# The Future of Money: A Statistical Analysis on Cryptocurrencies

Ampadu-Kissi, Ama

akampdu@ilstu.edu

Owusu, Isaac

iaowusu@ilstu.edu

March, 2018

# Contents

**Abstract** **ii**

**1 Background of the study** **1**

1.1 Introduction . . . . . 1

1.2 Literature Review . . . . . 2

1.2.1 Cryptocurrencies and their development . . . . . 2

1.2.2 Analysis on the Price of Cryptocurrencies . . . . . 5

**2 Data and Methods** **7**

2.1 Data . . . . . 7

2.2 Methods . . . . . 8

2.3 Assumptions and Limitations . . . . . 9

2.4 Hypothesis and Areas of Interest . . . . . 10

**3 Time Series Modeling** **12**

3.1 Training Data . . . . . 12

3.1.1 Test for Stationarity . . . . . 13

3.1.2 Data Transformation . . . . . 14

3.2 Tentative Model Building . . . . . . . . . . . . . . . . . . . . . 18
3.3 Model Fitting and Model Estimates . . . . . . . . . . . . . . . . 20
3.3.1 Residual Analysis . . . . . . . . . . . . . . . . . . . . . . 21
3.3.2 Outliers . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
3.4 Model Validation . . . . . . . . . . . . . . . . . . . . . . . . 25
3.4.1 Test for Overfitting . . . . . . . . . . . . . . . . . . . . . 25
3.4.2 Prediction . . . . . . . . . . . . . . . . . . . . . . . . . . 26
3.4.3 How many days ahead can we predict efficiently . . . . 28

**4 Relationship of Bitcoin with Altcoins and Other Investment Instruments 30**
4.1 Relationship Between Bitcoin and Litecoin . . . . . . . . . . . 32
4.2 Relationship Between Bitcoin and Ripple and Ethereum . . . 32
4.3 Relationship between Bitcoin and Wider Investment Market . 34

**5 Conclusion 37**

**Bibliography 40**

**A 41**
A.1 Residual Plots . . . . . . . . . . . . . . . . . . . . . . . . . . 41
A.2 Confidence Intervals for Predictions . . . . . . . . . . . . . . . 42
A.3 Table with significance of extra terms . . . . . . . . . . . . . . 43
A.4 Model Estimates . . . . . . . . . . . . . . . . . . . . . . . . . 44

## Declaration

This essay was written in the Department of Mathematics, Illinois State University, from October 2017 to March 2018, in partial fulfilment of the requirement for the completion of Statistical Competition.

Abstract

Subsequent to the 2008 financial crisis, the world has seen a massive rise in the use of cryptocurrencies. While opinions are still divided over the role of cryptocurrencies in the storied evolution of money, many analysts have concluded that cryptocurrencies will play a major part in the financial sector in the coming years. With their influence becoming more and more prominent in recent times, we attempt to analyze statistically, trends and relationships that exist among selected cryptocurrencies and between cryptocurrencies and other financial assets.

# Chapter 1

# Background of the study

## 1.1 Introduction

Since 2009 when Bitcoin, one of the first and most prominent cryptocurrencies launched, there have been a rapid increase in the number of this nouvelle financial tool. Various sources peg the number of these cryptocurrencies in circulation to be upward of nine hundred with the major players in the industry being Bitcoin, Ethereum, Bitcoin Cash, Ripple and Litecoin. In 2017, Bitcoin saw a more than 2000% increase in its exchange rate per US Dollar. Bitcoin rose from below $1,000 at the beginning of 2017 to about $20,000 by the end of the year. Other cryptocurrencies also recorded significant growth figures in their values over the same period of time, indicating a growing confidence in cryptocurrencies by the investment community. In 2018, however, the cryptocurrency space has experienced some setbacks which have adversely affected their prices. It remains to be what direction they will take going forward.

Cryptocurrencies use distributed ledger technologies to decentralize financial transactions making it globally accessible, faster and relatively cheaper due to the elimination of middlemen. The current climate makes for interesting research and this paper seeks to explore some of the emerging questions that are arising from the prominence of cryptocurrencies.

## 1.2 Literature Review

### 1.2.1 Cryptocurrencies and their development

"Cryptography has been used almost since writing was invented" Bellare and Rogaway (2005). The aim of cryptography has always been to ensure the security of information that is being passed on from one party to the other over an insecure channel. The technique is employed to ensure that no party other than the sender and intended receiver of the information can decode it. In modern times, cryptography has found its way into many applications and continues to gain prominence. The most recent and widely popular application of cryptography has been in the financial sector. However, in the few decades leading to this widespread application in the financial sector, cryptography was used, as is still being used, immensely in military communications. As the years have advanced, cryptography has become more complex inculcating knowledge from the academic disciplines of mathematics, computer science, electrical engineering, communication science, and physics[1].

During the dotcom bubble, many companies were formed to solve different problems. The growing use of the internet facilitated the operations of these companies. One of the areas that generated interest was virtual currency. In 1998 and 1999, Beenz and Flooz respectively, were founded to provide internet users the ability to transact business using virtual currencies. The models and technology used in those projects are different from the current iteration of virtual currencies as we have them today, but these represent the early stage of the virtual currency movement. It is also important to note that these early virtual currencies are not considered cryptocurrencies but rather electronic currencies.

According to Lansky (2018), a cryptocurrency is a system that meets all the following 6 conditions:

- The system does not require a central authority, distributed achieve consensus on its state.
- The system keeps an overview of cryptocurrency units and their ownership.
- The system defines whether new cryptocurrency units can be created. If new cryptocurrency units can be created, the system defines the circumstances of their origin and how to determine the ownership of these new units.
- Ownership of cryptocurrency units can be proved exclusively crypto-

[1] https://en.wikipedia.org/wiki/Cryptography

graphically.

- The system allows transactions to be performed in which ownership of the cryptographic units is changed. A transaction statement can only be issued by an entity proving the current ownership of these units.
- If two different instructions for changing the ownership of the same cryptographic units are simultaneously entered, the system performs at most one of them.

In his 2008 paper, "A Peer to Peer Electronic Cash System", Nakamoto, outlined the solution to inculcating pseudo anonymity, independence from central authority and prevention of double-spending into a currency. This birthed Bitcoin, the first cryptocurrency in 2009. Bitcoin rode on the back of ground-breaking research performed by Chaum (1983), Back (2002), Haber and Stornetta (1997) et cetera in the cryptology space.

Bitcoin, and many of the cryptocurrencies available today, use technology known as Blockchain technology to facilitate transactions. A blockchain is as distributed ledger technology (DLT). Blockchain is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way". Iansiti and Lakhani (2017). It uses cryptography to achieve the six criteria for a cryptocurrency. Blockchain technology is employed in other fields aside from the cryptocurrencies. Many industries are using the technology to improve their processes by leveraging on the speed, accuracy and transparency that the blockchain technology provides. The technology has been used in Supply Chain Management, Quality Assurance, Accounting, Smart Contracts, Voting, Stock exchange, Peer-to-Peer global transactions to name a few (Agrawal, 2018).

