SCSUG Paper 2018

Decyphering cryptocurrencies: Sentiments and prices

Priyansh Dadar, Oklahoma State University, Stillwater, OK

ABSTRACT

Cryptocurrencies like Bitcoin have recently gained a lot of attention from media as well as the market. The feature that made them stand out from standard currency was the fact that it was fluctuating with high variation. This fluctuation is not institutionalized but depends on how people perceive them. The impending question always has been centered around whether to treat them as a currency or a commodity. Their prices have been soaring the skies for the past year but took a nosedive this February. Analyzing the sentiments that people have for cryptocurrency might help us understand the underlying factors that affect its popularity and price. Social media like Twitter and other websites are a host to a lot of data to analyze these sentiments. These sources also provide the change in rates of cryptocurrencies throughout the period under consideration. Data from these sources will be used through web scrapping and APIs to procure the data and proceed further with the analysis.

INTRODUCTION

Cryptocurrencies have taken the financial market by storm by decentralizing the standard transaction procedure. The decentralization was beneficial in preventing the attenuation of the economy through fraud and corrupt practices by eliminating the middlemen. However, the variety of electronic currencies that have flooded the market has been raising concerns regarding the stability of the coin. Though there are more than 40 types of cryptocurrencies, Bitcoin reigns one-third of the digital currency market. Its price was $1,000 at the beginning of 2017 and reached till $14,000 in the same year. This 14-fold growth has helped bitcoin gain a lot of attention and speculation. The price also took a nosedive reaching a low of $7,000 and is struggling from then to soar high again.

These fluctuations are sporadic and driven primarily by the sentiments of the people around the world and not by institutionalized regulations. But as correlation is not causation, it could also be the other way around. These fluctuations can affect the perception of people for cryptocurrencies. Thus identifying the dependent and independent variable is crucial to see if the emotions are driving the prices if there is any correlation.

This analysis aims at first analyzing the prices of various cryptocurrencies separately which will help us in understanding if any cryptocurrency prices correlate. The sentiments will then be examined to see the correlation followed by the identification of dependent and independent variable in the relation.

GOALS AND OBJECTIVES

To analyze sentiments on cryptocurrencies, the variation in their prices, and determine whether the sentiments and the change in price correlate and Attempt to identify the independent and dependent variable between the two.

METHODOLOGY AND RESULTS

As the first step in the analysis, data of the daily price of bitcoin is procured with the help of web-scrapping through open source softwares from Quandl. Quandl is a platform that sources data from over 500 publishers delivering core financial and economic data. Quandl has price data from four sources: Kraken, Coinbase, Bitstamp, and Itbit. The mean of all the four prices is used to determine the Bitcoin(BTC) price for a particular day.

For altcoins (Nine of the top cryptocurrencies - Ethereum (ETH), Litecoin (LTC), Ripple (XRP), Ethereum Classic (ETC), Stellar (STR), Dash, Siacoin (SC), Monero (XMR), and NEM (XEM)), Poloniex is used as a data source. The data from Quadl and Poloniex is then merged to make a single data set. A logarithmic scale is used for prices so that all the cryptocurrencies fit in the single graph as the prices for Bitcoin are very high compared to the other cryptocurrencies. The variation of prices is graphically shown in figure 1.

