation between Market Returns and Bitcoin Returns:



```
Call:
lm(formula = Bitcoin$MarketReturn ~ Bitcoin$BitcoinReturn, data = BitcoinVol)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9717 -0.3642  0.0057  0.4309  3.6296 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)  
(Intercept)            0.05433    0.02370   2.292   0.0221 *
Bitcoin$BitcoinReturn -0.21904    0.46358  -0.472   0.6367  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7886 on 1113 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.0002005,	Adjusted R-squared:  -0.0006978 
F-statistic: 0.2232 on 1 and 1113 DF,  p-value: 0.6367
```

Interpretation:

This plot shows the correlation between Market return and the Bitcoin return. The X axis shows the Market return and the Y axis shows the Bitcoin return. We cannot see a parabola or a linear line in the plot. So, we conclude that they do not have a relationship. Thus, Bitcoin Return is not affected by changes in the Market Return.

We also plotted a linear model for Market Return and Bitcoin Return. P-value = 0.6367 which is greater than 5%. Hence, p-value is not significant, and both the returns do not hold a linear relationship.

```
> cor(MarketRET, BitcoinRET)
[1] -0.01416128
>
```

We confirmed our conclusion by performing the correlation of the two variables. The value obtained is -0.01416128, which is a very low value.

Thus, by including Bitcoin in their portfolio, investors can obtain a much well diversified, less risky investment.

4.3. Objective 3: Normality Tests

Testing if Bitcoin Returns follow normal distribution-

a) Jarque Bera Test:

The Jarque Bera test is a test for normality.

For Bitcoin Return:

S1: Hypothesis:

H0 (Null hypothesis): Bitcoin return comes from a normal distribution.

H1(Alternative Hypothesis): Bitcoin return does **not** come from a normal distribution.

S2: The p-value for Bitcoin return generated from computer is less than 2.2e-16.

S3: Decision Rule: Reject the null hypothesis when p-value <5%.

S4: Conclusion: Here we can see that, the p-value < 2.2e-16 of Bitcoin return is much lesser than 5% (alpha, the significance level), thus we reject the null hypothesis.

*This means that H1 is valid and Bitcoin return is **not** a normal distribution.*

b) The Lilliefors Test:

The Lilliefors test is a test for normality which is usually done when we do not know the mean or standard deviation. The test assumes that we have a random sample.

For Bitcoin Return:

S1: Hypothesis:

H0 (Null hypothesis): Bitcoin return comes from a normal distribution.

H1(Alternative Hypothesis): Bitcoin return does **not** come from a normal distribution.

S2: The p-value for IBM excess return generated from computer is less than 2.2e-16.

S3: Decision Rule: Reject the null hypothesis when p-value <5%.

S4: Conclusion: Here we can see that, the p-value 2.2e-16 of Bitcoin return is much lesser than 5% (alpha, the significance level), thus we reject the null hypothesis.

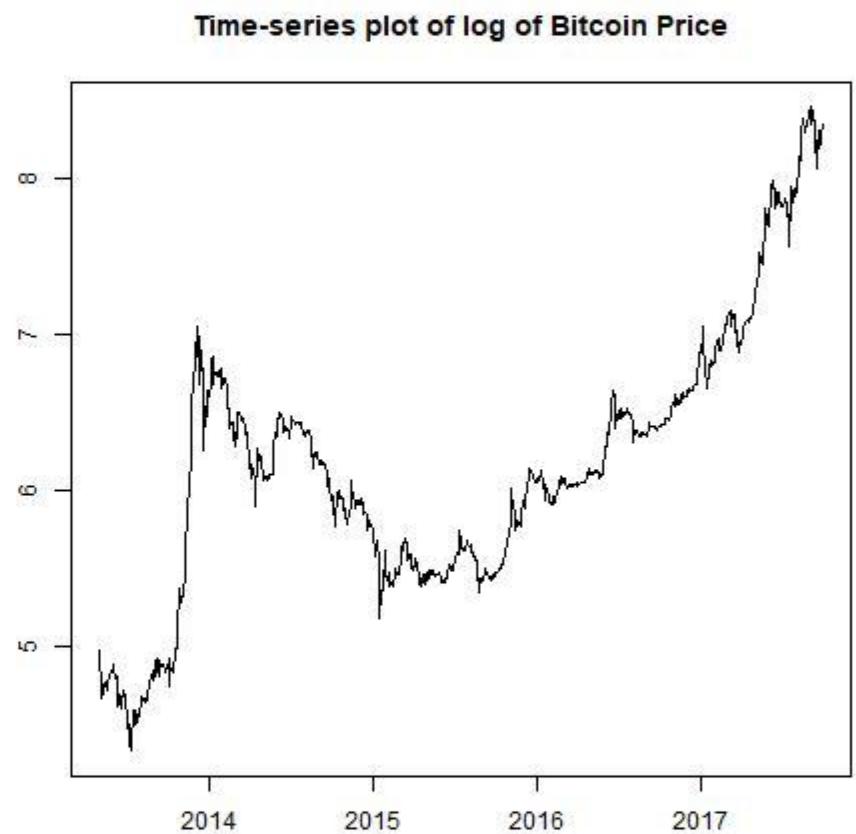
*This means that H1 is valid and Bitcoin return is **not** a normal distribution.*

4.4. Objective 4: Volatility and Forecasting

Studying the volatility models and estimating the future volatilities and correlations