Bitcoin launched at an exchange rate comparable to zero. Most users of Bitcoin at the time were Bitcoin enthusiasts who exchanged the coins as a hobby. It was not till Spring 2011 that the cryptocurrency gained parity with the USD selling at $1 per Bitcoin by April 2011. Since then, the cryptocurrency has risen in leaps and bounds. It was in 2017 that Bitcoin gained the most international media attention. The cryptocurrency started the year at about $1000 and almost hit the $20,000 mark by December 2017. At the time of this study, the Bitcoin sells at $7500. Together with other cryptocurrencies, Bitcoin has faced many setbacks during its lifetime and many enthusiasts attribute the recent fall in the cryptocurrency to fear, uncertainty and doubt (FUD) that surrounds all new innovations. Cryptocurrency enthusiasts view cryptocurrencies as the future of money.

### 1.2.2 Analysis on the Price of Cryptocurrencies

Most previous analytical work on cryptocurrencies have been centered on Bitcoin, the market leader. In their paper, "A Statistical Analysis of Cryptocurrencies", Chan et al (2017) attempt to find a parametric distribution that fits the major cryptocurrencies. They conclude that the general hyperbolic distribution provided the best fit for the data on Bitcoin and Litecoin. Their work is a corroboration of the work done by Chu et al. (2015) which also led to a similar conclusion on the parametric fit of Bitcoin data. Hencic and Gourieroux (2014) use a "noncausal autoregressive process with Cauchy errors in application to the exchange rates of the Bitcoin electronic currency against the US Dollar". They conclude that there may exist local trends in the daily exchange rate and these could be an indication of some speculative behavior stemming from online trading activities. Briere et al(2015) examine the inclusion of Bitcoin in a well diversified portfolio of investments and find out that Bitcoin had an exceptionally low correlation with other (non-crypto) assets. Following from this observation, the inclusion of Bitcoin in a well diversified portfolio holds the potential to greatly improve the risk-return tradeoff.

Bovaird (2017) examined the correlation between Bitcoin and other cryptocurrencies (also referred to as Altcoins). The conclusion drawn was a strong correlation between Bitcoin and Litecoin as well as a weak correlation between Bitcoin and Ripple prices over short intervals of time.

# Chapter 2

# Data and Methods

## 2.1 Data

The primary source of data used in performing the analysis in this project was retrieved from Coin Market Cap . This source contains relevant data on all cryptocurrencies on the market. The data collected includes the following information for each cryptocurrency:

- Date
- Opening price
- High and low prices
- **Closing price**
- Volume
- Market cap

Daily returns on prices are calculated from closing prices.
The closing price was the column of interest in the analysis. This is the price of the cryptocurrency at the end of the business day as reported from market activities. The closing price and the date are therefore the only data used during the time series modeling. The remaining data columns are utilized in the rest of the analysis. For this paper, all currency values are expressed as the exchange rate of the particular cryptocurrencies with the US Dollar.

We base our analysis on the five most notable cryptocurrencies at the time of our analysis. Our dataset includes data on Bitcoin, Bitcoin Cash, Ethereum, Litecoin and Ripple. Our assessment of their prominence is based on their market capitalization. These 5 cryptocurrencies contribute about 75% of the entire cryptocurrency market capitalization. Bitcoin, the market leader, has about 45% of the market capitalization. Ethereum, Ripple, Bitcoin Cash and Litecoin follow in that order.

For the time series analysis on Bitcoin, training data ranges from April 28, 2013 to January 31, 2018. This contained 1739 data points. The test data is made up of all Bitcoin prices from February 1, 2018 to March 16, 2018.

The compiled data for the entire analysis in this study can be accessed at `https://drive.google.com/open?id=1fXkv9Qh4GBN-QrDW_J8m8nkhHlUHggMA`.

## 2.2 Methods

The main statistical methods used in our analysis are Time series models.

## 2.3 Assumptions and Limitations

Before using the data, we review the data and conclude from our review that the data is accurate and free from errors. The data is deemed complete for our purpose.

Given that the primary model implemented is a time series model, modeling begins with a test of the assumptions on which a time series model is built. For a time series model, the following assumptions must hold:

1. The first assumption is that the series is stationary. This essentially means that the series is normally distributed with a constant mean and variance over a long period of time.

2. The next assumption is that of an uncorrelated random error, that is, errors are randomly distributed with a constant mean and variance over a long period of time.

3. We also assume that shocks, if present, are randomly distributed over time with a mean of 0 and a constant variance.

Contrary to the above assumptions, the following setbacks are encountered during the process of modeling:

The greatest limitation in our study is the breach of the assumption that the residuals are normally distributed. Although this is a strong assumption that should preferably not be defied, we are unable to conclude that the residuals are normally distributed and this is a major drawback to our analysis.

## 2.4 Hypothesis and Areas of Interest

In this paper, we study a number of areas of interest with relation to the major cryptocurrency, Bitcoin, and a few others. We also examine the relationship between Bitcoin and the wider investment market using the S&P 500 and the Dow Jones Industrial Average as indices for market performance.

The following hypotheses are tested:

- There exists some relationship between the movement of Bitcoin and the other major cryptocurrencies.
- Returns on Bitcoin is negatively correlated with at least one of the major stock indices (S&P 500 and the Dow).

In addition to testing the above hypotheses, the paper focused on building a simple time series model for Bitcoin. The behavior of the cryptocurrency is observed and various analysis made. Below is a detailed list of the areas explored in this paper regarding the time series model:

- A time series analysis is performed on the exchange rate per US Dollar of Bitcoin. The cryptocurrency is examined for autocorrelation, trends or seasonality that may exist in the data of the exchange rates. An attempt is made to understand the forces that underlie the behaviour of the exchange rate of the cryptocurrencies and to fit simple time series models to it. Further to this, the study forecasts the exchange rate per US Dollar of Bitcoin based on the model that fit the data. A determination is made as to which time frame realistic forecasts can be made with a degree of confidence about their accuracy.

- According to coinmarketcap.com which keeps data on the market capitalization of all cryptocurrencies active today, Bitcoin is the leader in the market boasting close to 45% of the entire market. While Bitcoin continues to remain the leader in the space, it is interesting to note that its market capitalization has fallen drastically from about 95% in April 2013 to its current value due to the proliferation of these cryptocurrencies. Ethereum, which was non-existent in April 2013 now boasts about 15% of the market. In 2017, Bitcoin saw a meteoric rise in its exchange rate. Altcoins also experienced an upsurge and the dynamics involved in their interactions with Bitcoin are looked at in this study.

- Lastly, we attempt to study the movement of Bitcoin in relation to the wider investment market. Our concentration remained on finding interesting relationships that may exist between Bitcoin and the S&P 500 and the Dow. We choose the Dow and S&P 500 because they present a high level assessment of the market. Historical data from both indices are analyzed together with data on Bitcoin and useful conclusions drawn.

# Chapter 3

# Time Series Modeling

The prices of Bitcoin are recorded on a daily basis. This provides us with data points that are indexed in a timely order and thus, we begin with a time series model to fit the data appropriately.

## 3.1 Training Data

The first step taken prior to doing any modeling is an analysis of the data. A plot of the data below shows the distribution of the closing prices of Bitcoin over the period studied.

