The Future of Money: A Statistical Analysis on Cryptocurrencies

Ampadu-Kissi, Ama
Owusu, Isaac

Illinois State University

akampdu@ilstu.edu
iaowusu@ilstu.edu

April 1, 2018


The Future of Money

Overview

Cryptocurrency

A cryptocurrency[1] 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.
- Ownership of cryptocurrency units can be proved exclusively cryptographically.
- The system allows transactions to be performed in which ownership of the cryptographic units is changed.
- If two different instructions for changing the ownership of the same cryptographic units are simultaneously entered, the system performs at most one of them.

[1] https://hbr.org/2017/01/the-truth-about-blockchain

Bitcoin

- Bitcoin is the first and most prominent cryptocurrency
- All other cryptocurrencies are generally referred to as Altcoins
- Ethereum, Bitcoin Cash, Ripple and Litecoin are some of the leading Altcoins in the market

Research Interest

- Build a simple time series model to fit Bitcoin data
- Make predictions with derived model
- Study relationships that exist between the major cryptocurrencies
- Study the relationship that exists between Bitcoin and market indices

Research Interests

Goal 1

Build a simple time series model to fit Bitcoin data

Goal 2

Make predictions with derived model

Goal 3

Study relationships that exist between the major cryptocurrencies

Goal 4

Study the relationship that exists between Bitcoin and market indices

Data

- Primary data on cryptocurrency is derived from Coin Market Cap
- Each cryptocurrency data contains information on
  - Date
  - Opening price
  - High and low prices
  - **Closing price**
  - Volume
  - Market cap
- Daily returns on prices are calculated from closing prices

Assumptions

Assumptions

Data is reviewed and we conclude that it is accurate.

1. The series is stationary.
2. Errors are randomly distributed with a constant mean and variance over a long period of time.
3. Shocks, if present, are randomly distributed over time with a mean of 0 and a constant variance.

