Market Sentiment Helps Explain the Price of Bitcoin

Yoshiharu Sato
http://yoshi2233.strikingly.com/

December 15th, 2017

Abstract. A simple model that accurately explains the market price of Bitcoin (BTC) based on Metcalfe's law and market sentiment on social networks is presented. A previously proposed model has shown that the number of unique addresses and the estimated transaction value per address explain most of the variation of the BTC price. We demonstrate that the model's accuracy can be further increased by introducing a market sentiment volume about BTC that is positively scored on Twitter and StockTwits.

1. Introduction

The exponential rise in the price of Bitcoin (BTC) in the past several months has attracted a great number of investors and speculators hoping to make capital gains in the market. As of December 8th, 2017, the market capitalization of Bitcoin surpassed that of Visa Inc., one of the world's largest card payment network, at more than US$300 billion. Many of the market participants seem to be buying bitcoins in the pure hope that the price of a bitcoin continues to rise in the future thanks to the cryptocurrency's deflationary nature. Nakamoto [1], the original creator of Bitcoin, defined the rate of block creation (i.e., "mining") to be adjusted every 2016 blocks with a two-week adjustment period, with which the amount of bitcoins created per block decreases by 50% per 210,000 blocks. The number of bitcoins in existence is therefore expected not to exceed 21 million. [2]

Despite this clearly defined supply model of BTC, the intrinsic value of a bitcoin is hard to estimate, inasmuch as the cryptocurrency does not generate any interest nor dividend, making it impractical to employ the existing valuing methods (e.g., discounted cash flow; DCF). The price of a bitcoin is determined predominantly by the chaotic process of price discovery that is taking place on online electronic exchanges based on the continuous double auction (CDA) mechanism (which is typically implemented as a limit order book; LOB). Attempts have been made to quantitatively model the market price of a bitcoin rather than to estimate its fair value. Significantly, Fundstrat has shown that most of the variation of the BTC price can be explained using two blockchain stats: 1) the number of unique addresses; and 2) the estimated transaction value per address. [3] Their model is based on Metcalfe's law, which states that the value of a telecommunications network is proportional to the square of the number of connected nodes.

In this short paper, we demonstrate that the variation of the BTC price can be more accurately explained by additionally introducing a market sentiment volume about BTC that is positively scored on Twitter [4] and StockTwits [5]. We use sentiment data of the cryptocurrency market aggregated by Decryptz [6]. Our model shows an improved accuracy in both in-sample and out-of-sample regressions.

2. Metcalfe's Law

Robert Metcalfe, co-inventor of the Ethernet network technology, formulated in early 1980s that the value of a telecommunications network is proportional to the square of the number of connected nodes of the network. Given n nodes the total number of connections that can potentially be established in the network is the triangular number $n(n\text{-}1)/2$, which is asymptotically n^2. The value of the network therefore grows on the order of $O(n^2)$ while the cost of the network increases linearly with n. Metcalfe has later modified the law using $n \times \log(n)$ instead of n^2. [7]

In the context of the BTC blockchain network, each node for Metcalfe's law can be represented by a unique "address" – i.e., an identifier of 26-35 (usually 33) alphanumeric characters representing a destination for a bitcoin payment. However, this is a compromise since each BTC user can have multiple BTC addresses. In fact, by design it is an unintended practice to use the same BTC address for multiple transactions. [8] For this reason the number of unique addresses is not the precise representation of the total number of unique users, which is impossible to know precisely at least from all publicly available information.

Figure 1 below shows the number of unique BTC addresses in blue (data obtained from [9]). For denoising purposes, the n for Metcalfe's law is represented by the 5-day simple moving average of the number throughout this paper.