The plot shows an increasing pattern in the prices. This is expected because the data used is made up of prices from the emergence of cryptocurrency, when many people did not have as much trust in cryptocurrency. The prices rose as the concept was embraced by the world better, with major fluctuations

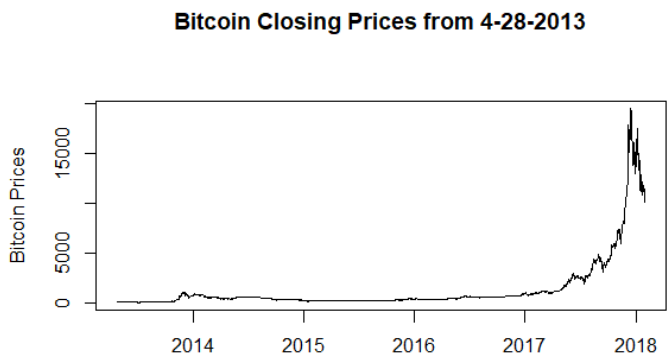



Figure 3.1: Time Series Plot of Bitcoin Prices

occurring after Bitcoin popularity in the market. The steady increase in prices is an indication that the data is not stationary. A formal test is carried out to confirm the stationarity or otherwise of the data.

### 3.1.1 Test for Stationarity

A unit root testing of stationarity is performed using the Dickey Fuller Test. The null and alternative hypothesis of this test is as follows:

$H_0$ : The time series is non-stationary ($\alpha = 1$)

$H_a$ : The time series is stationary ($|\alpha| < 1$)

A non-stationary data will suggest some data transformation.
The results of the Dickey Fuller test for stationarity resulted is a p-value of 0.99. This confirms the speculation of non-stationarity from the patterns observed in the time series plot. Stationarity is an important assumption.

In order to proceed with further analysis, the data is transformed to satisfy the stationarity assumption.

### 3.1.2 Data Transformation

Transformation of the time series data involves applying suitable functions to the data, with the goal of eliminating existing patterns and making the data stationary. In carrying out this step, the following transformations are randomly considered.

1. Log transformation
2. Difference transformation
3. Difference of log transformation

#### Log Transformation

The first transformation tested is a log transformation. The resulting impact of the plot is shown in the plot below.

Although the transformation eliminated some of the pattern observed in the original time series plot, it is evident that there still persists some non-stationarity. There is clearly still an increasing pattern in the data. The formal Dickey Fuller test results in an p-value of 0.9286. This confirms that the transformation does not remove the patterns in the data and does not make the time series stationary.

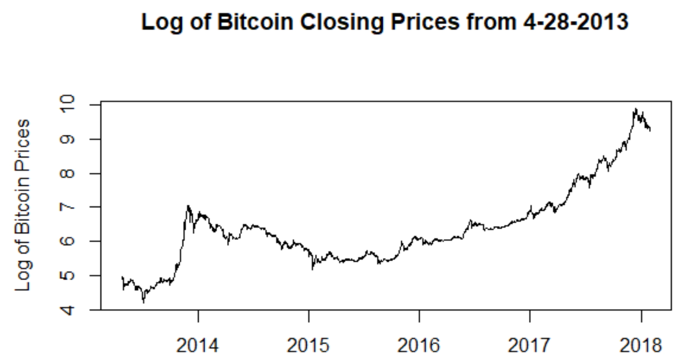



Figure 3.2: Time Series Plot of Bitcoin Prices after Log Transformation

### Difference Transformation

The next transformation is a difference transformation. The resulting time series plot is shown below:

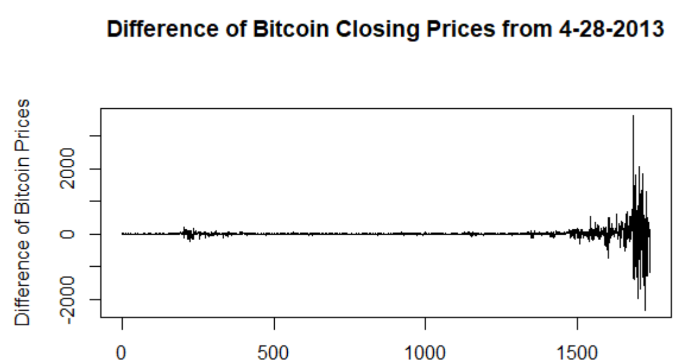



Figure 3.3: Time Series Plot of Bitcoin Prices after Difference Transformation

The plot above depicts a random distribution of the data points. This can be an indication that the transformed data is stationary. The Dickey Fuller test is performed to confirm this assumption. The test results in a p-value of 0.01. At a significance of 0.05, we reject the null hypothesis and in favor of the alternative that the transformed data is stationary.

### Difference of Log Transformation

A difference of log is the final transformation considered. The resulting time series plot of the transformed data is shown below.

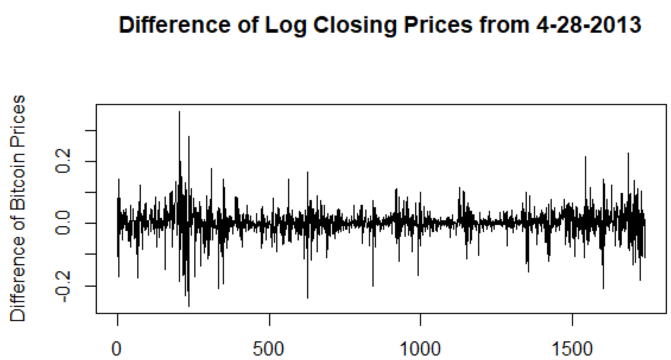



Figure 3.4: Time Series Plot of Bitcoin Prices after Difference of Log Transformation

From the plot above, there is no pattern and we can therefore assume that the transformed data is stationary. A formal Dickey Fuller test confirms that the transformation is stationary. The resulting p-value from the test is 0.01. At a significance level of 0.05, the transformed data is stationary.

### Selecting Appropriate Transformation

From section 3.1.2 above, the following transformations both appear to fulfill the the goal of stationarity of transformed data. In order to make an appropriate decision on which transformation best suits our purpose, another formal test of stationarity is introduced: the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for Stationarity. The hypotheses of the KPSS test are as follows:

$H_0$ : The time series is stationary

$H_a$ : The time series is non-stationary

The test is used as a 'tie-breaker' between the difference and difference of log transformations. The results of the test are shown below:

| Transformation | P-value |
|---|---|
| Difference Transformation | 0.1 |
| Difference of Log Transformation | 0.1 |

From the table above, the KPSS test confirmed stationarity in both models. It did not discriminate against any of the two models so we made a decision on which of them to work with. The final transformation selected and implemented in the rest of the analysis and this paper is the difference of log transformation.

## 3.2 Tentative Model Building

After ensuring the stationarity of the data to be used, the next stages involve the modeling process. Because of the nature of the data under consideration, a time series model is deemed best. Bitcoin prices are made up of data points that are in a timely order. The prices are on a daily basis and as such, all analysis performed are carried out with the notion of a daily frequency (Frequency = 1).

The processes implemented in selecting an appropriate model for the Bitcoin prices involved using different time series functions such as:

- ACF - Autocorrelation function
- PACF - Partial Autocorrelation function
- EACF - Extended ACF
- Subset

The outputs of the ACF, PACF, EACF and subset plots are below:

- The ACF plot in figure 3.5 does not indicate a clear cut-off at any lag. The plot gives no indication of the possible time series model that is appropriate for the data.
- The PACF plot in figure 3.5 is similar to the ACF plot. There is no clear indication of a cut at a particular lag. The two plots therefore do not contribute much to deciding which model best fits the data.
- The EACF plot gives some indication of possible models to fit the data.