Visualizing Data

1. Time series plot of the Bitcoin Prices (FIG)

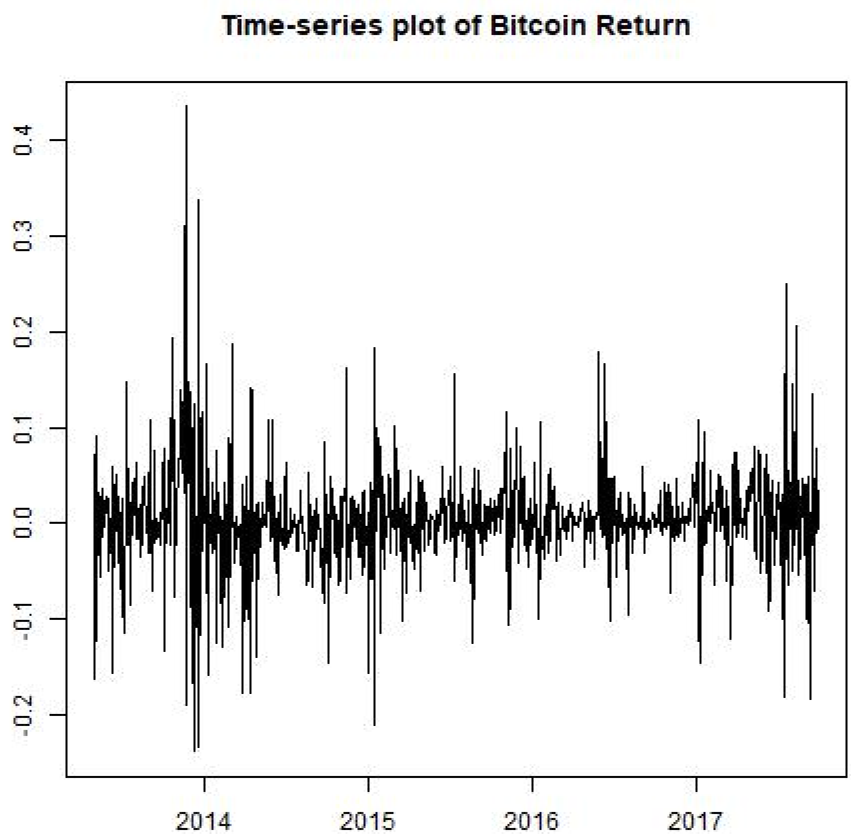


Interpretation: The above figure is time series plots for the log price of Bitcoin Price. The X-axis is the Date range from 2013-04-29 to 2017-09-29. The Y-axis is the log price of Bitcoin Price.

Here in this plot we observe that price goes up in 2014. We think this price rise is a combination of the increased usage and the no-VAT ruling in Europe, the end of the auctioning of Silk Road bitcoins, etc. In 2017, the price again has started to go up. Earlier this month, the US-based Commodity Futures Trading Commission (CFTC) allowed two exchanges to start trading in Bitcoin contracts. News reports suggest the CME Group, and CBOE Global Markets exchanges are expected to offer Bitcoin futures later this month, with rumors of even the New York-based NASDAQ offering the same. Another major reason for high can be attributed to the successful beta test of Lightning Network, an overlay network built on top of an existing Blockchain – in

this case, the Bitcoin Blockchain. It is a decentralized network using smart contract functionality in the Blockchain to enable instant payments across a network of participants.
The maximum point is 8.455692 on 2017-09- 01. The minimum point is 4.330733 on 2013-07-09.

2. Time series plot of Bitcoin Returns:



Interpretation: The plot drawn above, shows the Bitcoin return over the years. In this plot, X-axis represents the Date range from 2013-04-29 to 2017-09-29 and Y-axis represents the Bitcoin returns.
From the plot, we can see that, the returns are volatile. The returns tend to be more convergent around 2015 to 2017. On 2013-11-18, the return met the maximum value of 0.435. On 2013-12-06, the return met the minimum value of -0.238. Therefore, the volatility of the return was strong in the end of 2013. The returns are also getting volatile again 2017.

3. Unit Root Test

Step 1: Dickey Fuller testing (HT)

S1: *Hypothesis*:

H0 (Null hypothesis): $\phi = 1$. Log of Bitcoin price would be difference stationary.

H1 (Alternative Hypothesis): $|\phi| < 1$. Log of Bitcoin price would be trend stationary.

S2: After running Dickey Fuller test, we see that the p-value=0.9613.

S3: *Decision Rule*: Reject the null hypothesis when p-value <5%.

S4: *Conclusion*: Here we can see that, the p-value is greater than 5%, thus we fail to reject the null hypothesis.

Step 2: Augmented Dickey Fuller testing (HT)

S1: *Hypothesis*:

H0 (Null hypothesis): $\phi = 1$. Log of Bitcoin price would be difference stationary.

H1 (Alternative Hypothesis): $|\phi| < 1$. Log of Bitcoin price would be trend stationary.

S2: After running Augmented Dickey Fuller test, we see that the p- value= 0.8477

S3: *Decision Rule*: Reject the null hypothesis when p-value <5%.

S4: *Conclusion*:

Here we can see that, the p-value is greater than 5%, thus we fail to reject the null hypothesis.

4. Estimate an ARMA (p, q) Model (EST)

```
Call:
arima(x = ReturnX, order = c(p, 0, q))

Coefficients:
      intercept
         0.0044
s.e.     0.0015

sigma^2 estimated as 0.002595:  log likelihood = 1737.31,  aic = -3470.62
> coeftest(ARMA00)

z test of coefficients:

           Estimate Std. Error z value Pr(>|z|)
intercept 0.0043678  0.0015259  2.8625 0.004204 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From the R output, we can write the **estimated equation** is: *Est (*x_t*)* - 0.0044 = ε_t

P-value: In 95% confidence level, we hold that the parameter is significant when its p-value is less than 5%. In this estimated equation, the p-value of μ is lesser than 5%, therefore it is significant.

<u>**Select a model that minimizes AIC and BIC**</u>

Based on AIC and BIC values, we choose **ARMA (0,0)**.

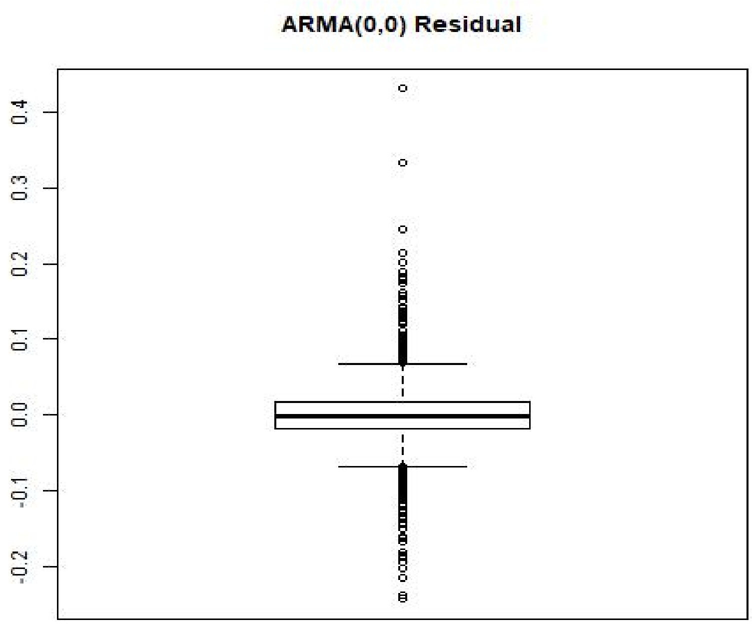
```
> ARMAt_AIC
          0         1         2
0 -3470.616 -3468.640 -3467.166
1 -3468.639 -3466.639 -3465.752
2 -3467.213 -3465.868 -3468.180
>
```