Time Series Plot

Initial Time Series plot of Bitcoin Prices

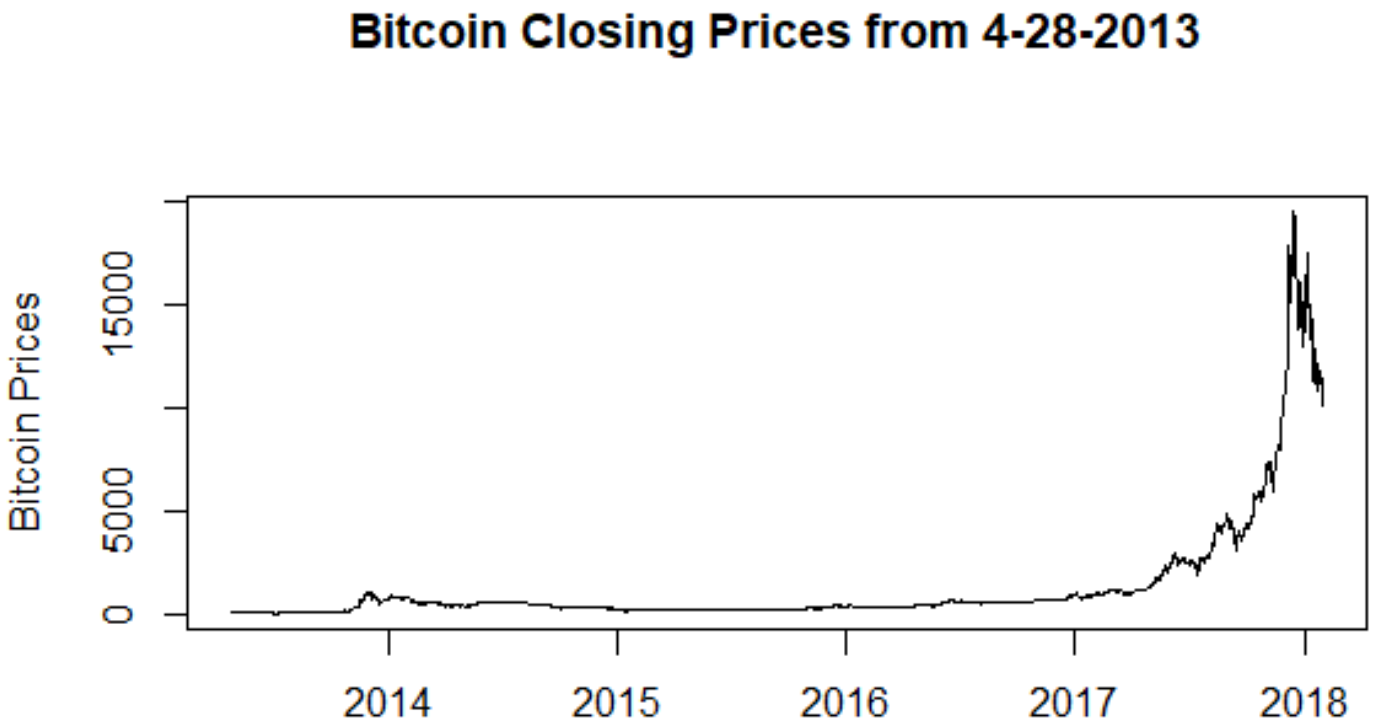


Figure: Time Series Plot of Bitcoin Prices

Test for Stationarity

Dickey-Fuller Test for Stationarity:

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

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

Raw data is non-stationary data and therefore suggests some data transformation is needed. Transformations considered:

- Log Transformation
- Difference Transformation
- Difference of Log Transformation

Transformation

Difference of log served as the best transformation.
Resulting data is stationary with no particular pattern.