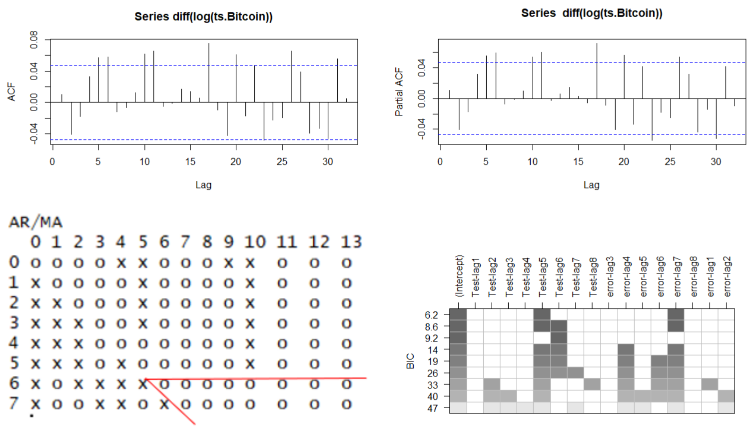



Figure 3.5: Plots of ACF, PACF, EACF and Subsets

From the plot, the following models are deduced

- an Autoregressive Moving Average model (ARMA (5,5)) may be an appropriate fit for the data. Given that in transforming the data we take the difference, the suggested model is therefore an ARIMA (5,1,5).

- an Autoregressive Moving Average model (ARMA (5,5)) may be an appropriate fit for the data. Given that in transforming the data we take the difference, the suggested model is therefore an ARIMA (5,1,5).

- The subset function finds a number of subset ARMA models. The plausible models are ordered by their resulting BICs. From figure 3.5 above, the model with the least BIC of 6.2 is an Autoregressive model with a lag of 5 and a Moving average of lag 7 (ARMA(5,7)). For the reason that the data is transformed by taking a difference, this model becomes an ARIMA(5,1,7).

From the outcomes of the plots and our analysis, we select the

- Model 1 - ARIMA(5,1,5)
- Model 2 - ARI(6,1)
- Model 3 - ARIMA(6,1,6); and
- Model 4 - ARI(5,1,7)

as 4 models that will potentially fit the data appropriately.

## 3.3 Model Fitting and Model Estimates

The selected models are fitted and their coefficients estimated using maximum likelihood. It is worthy to note that these models and their estimates are derived from the transformed data. It is important to take cognizance of this because the transformations will be reversed at the end of the process in order to interpret the outputs of the model appropriately. The resulting models with the estimates are shown below.

ARIMA(5,1,5):$Y_t = -0.0503Y_{t-1} + 0.3657Y_{t-2} - 0.3884Y_{t-3} + 0.0231Y_{t-4} + 0.9269Y_{t-5} + e_t - 0.0909e_{t-1} - (-0.3883)e_{t-2} - 0.03738e_{t-3} - 0.0301e_{t-4} - (-0.9042)e_{t-5}$

ARIMA(6,1,6):$Y_t = -0.2703Y_{t-1} - 0.2256Y_{t-2} + 0.1021Y_{t-3} - 0.0495Y_{t-4} + 0.5059Y_{t-5} + 0.7524Y_{t-6} + e_t - 0.2815e_{t-1} - 0.2220e_{t-2} - (-0.1000)e_{t-3} -$

$0.0670e_{t-4} - (-0.4578)e_{t-5} - (-0.7015)e_{t-6}$

ARI(6,1):$Y_t = 0.0081Y_{t-1} - 0.0375Y_{t-2} - 0.0116Y_{t-3} + 0.0374Y_{t-4} + 0.0597Y_{t-5} + 0.0643Y_{t-6} + e_t$

ARIMA(5,1,7):$Y_t = -0.1243Yt - 1 + 0.4342Yt - 2 - 0.5193Y_{t-3} + 0.1107Y_{t-4} + 0.8582Y_{t-5} + e_t - 0.1294e_{t-1} - (-0.4705)e_{t-2} - 0.5038e_{t-3} - (-0.0455)e_{t-4} - (-0.8381)e_{t-5} - 0.0460e_{t-6} - 0.0393e_{t-7}$

### 3.3.1 Residual Analysis

After deriving tentative models for the data, residual analysis are performed. These diagnostics are performed to certify the assumption that the residuals are random and unpredictable. The randomness and unpredictability of residuals will validate the model in order to carry on with other analyses. Residual plots of the model are derived and analyzed for this purpose.

The graph of the standardized residuals shows a random distribution of residuals around the horizontal line at 0. This is a good indication that no particular patterns exist within the residual plots. We can confidently say from this plot that the mean of the residuals of both models up to this point is zero. It is worthy to note that there is an indication of the existence of outliers within the data. This is analyzed in the next section.

The plot also gives an indication of constant variance considering the non-

Figure 3.6: Residual Plots- Plots of Standardized Residuals

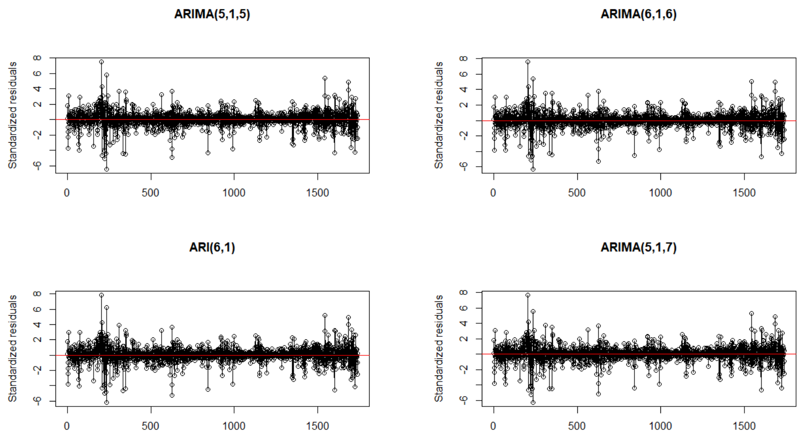
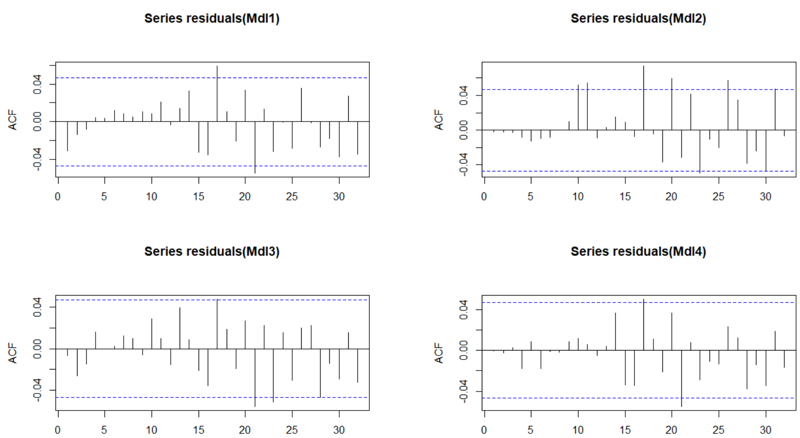



Figure 3.7: Residual Plots - ACF Plots of Residuals




existence of a particular pattern.