```
> ARMAt_BIC
          0         1         2
0 -3460.583 -3453.590 -3447.100
1 -3453.589 -3446.573 -3440.669
2 -3447.147 -3440.785 -3438.081
>
```

5. ARMA (0,0) Diagnosis Test (5FIG + 3HT)

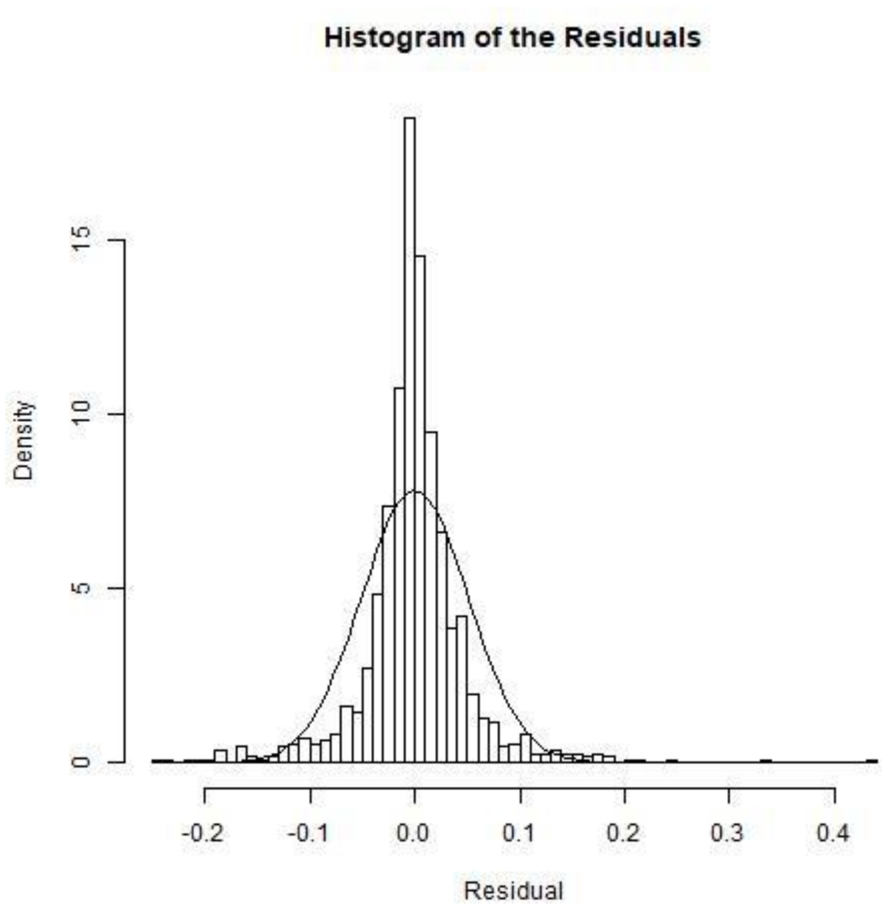
Do disturbances term εt follow a normal distribution?

5.1. Box-Plot



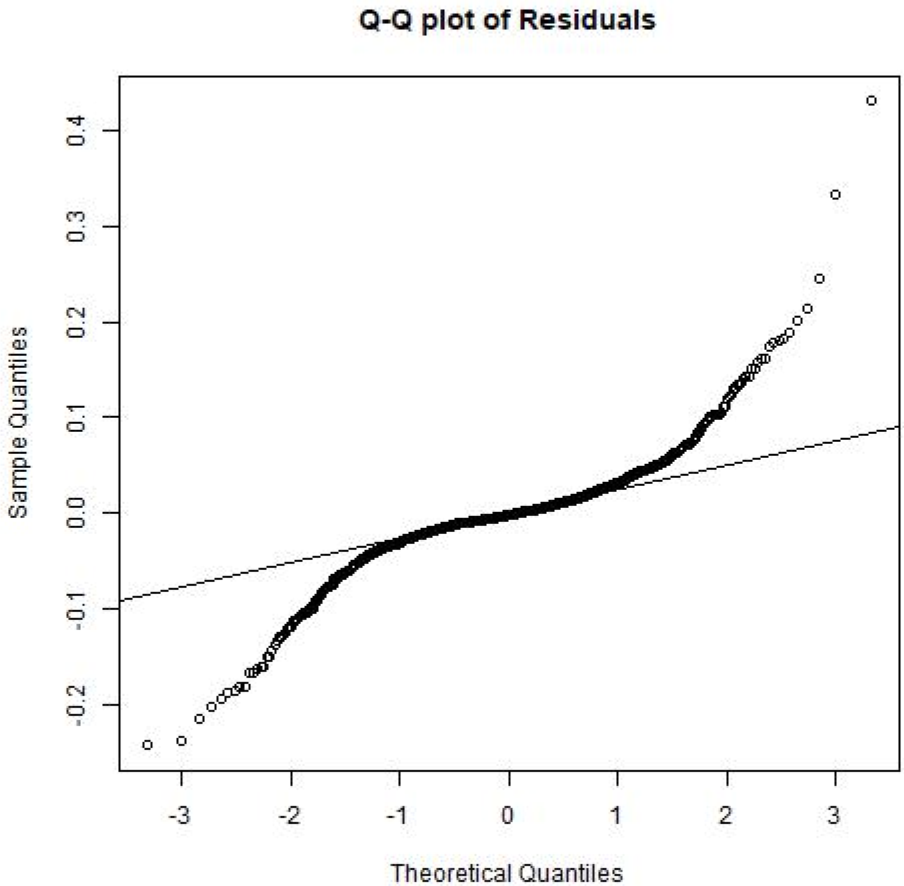
Interpretation: The box plots given above show the distribution for the residuals. Y-axis represents the Residuals. The data is sorted and divided into 4 equal size groups, i.e. 25% of the data is placed in each group. In this plot, the minimum (the lowest point), first quartile (the lower bound of the rectangle), median (the black line in the center), third quartile (the upper bound of the rectangle) and maximum (the highest point) are clearly displayed. In the plot, median of residuals is -0.001969087. 50% of the returns are between -0.017490 (the first quartile) and 0.017008(the third quartile). From this plot, we can see outliers lie under the lower short line and above the higher short line.

5.2. Histogram plot



Interpretation: In this plot, X-axis represents the Residuals and Y-axis represents the Density. Most dots lie around the range between -0.2 to 0.2. However, the bars in the middle part of the graph are higher than the normal distribution curve. On either side of the middle region, the bars are slightly lower than the normal distribution curve. At the two tails of normal distribution, the histogram is slightly higher than the normal distribution curve. Thus, we conclude that it is not a normal distribution.

5.3. QQ Plot