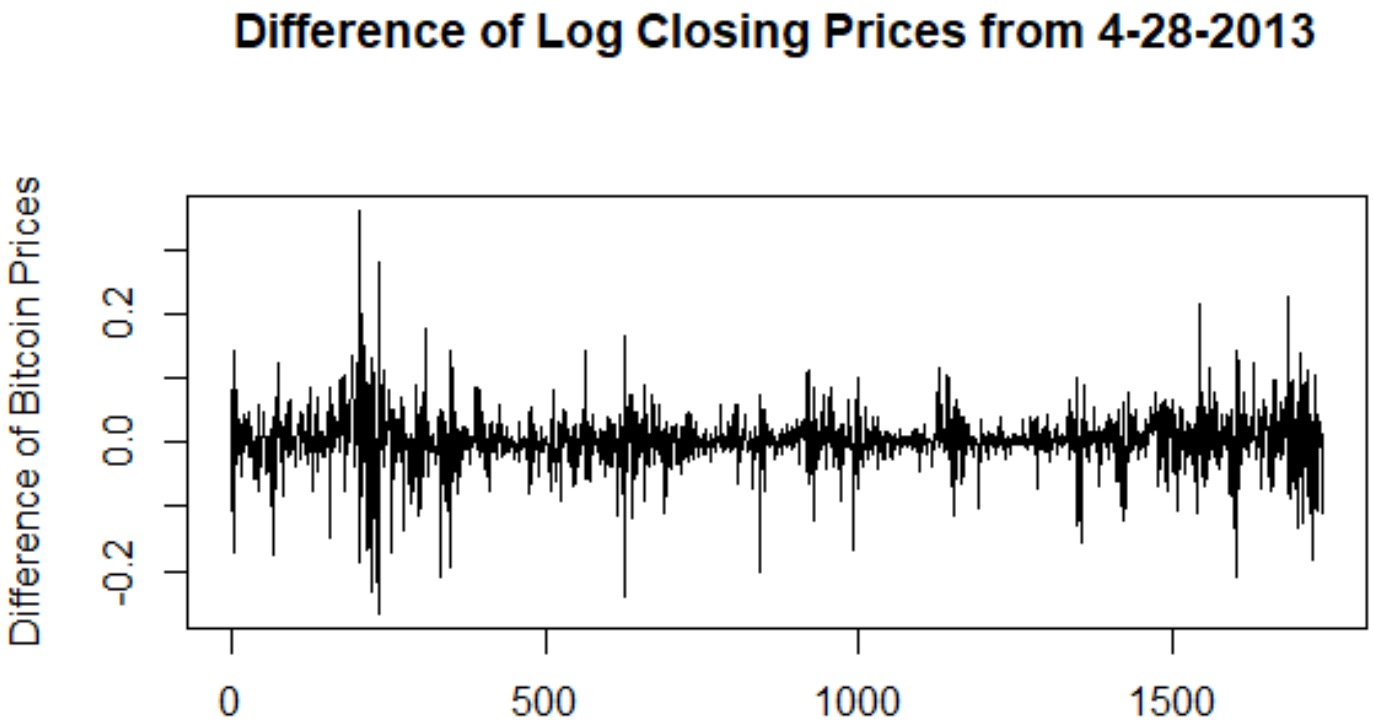


Figure: Time Series Plot of Bitcoin Prices after Difference of Log Transformation

Model Building

Functions and Models[2]