The graph of the ACF of the residuals reveals if there is any autocorrelation in the residuals. This will usually suggest whether or not there is information that has not been accounted for by the model. From the graph above in figure 3.7, there is no indication of the existence of autocorrelation within the residuals of the models. We can safely assume therefore that the residuals of both models are free from autocorrelation. A formal test is carried out to confirm this derivation.

Ljung-Box Tests are performed on all models with the following hypothesis.

$H_0$ : The data are independently distributed; there is no correlation.

$H_a$ : The data are not independently distributed; they exhibit serial correlation.

The resulting p-values from the tests on all models are as follows:

| Model | X-squared | DF | P-value | Decision |
|---|---|---|---|---|
| ARIMA(5,1,5) | 7.8028 | 5 | 0.1674 | Fail to reject null |
| ARI(6,1) | 11.44 | 9 | 0.2467 | Fail to reject null |
| ARIMA(6,1,6) | 8.352 | 9 | 0.4991 | Fail to reject null |
| ARIMA(5,1,7) | 6.1586 | 9 | 0.724 | Fail to reject null |

At a significance level of 0.05, we fail to reject the null hypothesis and conclude that for both models, there are no signs of autocorrelation of residuals.

As stated earlier in the limitations, our tests for normality failed for all transformations and models. Multiple transformations were attempted but failed to satisfy the normality assumption. QQ-plots are shown in the appendix. This is the greatest setback of our modeling. It is therefore worth noting that the model is to be used with caution.

### 3.3.2 Outliers

Prior to carrying out any further step, outlier analysis is performed now that potential models exist. Two possible outliers are assumed to exist and thus tested.

- Additive outlier - these are those outliers that are as a result of some additive effect. Thirty outliers are detected in the data.
- Innovational outlier- these outliers are those that do not result from some mere additive effect and can be more complicated reversing. Twenty-nine innovative outliers are detected.

From analysis, we observed that the outliers detected as innovative outliers are a complete subset of those detected as additive outliers. We therefore assume that there are thirty outliers in the dataset. It is needful to analyse these outliers to make a decision on how to treat them. A careful analysis of outliers indicated that the model interprets spikes, which are to be expected in this kind of dataset, as outliers. For example, after a run of daily prices in the region of $300, the model flags an amount of $600 immediately following this run as an outlier because it deems $600 an unusually high closing figure to follow a run of $300s. However, this phenomenon actually characterizes such a volatile financial instrument like Bitcoin. We therefore refrain from eliminating these reported outliers from our analysis and proceed without modifying our dataset.

## 3.4 Model Validation

### 3.4.1 Test for Overfitting

The next stage in our analysis was to check for the need for extra coefficients in the model. We perform this check to ascertain whether or not there is the need to include extra lags and thus extra coefficients (which make the model more complicated) in order for the model to fit the data more perfectly. This is done by increasing the various lags of the proposed models consecutively. The significance of the estimates of these models are observed. If the estimate is deemed insignificant, then we can confidently say that there is no need to adjust the models to reflect the extra lags.

An estimate is insignificant if —coefficient/standard error—$< 1.96$.

The results of this test shown in the appendix indicate that there is no need for extra coefficients for any other the models. We therefore carry on with

- Model 1 - ARIMA(5,1,5)
- Model 2 - ARI(6,1)
- Model 3 - ARIMA(6,1,6); and
- Model 4 - ARI(5,1,7)

### 3.4.2 Prediction

The essence of having a model for Bitcoin is to be able to effectively predict future prices with little or no errors. The existence of errors within a model's prediction, though it cannot be completely eliminated, should be as minimal as possible. In the beginning of this paper, in the data section, it is stated that the initial data ranging from April 28, 2013 to January 31, 2018 is used to train the model. The remaining data therefore remains as the test data.

Given that the prices of Bitcoin are very time sensitive, we make predictions within a thirty (30) day bracket. Predictions beyond this bracket will be very erroneous because of the nature of the market.

The results from predictions are then compared with the actual data which is kept from the model. The comparison of the results for both models is used to determine which of the two models works better or is more efficient for the purpose of predicting prices.

A plot of the predicted prices against the actual prices in Figure 3.8 shows the accuracy of all four models.

A formal test of accuracy is performed on the predicted values. The following metrics are derived from these values and compared between the models.

The table below shows the results from prediction and compares the model predictions with the test data.

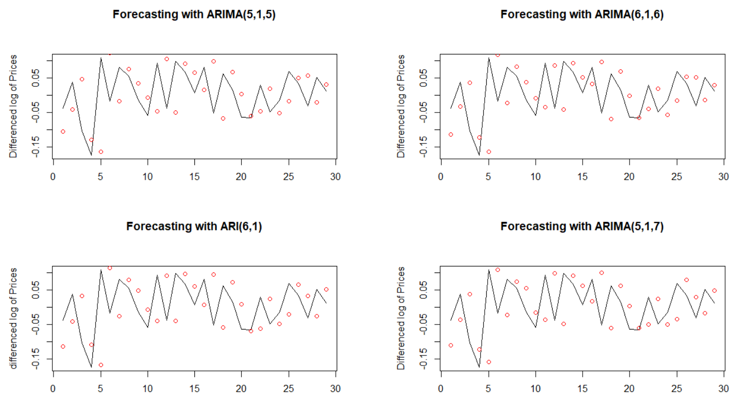



Figure 3.8: Plot of Actual Prices vs Predicted Prices

| Model | ME | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| ARIMA(5,1,5) | 53.90069 | 615.2435 | 540.5356 | 0.3309236 | 5.830268 |
| **ARI(6,1)** | **31.03189** | **620.2905** | **536.1872** | **0.0953679** | **5.798926** |
| ARIMA(6,1,6) | 45.8499 | 610.0216 | 531.1792 | 0.2561366 | 5.736502 |
| ARIMA(5,1,7) | 44.66699 | 620.5146 | 536.8497 | 0.2555175 | 5.799093 |

From the table above, we conclude that Model 3, that is, an ARI(6,1) best fits the data.A plot of predictions (in red) against the actual prices from February 1 2018 to March 2 2018 using this model is shown below:

95% confidence intervals of these predictions are attached in section A.2 of the appendix.

It is evident from the plot that although the predictions are not perfect, they generally follow the same pattern as the actual prices.

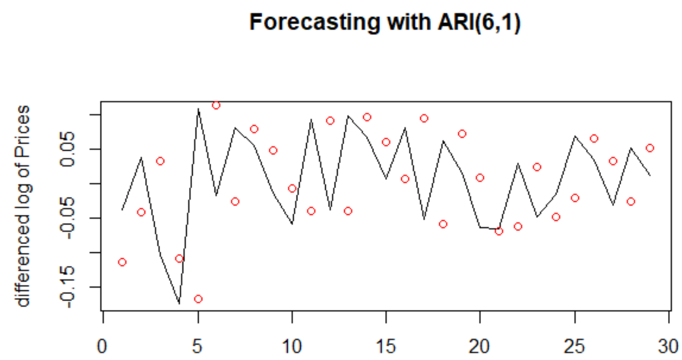



Figure 3.9: Plot of Actual Prices vs Predicted Prices

### 3.4.3 How many days ahead can we predict efficiently

It is essential to know how many days ahead the model effectively forecasts. Although it will be preferred that the prices of the Bitcoin is known way ahead of time in order to avoid getting into future losses, it is unrealistic to have this expectation. The market is highly uncertain and as such, the behavior of prices cannot be predicted too far into the future. We set an initial cap of 30 days as the longest a prediction should be made.