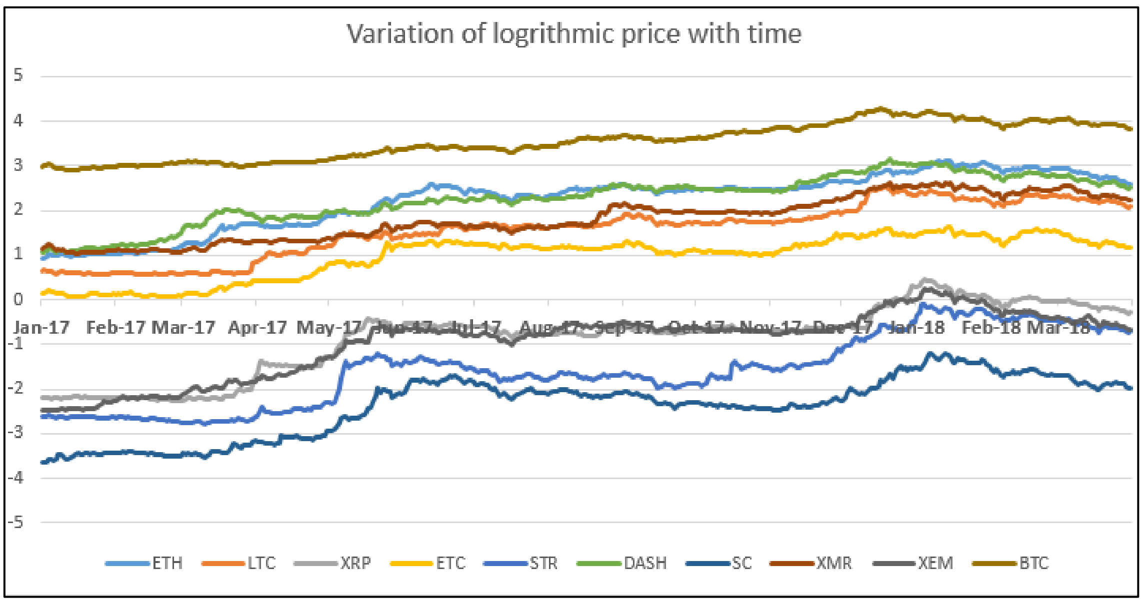

Figure 1: Variation of logarithmic price of cryptocurrencies with time

Tweets with hashtag of cryptocurrency are scraped with the help of python and some datasets are procured from Kaggle. The sentiments are considered for the period between April 2017 and February 2018. The dataset originally has 6.25 million records but for the purpose of analysis 500,000 records are randomly picked by simple random sampling using Base SAS.

Text analytics is done with the help of SAS Viya. SAS Viya gives a sentiment score(polarity) to each tweet between 0 to 1. The score between 0 to 0.4 is a negative tone, between 0.4 to 0.6 is a neutral tone and 0.6 to 1.0 is a positive tone. So the tweets are classified into tones based on these cut-offs. Total number of tweets and the number of tweets of each tone on a daily basis is shown in figure 2.

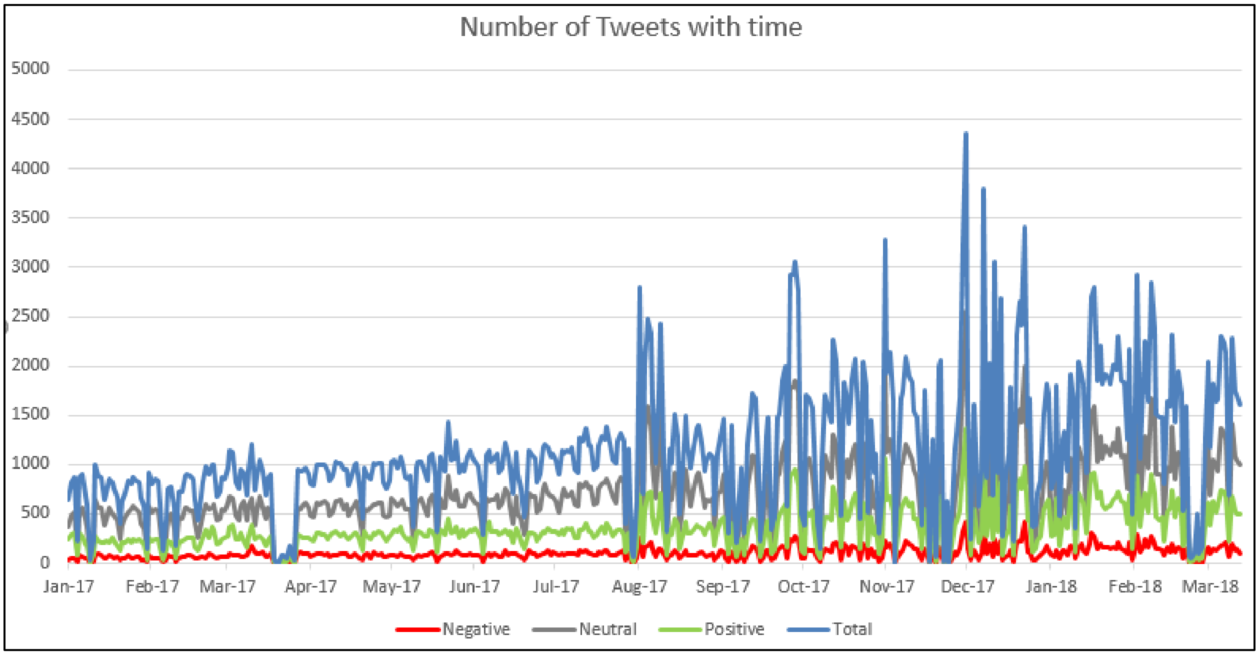

Figure 2: Variation of number of tweets of different tones with time

A pearson correlation matrix is prepared with price and the average polarity for the day as well as the week . Upon analyzing the matrix, it is found that:

1. The correlation between the daily polarity and the price on that day is week and positive.
2. The correlation between the weekly price and the polarity is moderate and positive.