Model selection functions

1. Autocorrelation function (ACF)
2. Partial ACF (PACF)
3. Extended ACF (EACF)
4. Subset function

Resulting Models

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

[2] ARIMA - Autocorrelation/Moving Average

Plots of Functions

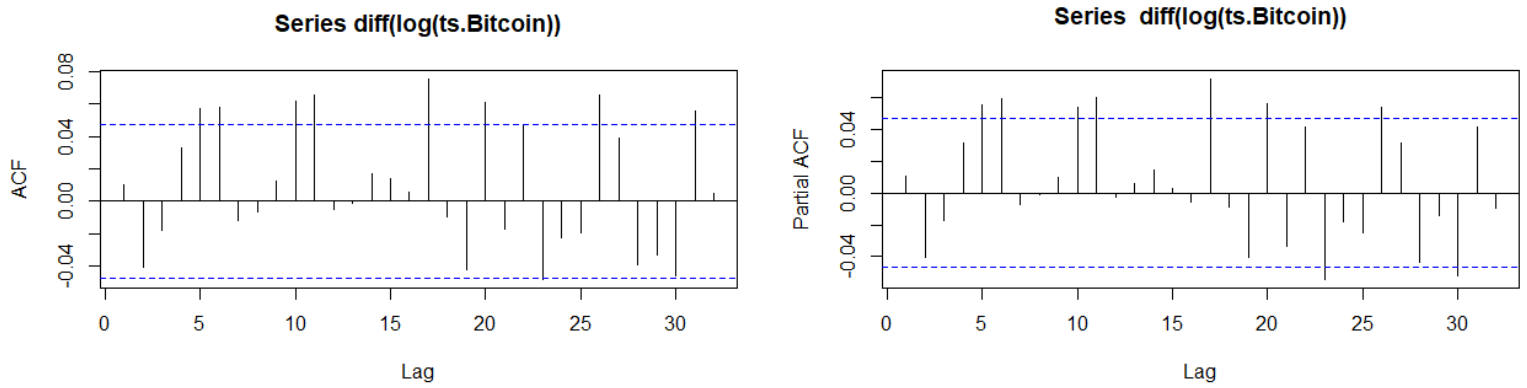


Figure: Plots of ACF and PACF

Plots of Functions

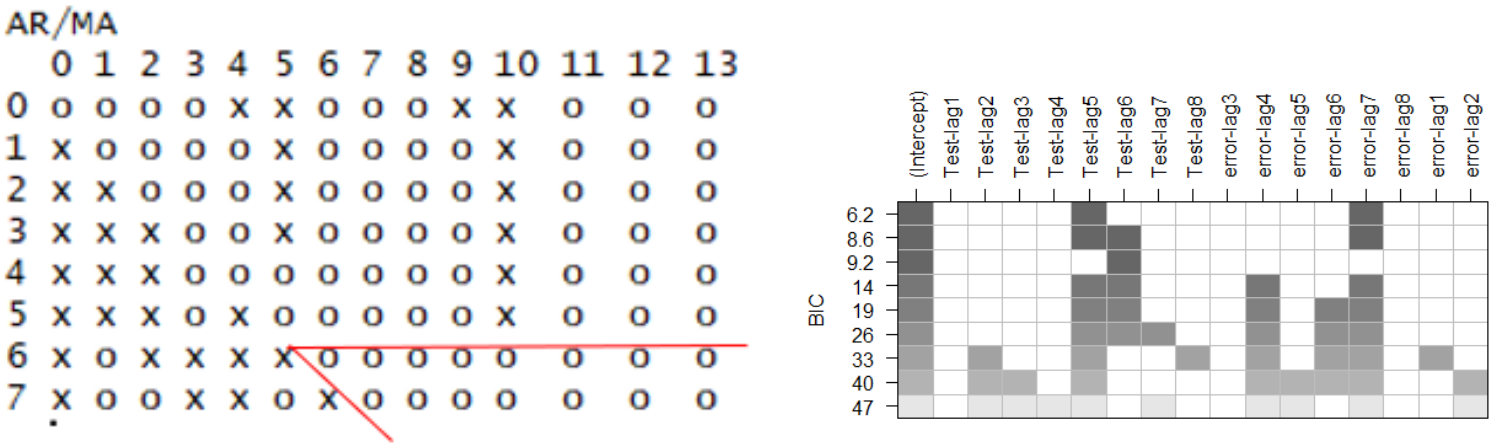


Figure: Plots of EACF and Model Subsets

Tentative Model Estimates

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

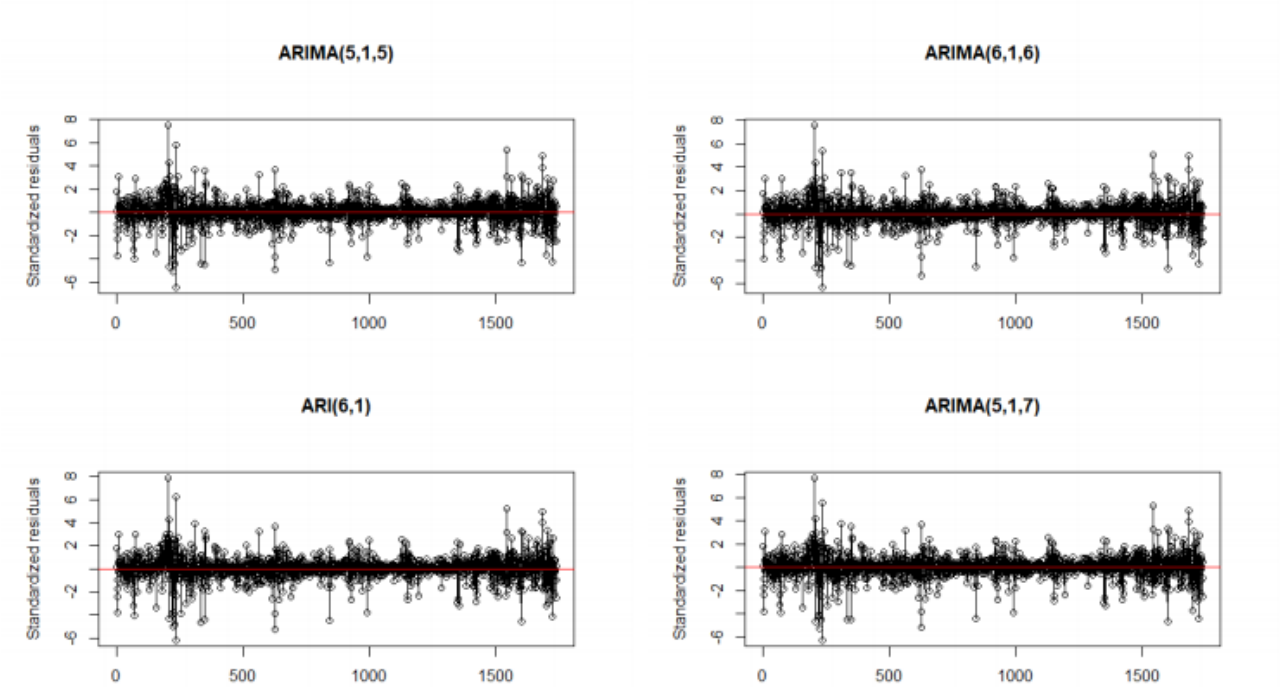
$$\text{ARIMA}(6,1,6): Y_t = -0.2703Y_{t-1} - 0.2256Y_{t-2} + 0.1021Y_{t-3} - 0.0495Y_{t-4} + 0.5059Y_{t-5} + 0.7524Y_{t-6} + e_t - 0.2815e_{t-1} - 0.2220e_{t-2} - (-0.1000)e_{t-3} - 0.0670e_{t-4} - (-0.4578)e_{t-5} - (-0.7015)e_{t-6}$$

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

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

Residual Analysis

Plots of standardized residuals are randomly distributed indicating a constant variance and a mean approaching 0.