In this section, we want to decide on the level of effectiveness of the model as the length of prediction changes. We assume that we do not have access to any data other that what is available today and make predictions for one day ahead, two days ahead, all the way to thirty days ahead. The resulting metrics are compared and at a reasonable tolerance level, a decision can be made on how far predictions should be made.

The resulting cumulative error metrics when predictions are made 30 days in advance are much worse that if they were made one day at a time, for 30 days. The metrics considered are

- Mean Error (ME)
- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- Mean Percentage Error (MPE)
- Mean Absolute Percentage Error (MAPE)

The results are shown in the table below:

| Prediction Method | ME | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| 30 days ahead | 552.477 | 1325.68 | 1077.577 | 5.434173 | 10.63176 |
| 1 day at a time | 31.03189 | 620.2905 | 536.1872 | 0.0953679 | 5.798926 |

The one day at a time prediction is carried out by making predictions only for the next day. The results of that day is then known at the close of the day and then we can make the prediction of the day that follows. This is done for 30 days.

The 30 day ahead prediction on the other hand is performed by predicting 30 days from the onset. At a tolerance level of $+/-\$1000$, we believe that it is too risky to predict Bitcoin prices beyond one day.

# Chapter 4

# Relationship of Bitcoin with Altcoins and Other Investment Instruments

At the beginning of this research, we made assumptions relating to the possible relationships that may exist between the major cryptocurrency, Bitcoin, and some of the other cryptocurrencies that are leading in the market. The cryptocurrencies selected for this purpose are selected on the basis of which has the highest market share. We select

- Bitcoin;
- Bitcoin cash;
- Litecoin;
- Ethereum; and

- Ripple.

These cryptocurrencies together make up about 75% of the total market share of cryptocurrencies and therefore dominate the market.

To analyze the relationships between the major cryptocurrencies, we consider the correlation between the daily returns of Bitcoin, Litecoin, Ethereum and Ripple over time. We eliminate Bitcoin Cash at this stage because it does not offer as much data as needed to draw any meaningful conclusions. We calculate the daily returns from the daily price data collected from coinmarketcap.com. We go ahead to segregate the returns data into sections containing 60 data points on each of the cryptocurrencies except Bitcoin Cash with the justification stated above. We elect to analyze the relationships over 60 days to give us a fair reflection of the correlations over a reasonable amount of time. A similar analysis conducted by coindesk.com on its website titled "Follow the Leader: Analyzing Cryptocurrency Price Correlations" uses quarterly data to analyze relationships. Using a shorter time frame helps us to compare and contrast their observations and make conclusions either to corroborate or contradict those results. Our first sub-dataset begins on September 29, 2015 and ends on November 27, 2015. The subsequent sub-datasets follow in that order till we reach March 16, 2018. Segregating the dataset in this fashion produces 15 sub-datasets to analyze. After running the correlation, we observe the following:

## 4.1 Relationship Between Bitcoin and Litecoin

**Bitcoin returns are generally positively correlated with Litecoin returns**

In all of the 15 sub-datasets produced, we observe that Bitcoin records a negative correlation with Litecoin only once. There are few sub-datasets that have especially low correlation values. In general, Bitcoin and Litecoin are fairly positively correlated. In fact, 8 of the 15 sub-datasets produced produced correlation values that are 70% or higher. This is an interesting phenomenon and corroborates the conclusion reached by the coindesk.com article. Many refer to Litecoin as the silver to Bitcoin's gold. This is because the two coins are similar in many respects. Even though the correlation between the two, in our analysis and in many others conducted by other researchers, indicate an overall positive relationship, this is not always the case as there exist some periods when Litecoin has not responded to Bitcoin's surge in similar fashion in terms of magnitude.

## 4.2 Relationship Between Bitcoin and Ripple and Ethereum

**Bitcoin versus Ripple and Ethereum**

It is the assertion among researchers that Ripple and Bitcoin have tradition-

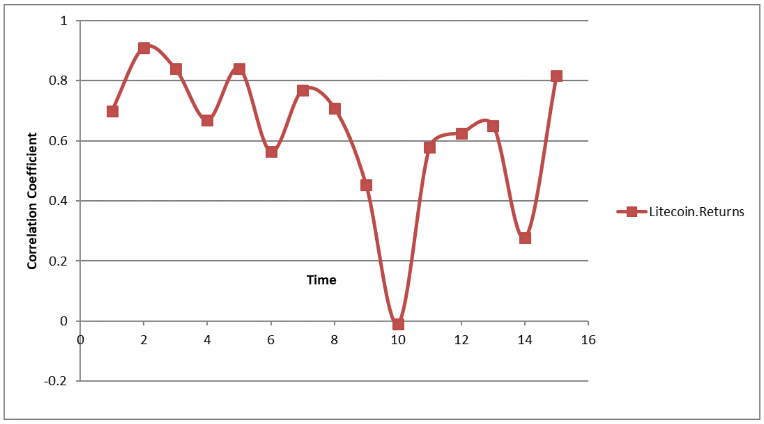



Figure 4.1: Plot of Correlation between Bitcoin and Litecoin

ally been negatively correlated. The fundamental difference in the model of Bitcoin and Ripple, the technology on which both coins are based and the audience they both target are cited as justification for this negative correlation. Ripple has an especially low correlation with Bitcoin. Out of the 15 sub-datasets on which correlation of returns was done, 13 sub-datasets recorded correlation coefficients lower than 50%. In fact, 11 of these are 30% or below. The only reasonably high correlation coefficient is a 76% correlation in the 15th sub-dataset. The assertion that Ripple offers an alternative for cryptocurrency investors to move their money when shocks in Bitcoin prices are experienced is corroborated.

The experience of Ethereum is not easily discernible. While there has been as high correlation as 90% in one sub-dataset, most of the correlation coefficients over the sub-datasets have fallen between -20% and 45%. The Coindesk analysis posits that using a shorter time frame of 7 days creates correlation over 90% in one 7-day period.

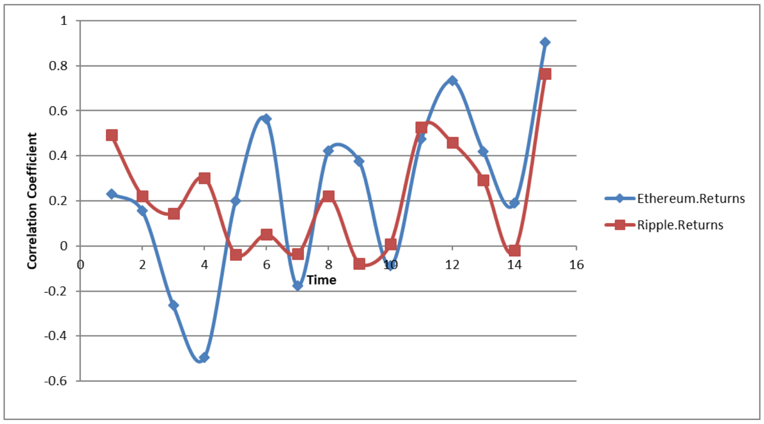



Figure 4.2: Plot of Correlation between Bitcoin and Ripple and Ethereum

## 4.3 Relationship between Bitcoin and Wider Investment Market

**Bitcoin is generally not indicative of the wider market position** Burr, Hong and Lee (2017) concluded that Bitcoin had different characteristics from other investment assets and could be used for diversification purposes. They also assert that majority of Bitcoin holders hold the instrument for investment purposes rather than for transactional purposes.