The correlations are tabulated in figure 3:

Time period	Correlation	Strength
Daily	0.27	Weak
Weekly	0.40	Moderate

Figure 3: Weekly and daily correlation between price and polarity

Figure 3 clearly depicts that there is some correlation between price and sentiments. However, it is also important to determine the independent and the dependent variable between the two. For this purpose, the polarity and price are taken as two time-series to see if there is some cross-correlation between the two.

Looking at the cross-correlations plots (ccf plots) for the daily data in figure 4 and figure 5, it is observed that the price and polarity have high cross-correlation as the bars in the plots are above the tolerance limit indicated by the shaded portion along the x-axis. However, the plots are not helpful in determining the dependent variable as the ccf does not die out after a certain lag. It is to be noted here that only the positive lag is focused for the analysis as a negative lag indicates a lead and hence has been omitted from analysis.

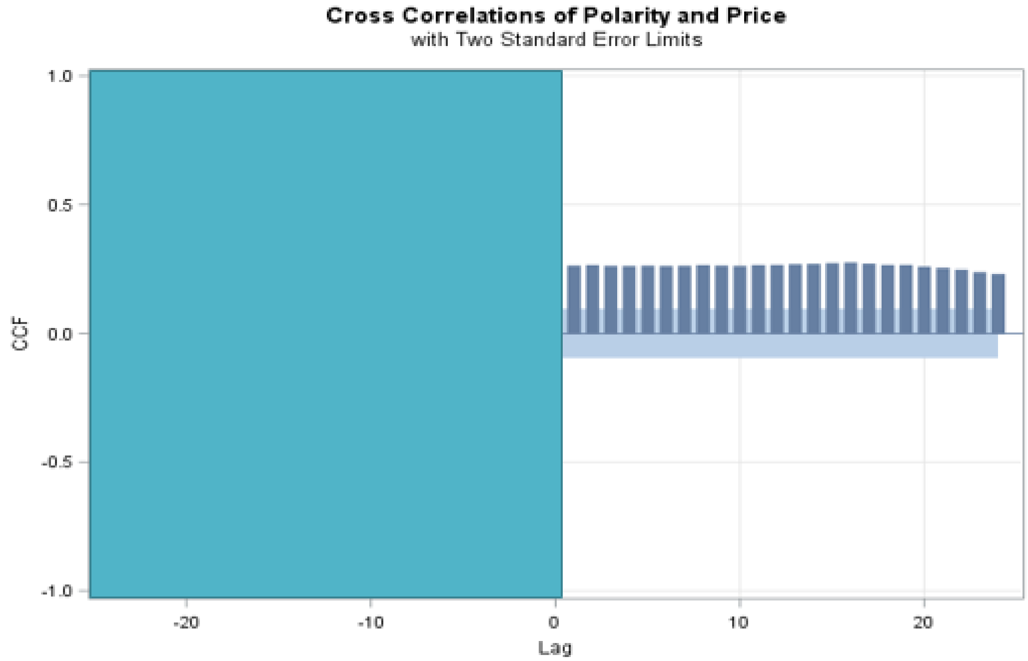

Figure 4: Cross-correlation plot taking daily polarity as the target variable