Interpretation: In this plot, the X-axis is the theoretical quantiles from -3 to 3, and Y-axis is the quantile of residuals. The dots in this plot do not construct a straight line. The lower region on the left and higher region on the right show large deviations from normal distribution, in other words, dots are far away from the normal distribution line. Also, there are outliers in this plot, which means they both have fatter tails compared to normal distribution. Thus, we infer that it is not a normal distribution.

6. Jarque-Bera test

The Jarque-Bera test is a test for normality.

S1: *Hypothesis*: H0 (Null hypothesis): Residuals follow normal distribution.

H1 (Alternative Hypothesis): Residuals do not follow normal distribution.

S2: From the computer output, we see that for residuals, the p-value < 2.2e-16.

S3: *Decision Rule*: Reject the null hypothesis when p-value <5%.

S4: *Conclusion*:

Here we can see that, the p-value<2.2e-16 of Residuals, is much lesser than 5% (alpha, the significant level), thus we reject the null hypothesis. This means that H1 is valid and residuals are not normal distributed.

7. Lilliefors test

The Lilliefors test is a test for normality which is usually done when we do not know the mean or standard deviation.

S1: *Hypothesis*:

H0 (Null hypothesis): Residuals are a normal distribution.

H1 (Alternative Hypothesis): Residuals are not a normal distribution.

S2: From the computer output, we see that for the residuals, the p- value< 2.2e-16.

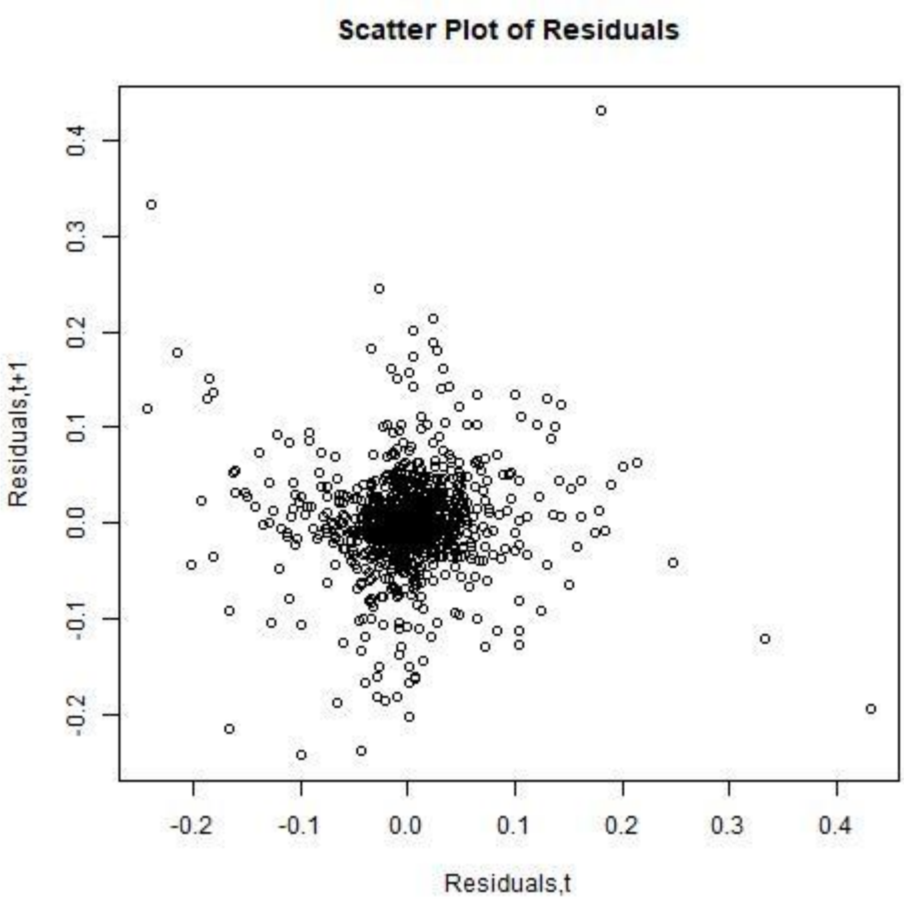
S3: *Decision Rule*: Reject the null hypothesis when p-value <5%.

S4: *Conclusion*:

Here we can see that, the p-value< 2.2e-16, which means that it is lesser than 5% (alpha, the significant level). Thus, we reject the null hypothesis, and conclude that residuals are not a normal distribution.

8. Are disturbances term εt serially uncorrelated?

8.1. Scatter Plot:



Interpretation: The scatter plot drawn above helps to identify if the residuals are independent of each other. The X-axis is the Residuals in time t and Y-axis is the Residuals in time t+1. It seems like most of the points are at the middle of the distribution. We cannot see a parabola or a linear line in the plot. So, we conclude that the residuals do not have a relationship. Hence, uncorrelated.

8.2. Ljung-Box Q-Test

S1: *Hypothesis*:

H0 (Null hypothesis): Residuals are serially uncorrelated.

H1 (Alternative Hypothesis): Residuals are serially correlated.

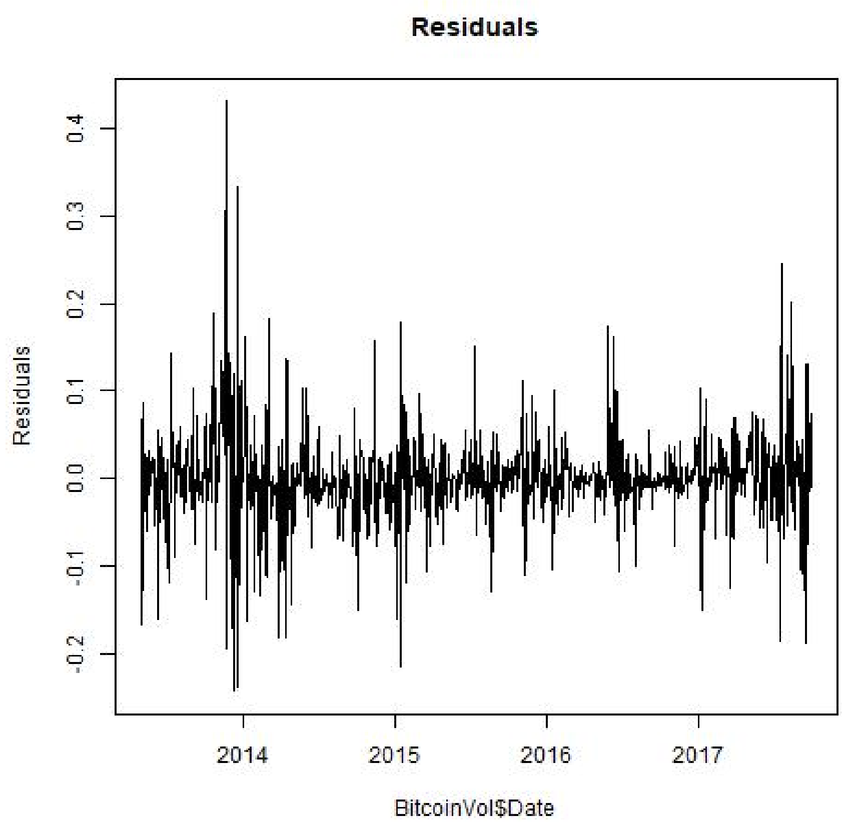
S2: From the computer output, we see that for the residuals, the p- value= 0.0229.

S3: *Decision Rule*: Reject the null hypothesis when p-value <5%.

S4: *Conclusion*:

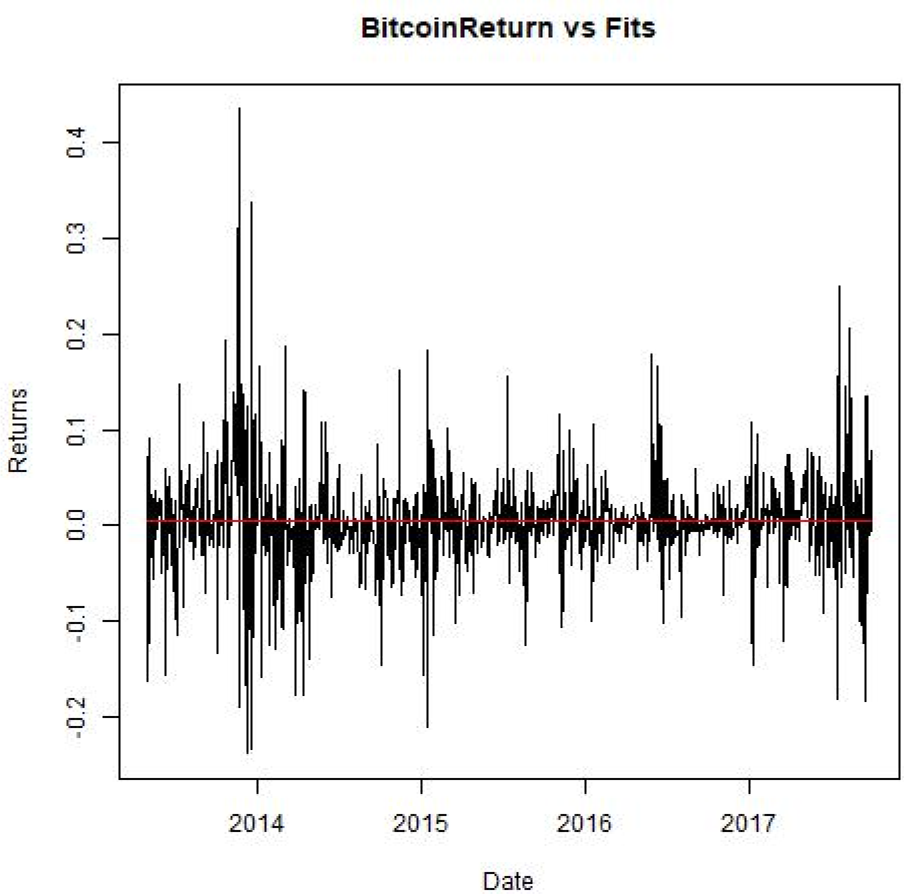
Here we can see that, the p-value is 0.0229 < 5% (alpha, the significant level), thus we reject the null hypothesis, and conclude that the residuals are serially correlated.

9. Is there any evidence of Heteroskedasticity?



Interpretation: The plot shows the residuals over the years. In this plot, X-axis represents the Date range from 2013-04-29 to 2017-09-29 and Y-axis represents the Residuals. The residual values lie around -0.2 to 0.4 from 2014 to 2017, then the volatility of residuals became smaller around 2014 to 2017. On 2013-11-18, the value of residuals met the maximum level of 0.4309. On 2013-12-06, the value of residuals met the minimum point at -0.2428413. Therefore, the volatility of the residual is great around 2014. Overall this plot has a cone shape, which means that residuals have heteroscedasticity.

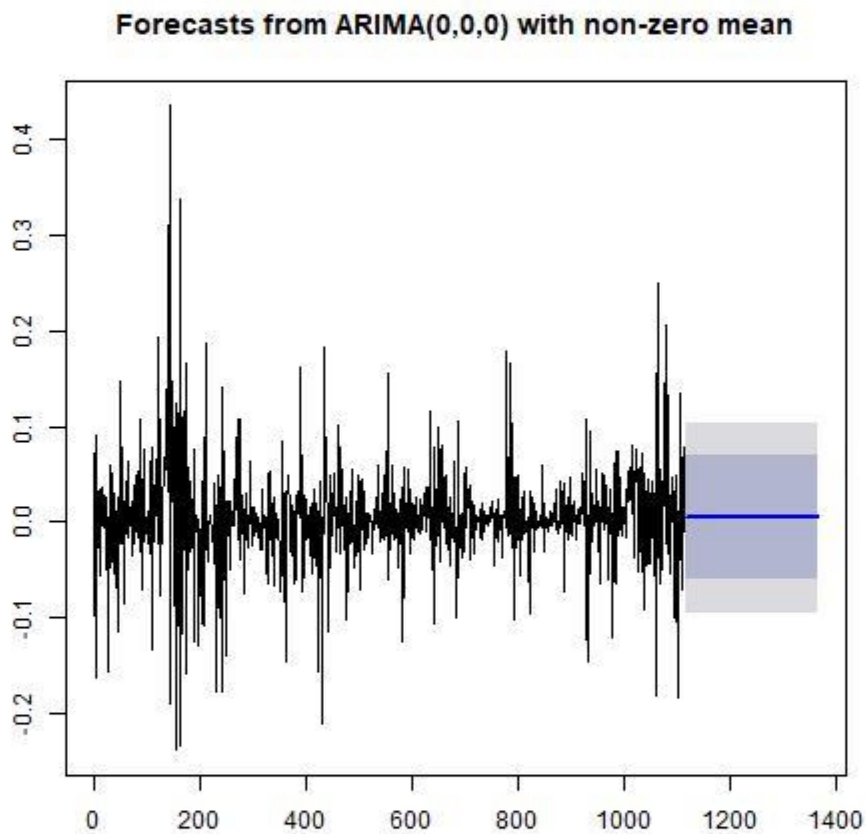
10. ARMA (0,0) Predicted Value (FIG + TAB)



Interpretation: This plot displays the actual value and the fitted value of Bitcoin Return. In this plot, the X-axis represents the Date range 2013-04-29 to 2017-09-29 and the Y-axis represents the Bitcoin Return. The black line is the time-series plot of actual return of Bitcoin and the red line is the time-series plot for the fitted values using the ARMA model. The fitted values fluctuate around 0. But the predicted part does not fit well with the actual part. Therefore, the model does not explain the return very well.