We employ a similar approach to that used to analyze relationships between Bitcoin and the major Altcoins to examine the relationship between Bitcoin and the wider investment market. We elect to use the returns on the S&P 500 and the Dow Jones to measure the performance of the market. Returns data segregation for Bitcoin and the S&P 500 into sections of 60 data points each is done and analysis on the correlation between the returns in those sub-datasets is performed. Our sub-datasets begin from March 18 2013 to

March 16 2018, a time span of 5 years. The methodology for segregating the data produces the sub-datasets on which the correlation of prices of Bitcoin and the S&P 500 and the Dow Jones are measured.

The first observation we make is that correlation between Bitcoin and the S&P 500 mimics the correlation between Bitcoin and the Dow. This is to be expected as the S&P 500 and the Dow both measure market performance. From Figure 4.3. below, both market indices move in tandem with each other so far as their correlation with Bitcoin during the various time periods are concerned.

Generally, the correlation between Bitcoin and both S&P 500 and the Dow are low. This indicates that we cannot be emphatic that the market has a strong relationship with Bitcoin. There are a few sub-datasets that record negative correlation, albeit low. Bitcoin seems to not be a direct alternative to investing in the market. However, there are certain instances where a loss of confidence in the financial system has led to people shifting their attention to the cryptocurrency as a store of value. An example is when Zimbabwe recorded high Bitcoin prices as its citizens lost confidence in the financial system and sought to use Bitcoin as a store of value as cited on bitcoin.com

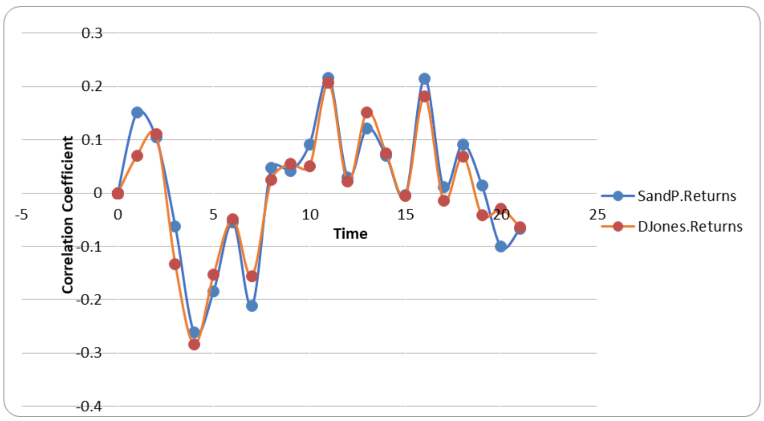



Figure 4.3: Plot of Correlation between Bitcoin and S&P 500 and Dow Jones Index

# Chapter 5

# Conclusion

The aim of this study was to examine the behavior of Bitcoin, the foremost cryptocurrency, through a statiscal lens. Our analysis centered on finding a time series model that aptly describes the movement of Bitcoin and that would enable us to make predictions as to what the price would be over a period of time.

We also tested the veracity of the following hypothesis:

- That there exists some relationship between the movement of Bitcoin and the other major cryptocurrencies.
- That returns on Bitcoin is negatively correlated with at least one of the major stock indices (S&P 500 and the Dow).

To answer these questions, we performed a time series analysis of the daily closing price of Bitcoin over a 5 year period and came to the following conclusions:

- That an ARI(6,1) is an accurate enough time series model to fit the data experienced during the 5 year period
- That the most realistic time frame to base the prediction of Bitcoin prices was daily. Any prediction beyond a day would be subject to too much volatility in the price and therefore is not advisable.

The other hypotheses tested also revealed that:

- The Altcoin that is most correlated with Bitcoin is Litecoin. This cryptocurrency mimics Bitcoin's movement more than all the other Altcoins examined
- Ripple has a poor linear correlation with Bitcoin and offers an alternative to cryptocurrency investors looking to redirect their investment from Bitcoin while still staying the in the cryptocurrency space
- The S&P 500 and the Dow, both widely used to measure the health of the financial market, do not enjoy a good linear correlation with Bitcoin.

It has to be stated that due to the very volatile nature of cryptocurrencies, these results are susceptible to change in the short to medium term. As such, the results discussed should be relied upon cautiously.

# Bibliography

[1] Agrawal, J.(2018). 8 Benefits of Blockchain to Industries Beyond Cryptocurrency, Entrepreneur, Available at: https://www.entrepreneur.com/article/306420(Accessed on March 1,2018).

[2] Back, A. (2002). Hashcash - A Denial of Service Counter-Measure. Available at: http://www.hashcash.org/papers/hashcash.pdf

[3] Bellare, M., Rogaway, P. (2005). Introduction to Modern Cryptography, Available at: http://web.cs.ucdavis.edu/~rogaway/classes/227/spring05/book/main.pdf

[4] Bovaird, C. (2017). Follow the Leader: Analyzing Cryptocurrency Price Correlations, Coindesk, Available at: https://www.coindesk.com/follow-leader-analyzing-bitcoin-cryptocurrency-price-correlations/ (Accessed on March 1,2018).

[5] Chaum, D. (1983). Blind signatures for untraceable payments. Advances in Cryptology Proceedings of Crypto 82 (3). Available at: http://sceweb.sce.uhcl.edu/yang/teaching/csci5234WebSecurityFall2011/Chaum-blind-signatures.PDF

[6] Haber, S., Stornetta, W. S., (1997). Secure names for bitstrings. Proceedings of the 4th ACM Conference on Computer and Communication Security. Available at: http://nakamotoinstitute.org/static/docs/secure-names-bit-strings.pdf

[7] Iansiti, M., Lakhani, K. R. (January 2017). The Truth About Blockchain, Harvard Business Review, Harvard University. Available at: https://hbr.org/2017/01/the-truth-about-blockchain

[8] Available at:

[9] Nakamato, S. (2008). A peer-to-peer electronic cash system.Available at: https://bitcoin.org/bitcoin.pdf (Accessed on March 1, 2018).