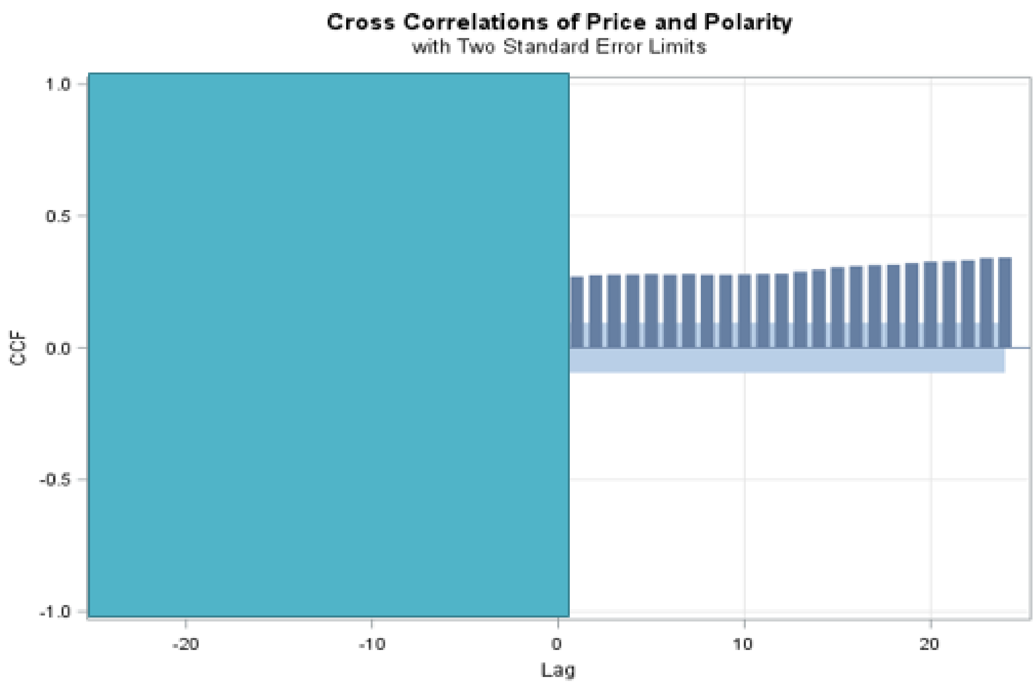

Figure 5: Cross-correlation plot taking daily price as the target variable

Looking at the ccf plot (figure 6) with price as target for the weekly data, it is observed that cross-correlation between the price and polarity can be deemed white noise as the bars in the plots are below the tolerance limit indicated by the shaded portion along the x-axis. The first lag cross-correlation is above the tolerance interval but still the bar is not high enough to be taken into account. Hence, the polarity is not determining the price.

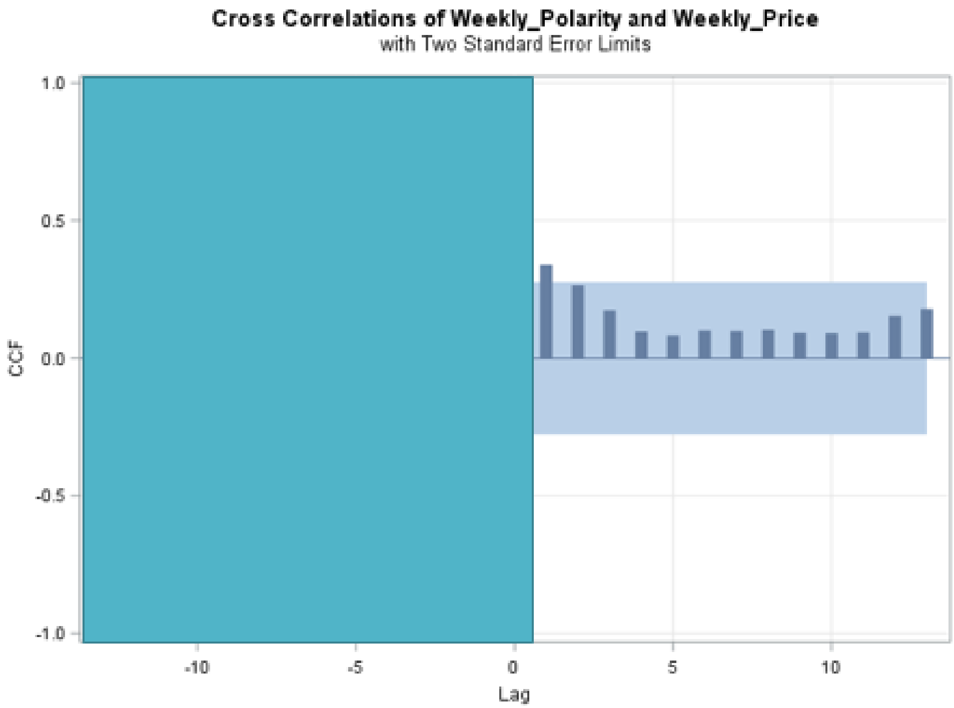

Figure 6: Cross-correlation plot taking weekly price as the target variable

Looking at the ccf plot (figure 7) with polarity as target for the weekly data, it is observed that cross-correlation between the price and polarity is significant as the bars in the plots are above the tolerance limit indicated by the shaded portion along the x-axis. Hence, the price for the past five weeks is determining the polarity for the current week. This is because the cross-correlation cuts-off after the fifth lag.