11. Training and test sets (3FIG + 1TAB)

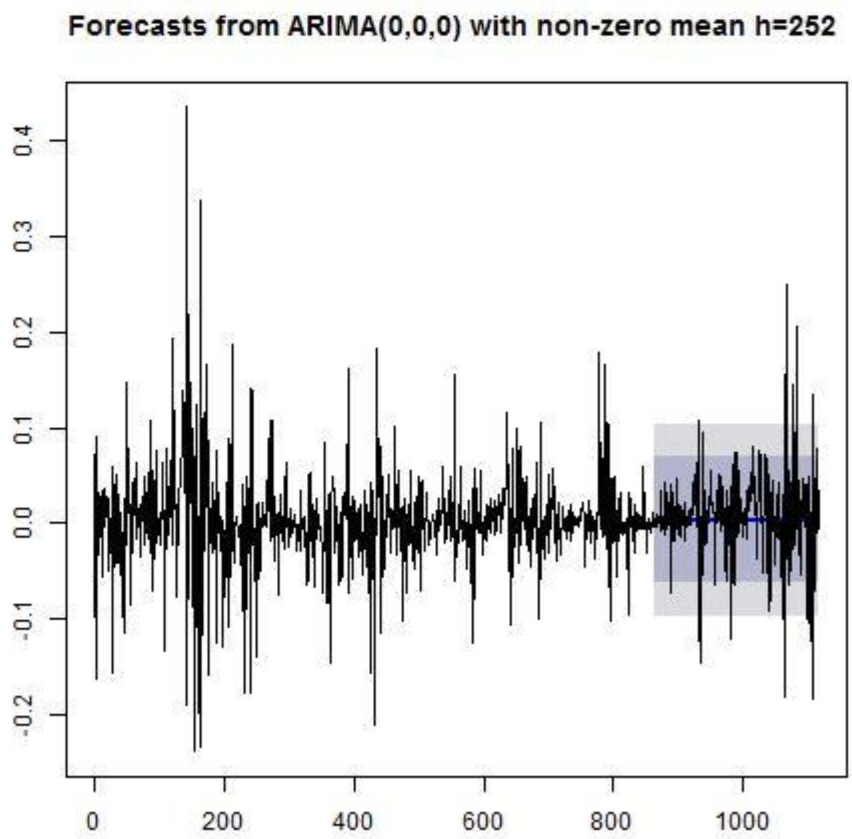
Case 1



Interpretation: The plot drawn above, shows the time series of Bitcoin return and the forecasted returns for the next 252 days. In this plot, X-axis represents the time (ie, t=1, t=2) and Y-axis represents the returns.

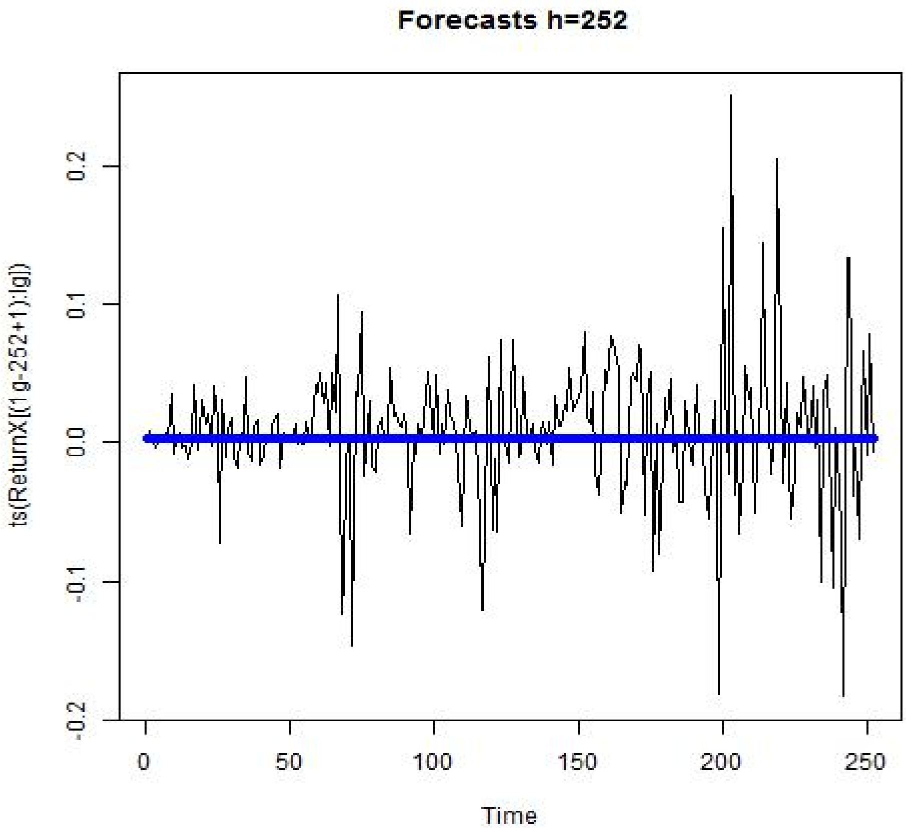
In the first case, we use 1115 actual Bitcoin returns to build ARMA (0,0) model and use this model to forecast the next 252 days' Bitcoin returns after t=1116. Here, we have a training set of size 1115 and test set of size 0. The time-series plot shows the observable returns from t=0 to t=1115 and the fuzzy part of the graph behind t=1115 shows the estimated range of Bitcoin returns. The dark fuzzy part represents the 80% confidence level of the point estimation and the light fuzzy part represents the 95% confidence level of the point estimation.

Case 2



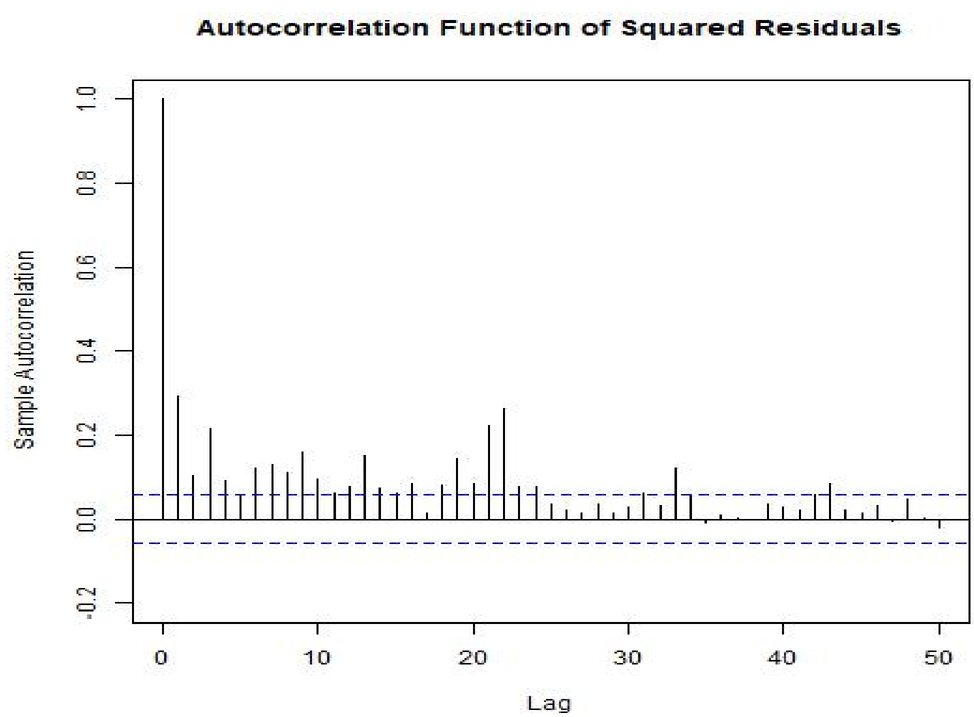
Interpretation: The plot drawn above, shows the time series of Bitcoin return and the forecasted returns of next 252 days. In this plot, X- axis represents the time (i.e., t=1, t=2) and Y-axis represents the returns.

In the second case, we use part of the actual return data (the first 863 data) to build ARMA (0,0) model and forecast the last 252 days' Bitcoin returns in our sample. Here, we have a training set of size 4773 (1115-252=863) and test set of size 252. The fuzzy part of the graph behind t=863 shows the estimated range of Bitcoin returns and the black line is the actual observed returns.



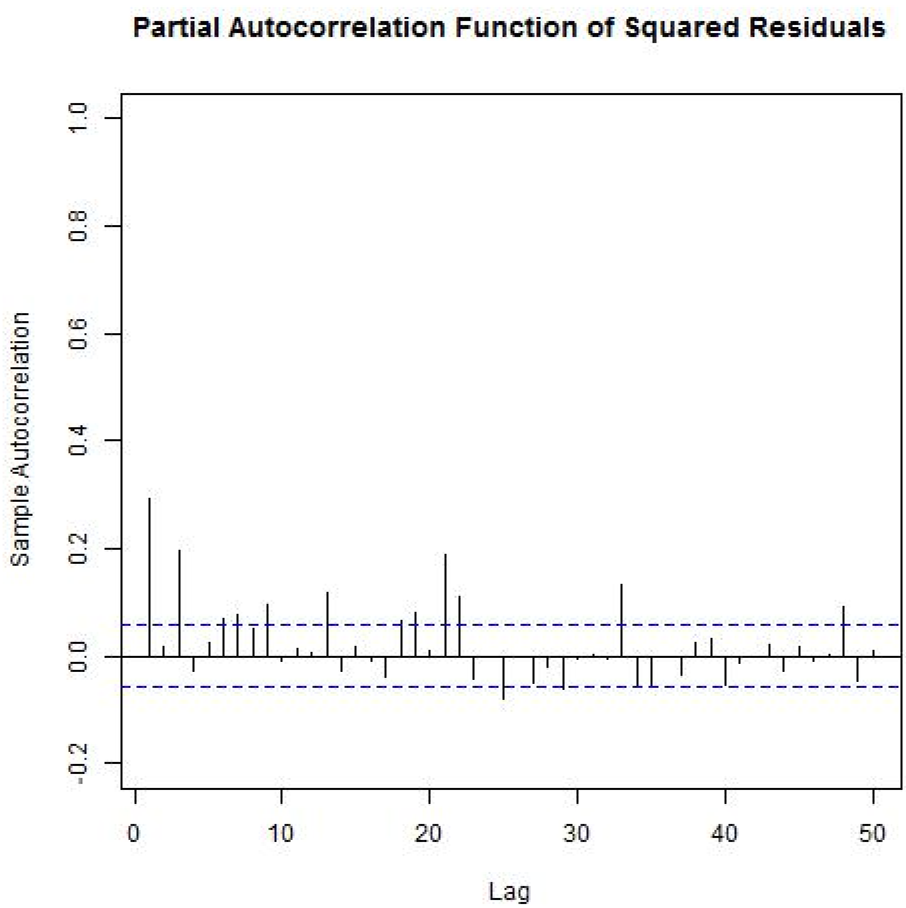