Test of Independence of Residuals

Ljung-Box Tests are performed on 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

$\therefore$ Residuals are all uncorrelated.

Outliers Investigation

- Outliers detected as innovative outliers are a complete subset of those detected as additive outliers
- Careful analysis 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
- We therefore refrain from eliminating these reported outliers.

Prediction vs Actual of Different Models

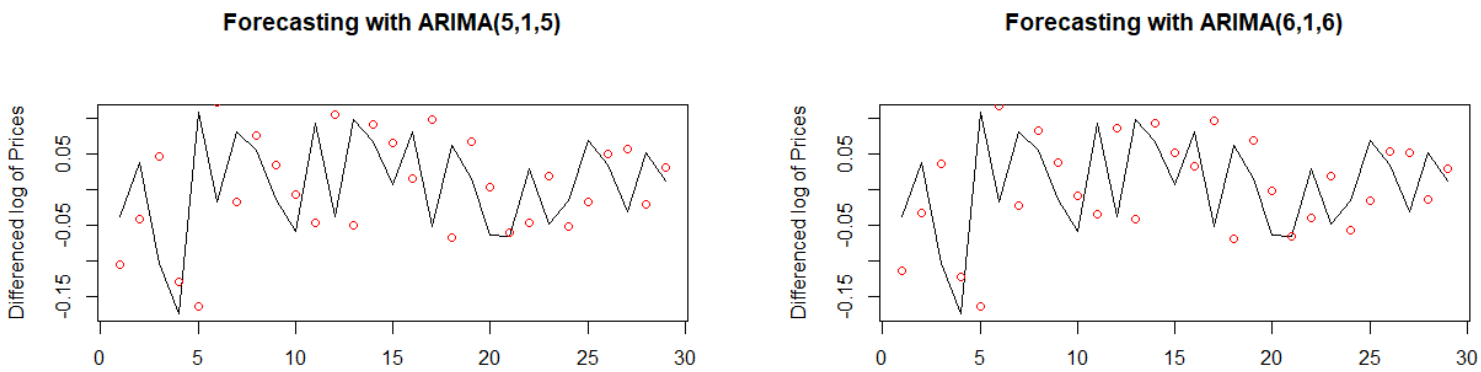


Figure: Plot of Actual Prices vs Predicted Prices

Prediction vs Actual of Different Models

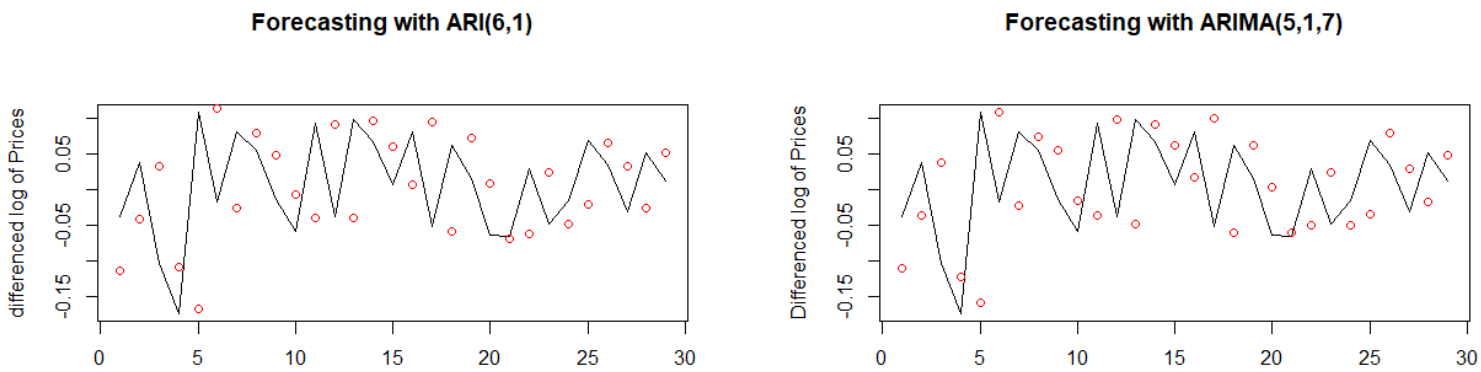


Figure: Plot of Actual Prices vs Predicted Prices

Model Selection using Error Metrics

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

ARI(6,1) turns out to be the best model that fits the data.