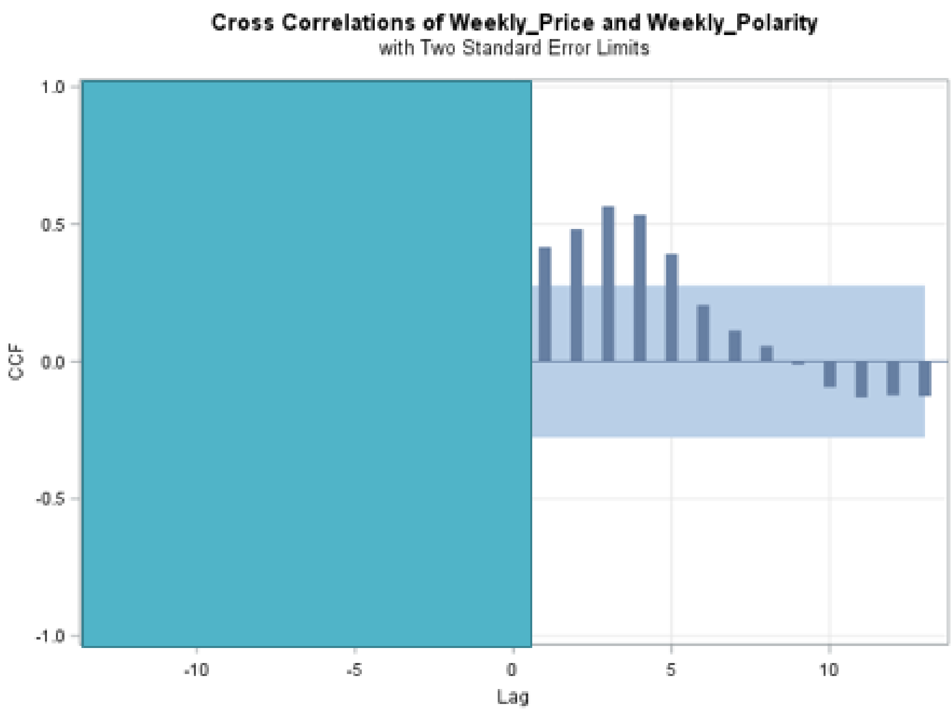

Figure 7: Cross-correlation plot taking weekly polarity as the target variable

CONCLUSION

This paper uses the power of SAS to analyze the sentiments in text and determines the factors that affect the sentiments and the factors that are affected by the sentiments. The insights gained from the analysis are:

1. Correlation is more pronounced in the weekly data and helps in determining the independent variable.
2. The price for the past 5 weeks is affecting public perception.
3. Factors other than people's opinion drives the prices.
4. Monthly aggregation with more data can provide further scope for analysis.

REFERENCES

[1] Rishanki Jain, Rosie Nguyen, Linyi Tang, and Travis Miller. 2018. "Bitcoin Price Forecasting using Web Search and Social Media Data" *Proceedings of the SAS Global 2018 Conference*, 3601-2018. Oklahoma State University, Stillwater. Available at https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/3601-2018.pdf

[2]Patrick Triest. 2017. "Analyzing cryptocurrency markets using python." Article available at https://blog.patricktriest.com/analyzing-cryptocurrencies-python/

[3] Kaggle. (2017-2018): https://www.kaggle.com/

[4] Twitter. (2017-2018): https://twitter.com/?lang=en

ACKNOWLEDGMENTS

I thank Dr. Goutam Chakraborty, SAS® Professor of Marketing Analytics and Dr.Miriam McGaugh, Clinical Professor at Oklahoma State University for their guidance and support throughout the research.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

1. Priyansh Dadar, Oklahoma State University, Stillwater, OK
   Email: priyansh.dadar@okstate.edu