12. ARCH model

12.1. Auto-correlation plot for the squared residuals et^2 (FIG)?



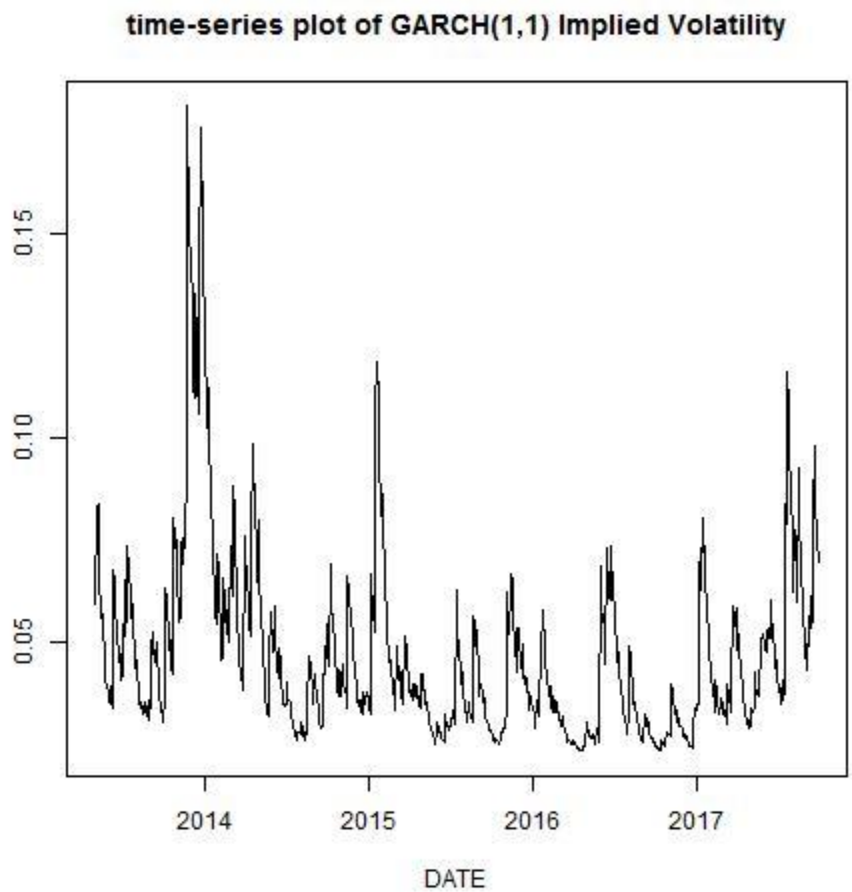
Interpretation: The plot drawn above, shows the autocorrelation of squared residuals of ARMA (0,0) model. The x-axis is lag and y-axis is sample autocorrelation. The autocorrelation function measures the correlation between yt and yt+k (k is the lag). The area within the two dash lines is the confidence band. If the bar is near 0, in other words, the bar lies within the confidence band, εt2 are not auto-correlations. We observe that the bars are all outside the confidence band. Hence, we strongly reject that the disturbance terms εt are independent.

12.2. Partial auto-correlation plot for the squared residuals et^2 (FIG)



Interpretation: The plot drawn above, shows the autocorrelation of squared residuals. The X-axis is lag, and Y-axis is sample autocorrelation. The area within two dash line is confidence band. If the bar is near 0, in other words, the bar lies within the confidence band, εt2 are not auto-correlations. Therefore, we fail to reject that disturbance terms εt are independent in these lags. On the other hand, most of the bar are outside the confidence band, there we can strongly reject that the squared disturbance term ε2t are serially uncorrelated.

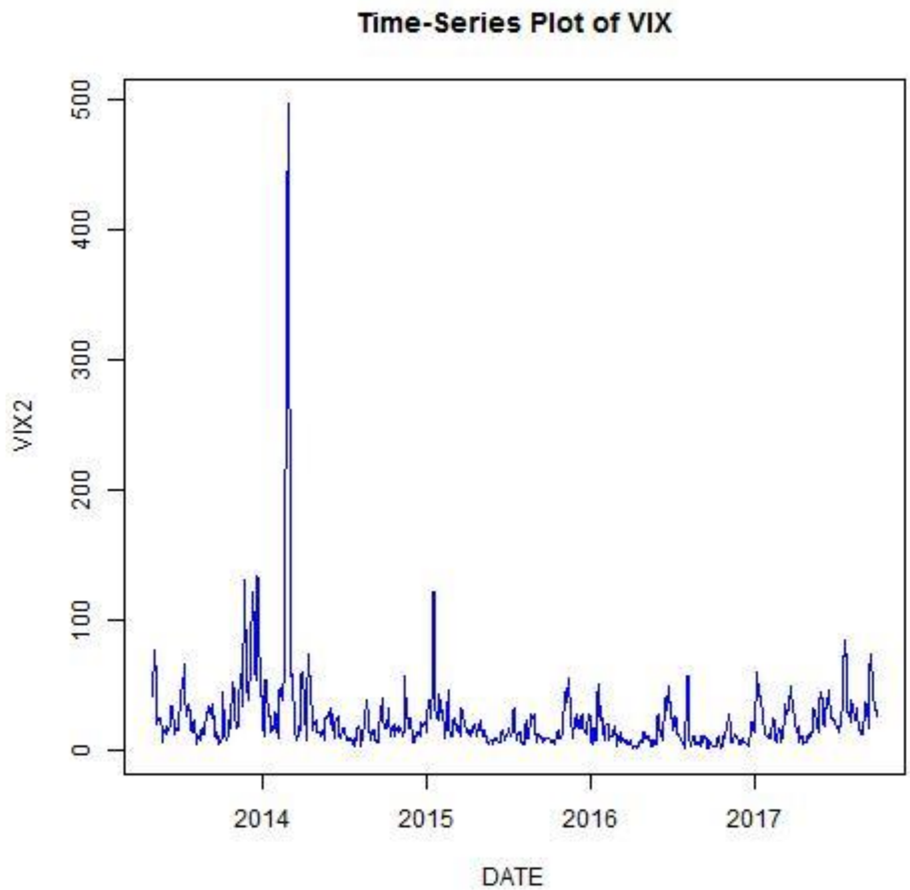
13. GARCH family
GARCH (p, q) (EST+FIG)



Interpretation: The graph is time series plots of GARCH (1,1) implied volatility. X-axis represents the Date range from 2013-04-29 to 2017-09-29 and Y-axis represents conditional predicted standard deviation of GARCH (1,1).

The implied data meet minimum value of 0.0232 on 2016-04-19 and it meet maximum value of 0.1810 on 2013-11-20. From the plot we can see that compared with less volatility after 2014 the volatility before 2015 is strong. This is because FBI took down Silk Road, the biggest online marketplace to buy anything illegal, they seized over 26,000 Bitcoins.