How many days ahead to predict

Caution!

Making predictions beyond one day at a time results in very huge losses.

Volatile

Voalatility measured by standard deviation is $1017.

Tolerance

Consider your tolerance level before making decisions from predictions that go too far into the future.

Bitcoin vs Altcoins

An analysis of the daily returns on Bitcoin versus that of the major alternatives (also referred to as Altcoins) revealed the following:

- The Altcoin that has the most positive correlation with Bitcoin over the period analysed is Litecoin
- Bitcoin and Ripple share a weak correlation in their daily returns figures over the period of the analysis
- Ethereum and Bitcoin do not have any discernible relationship, sometimes sharing a good correlation and at other times showing low correlation

The analysis segregated cryptocurrency daily returns into 60 day periods and examined correlations in those time periods.

Bitcoin vs Altcoins Cont'd

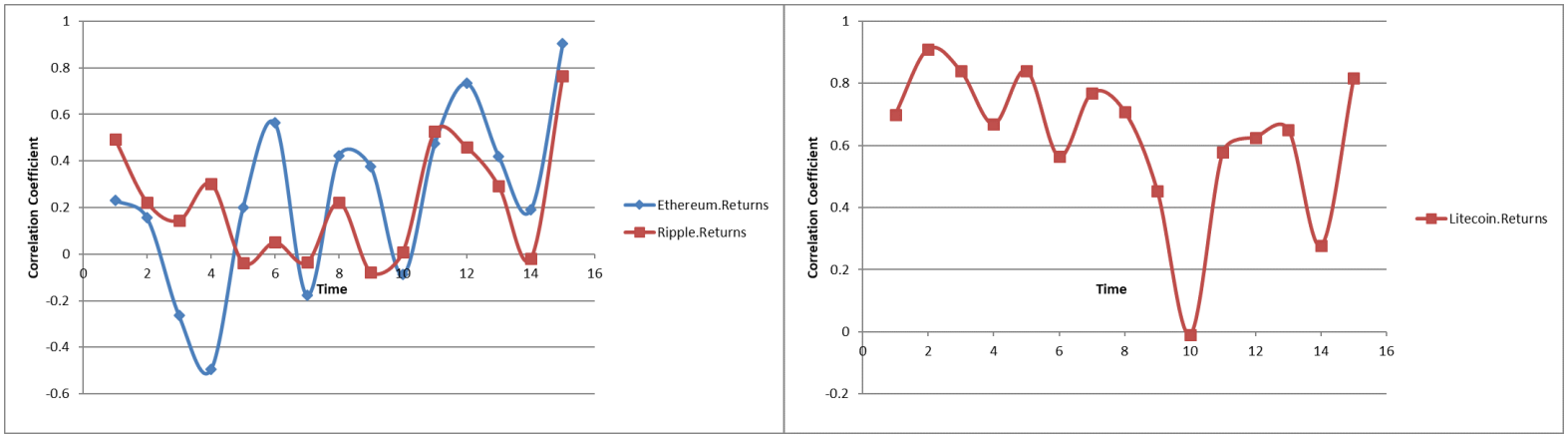


Figure: Bitcoin Relationship with Altcoins

Bitcoin vs Market Indices

A similar analysis to Bitcoin vs Altcoins is conducted between Bitcoin and the S&P 500 and the Dow Jones Industrial Average. The same data segregation method is used as described in the previous analysis.