# Appendix A

## A.1 Residual Plots

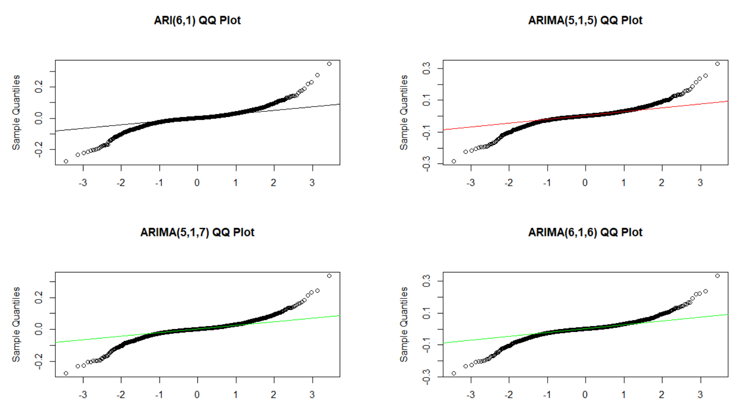



Figure A.1: QQ Plot of Residuals of ARI(6,1)

## A.2 Confidence Intervals for Predictions

| Day | Prediction | Low | High | Actual Price |
|---|---|---|---|---|
| 1 | 10,290.59 | 9,857.53 | 10,723.66 | 9,170.54 |
| 2 | 9,185.78 | 8,752.72 | 9,618.84 | 8,830.75 |
| 3 | 8,817.87 | 8,384.81 | 9,250.94 | 9,174.91 |
| 4 | 9,118.25 | 8,685.19 | 9,551.32 | 8,277.01 |
| 5 | 8,175.76 | 7,742.69 | 8,608.82 | 6,955.27 |
| 6 | 6,914.84 | 6,481.77 | 7,347.90 | 7,754.00 |
| 7 | 7,749.56 | 7,316.50 | 8,182.63 | 7,621.30 |
| 8 | 7,557.04 | 7,123.97 | 7,990.10 | 8,265.59 |
| 9 | 8,188.77 | 7,755.71 | 8,621.83 | 8,736.98 |
| 10 | 8,597.46 | 8,164.39 | 9,030.52 | 8,621.90 |
| 11 | 8,541.30 | 8,108.24 | 8,974.37 | 8,129.97 |
| 12 | 8,208.34 | 7,775.28 | 8,641.41 | 8,926.57 |
| 13 | 8,997.12 | 8,564.05 | 9,430.18 | 8,598.31 |
| 14 | 8,645.07 | 8,212.00 | 9,078.13 | 9,494.63 |
| 15 | 9,519.90 | 9,086.83 | 9,952.96 | 10,166.40 |
| 16 | 10,124.71 | 9,691.64 | 10,557.77 | 10,233.90 |
| 17 | 10,196.24 | 9,763.17 | 10,629.30 | 11,112.70 |
| 18 | 11,208.57 | 10,775.51 | 11,641.64 | 10,551.80 |
| 19 | 10,572.73 | 10,139.66 | 11,005.79 | 11,225.30 |
| 20 | 11,377.27 | 10,944.20 | 11,810.33 | 11,403.70 |
| 21 | 11,478.42 | 11,045.35 | 11,911.48 | 10,690.40 |
| 22 | 10,721.92 | 10,288.86 | 11,154.99 | 10,005.00 |
| 23 | 10,075.10 | 9,642.04 | 10,508.17 | 10,301.10 |
| 24 | 10,328.88 | 9,895.82 | 10,761.95 | 9,813.07 |
| 25 | 9,838.18 | 9,405.12 | 10,271.25 | 9,664.73 |
| 26 | 9,634.94 | 9,201.87 | 10,068.00 | 10,366.70 |
| 27 | 10,296.02 | 9,862.96 | 10,729.09 | 10,725.60 |
| 28 | 10,649.04 | 10,215.98 | 11,082.11 | 10,397.90 |
| 29 | 10,373.81 | 9,940.75 | 10,806.88 | 10,951.00 |
| 30 | 10,939.01 | 10,505.94 | 11,372.07 | 11,086.40 |

Figure A.2: Predictions and Confidence Intervals

## A.3 Table with significance of extra terms

| Model | Overfitting | Estimate of Extra Term | S.E. of Extra Term | \|Est./S.E.\| |
|---|---|---|---|---|
| ARIMA(6,1,6) | ARIMA(6,1,7) | 0.0271 | 0.0389 | 0.6967 |
| | ARIMA(7,1,6) | 0.0024 | 0.0275 | 0.0873 |
| ARI(6,1) | ARI(7,1) | -0.0045 | 0.0242 | 0.186 |
| ARIMA(5,1,7) | ARIMA(6,1,7) | -0.441 | 0.9486 | 0.4649 |
| | ARIMA(5,1,8) | -0.0093 | 0.028 | 0.3321 |

## A.4 Model Estimates

```
Call:
arima(x = log(ts.Bitcoin), order = c(5, 1, 5), xreg = NULL, include.mean = TRUE,
    init = NULL, method = "ML")

Coefficients:
          ar1     ar2      ar3     ar4     ar5     ma1      ma2     ma3     ma4
      -0.0503  0.3657  -0.3884  0.0231  0.9269  0.0909  -0.3883  0.3738  0.0301
s.e.   0.0128  0.0158   0.0120  0.0183  0.0207  0.0183   0.0083  0.0141  0.0255
          ma5
      -0.9042
s.e.   0.0257

sigma^2 estimated as 0.001929:  log likelihood = 2966.18,  aic = -5912.37
```

Figure A.3: Model Estimates of ARIMA(5,1,5)

```
Call:
arima(x = log(ts.Bitcoin), order = c(6, 1, 6), xreg = NULL, include.mean = TRUE,
    init = NULL, method = "ML")

Coefficients:
          ar1      ar2     ar3      ar4     ar5     ar6     ma1     ma2      ma3
      -0.2703  -0.2256  0.1021  -0.0495  0.5059  0.7524  0.2815  0.2220  -0.1000
s.e.   0.2755   0.1412  0.1448   0.1557  0.1495  0.2420  0.2814  0.1462   0.1438
          ma4      ma5      ma6
       0.0670  -0.4578  -0.7012
s.e.   0.1516   0.1580   0.2278

sigma^2 estimated as 0.001936:  log likelihood = 2964.14,  aic = -5904.28
```

Figure A.4: Model Estimates of ARIMA(6,1,6)

```
Call:
arima(x = log(ts.Bitcoin), order = c(6, 1, 0), xreg = NULL, include.mean = TRUE,
    init = NULL, method = "ML")

Coefficients:
         ar1      ar2      ar3     ar4     ar5     ar6
      0.0081  -0.0375  -0.0116  0.0374  0.0597  0.0643
s.e.  0.0239   0.0239   0.0240  0.0241  0.0241  0.0242

sigma^2 estimated as 0.001969:  log likelihood = 2949.56,  aic = -5887.12
```

Figure A.5: Model Estimates of ARI(6,1)

```
Call:
arima(x = log(ts.Bitcoin), order = c(5, 1, 7), xreg = NULL, include.mean = TRUE,
    init = NULL, method = "ML")

Coefficients:
          ar1     ar2      ar3     ar4     ar5     ma1      ma2     ma3      ma4
      -0.1243  0.4342  -0.5193  0.1107  0.8582  0.1294  -0.4705  0.5038  -0.0455
s.e.   0.1345  0.1254   0.0808  0.0487  0.0680  0.1378   0.1281  0.0634   0.0638
          ma5     ma6     ma7
      -0.8381  0.0460  0.0393
s.e.   0.0592  0.0386  0.0459

sigma^2 estimated as 0.001932:  log likelihood = 2966.02,  aic = -5908.04
```

Figure A.6: Model Estimates of ARIMA(5,1,7)