14. Time-Series plot for VIX



Interpretation: The graph is time series plot for CBOE Volatility Index (ticker symbol VIX) traded on the Chicago Board Options Exchange (CBOE). The x-axis represents the Date ranging from 2013-04-09 to 2017-09-29 and y-axis represents VIX. The data meets minimum value 1.47 on 2016-09-16. In addition, the data meets maximum value 495.99 on 2014-02-25 showing profound change in volatility due to the downfall of Silk Road and ceasing of 26000 bitcoins. The reason for the spike after 2017 are the issues with regulations around the world. For example, China has recently blocked the registration of new users, whereas Japan has allowed the usage of Bitcoins. The increase in adoption of Bitcoins in the mainstream financial economy has affected the volatility as well.

5. Conclusion:

We have analyzed the price trends on time-series plot for Ethereum, Bitcoin and Litecoin. And all of them showed an upward trends in its price. We started our further analysis on Bitcoin as it showed maximum variations in its price. Moving forward with finding correlation between Market and Bitcoin Returns, we observed that there was no correlation between the two. So, we would recommend investors to include Bitcoin in their portfolio to reduce the risks. We also performed normality tests on Bitcoin Price data and found that the data did not follow normal distribution. After running the volatility models, we found out that the prices were volatile in 2014 and are getting volatile again in 2017. In the past 24 months, the USD-BTC exchange rate increased more than 50-fold. At last, we forecasted Bitcoin Price for 252 days and observed 80% accuracy in the predicted values.