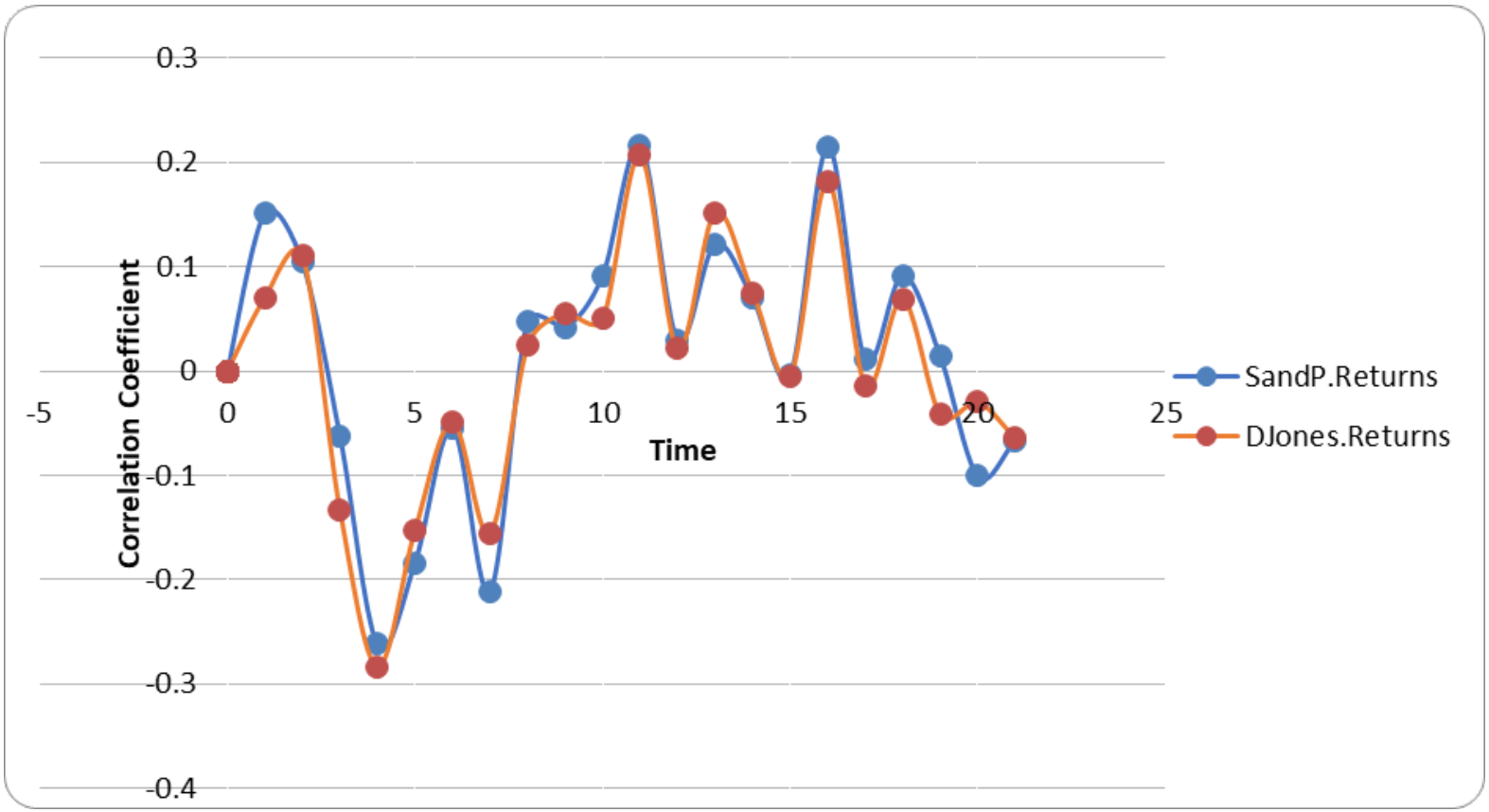


Figure: Bitcoin Relationship with Market Indices

Bitcoin vs Market Indices Cont'd

The findings are as follows:

- As expected, the S&P 500 and Bitcoin move in tandem as far their correlation with returns on Bitcoin is concerned
- Bitcoin has a poor correlation with the two market indices
- Despite this lack of a strong correlation, Bitcoin has been used in some financial markets to store value when inflation in fiat currencies have escalated

[3] Credit : TechCrunch