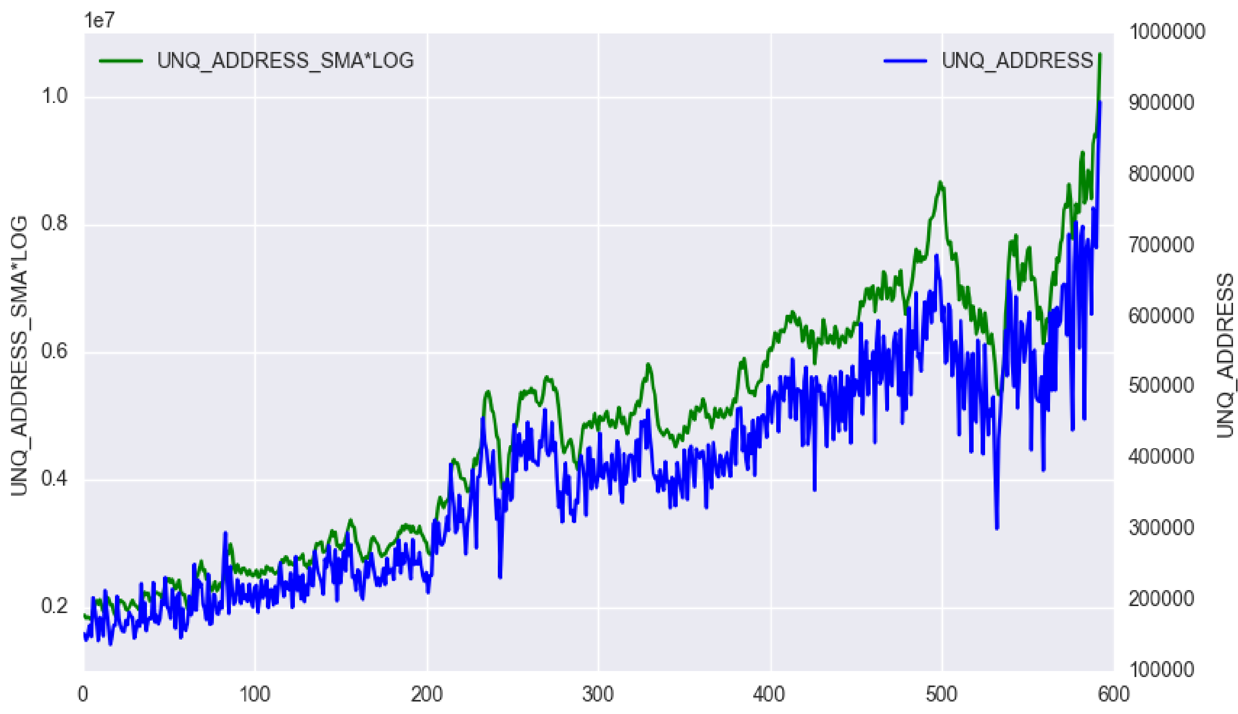

Figure 1: Number of unique BTC addresses (n) [9] and $\text{SMA}_5(n) \times \log(\text{SMA}_5(n))$ (from September 1st, 2014 to December 1st, 2017).

3. Market Sentiment

Since BTC does not make any cash flows (i.e., it does not accrue interests nor pay dividends), the price of BTC is determined predominantly by the chaotic process of price discovery that is taking place on online electronic exchanges, driven continuously 24/7 by the dynamic demand-supply balance of buyers and sellers. Such balance is considered to be significantly affected by the sentiment of investors and speculators participating in the market. We hypothesize that such market sentiment, if properly quantified, can help explain the BTC price.

In order to measure the sentiment quantitatively, we use a sentiment dataset aggregated by Decryptz [6] from Twitter [4] and StockTwits [5]. Their natural language processing (NLP) engine ingests those social networks' data streams everyday with a filter identifying and collecting the volume of conversations for each cryptocurrency based on cashtags ("$"), hashtags ("#") and keywords. The engine subsequently analyzes the language in each conversation to determine the sentiment as well as its intensity level. The engine assigns an intensity score to each conversation in a range of -4 (extreme negative sentiment) to +4 (extreme positive sentiment). Figure 2 below shows the daily BTC price and the positively-scored market sentiment volume about BTC collected by the engine. The high correlation between the two can be immediately observed.

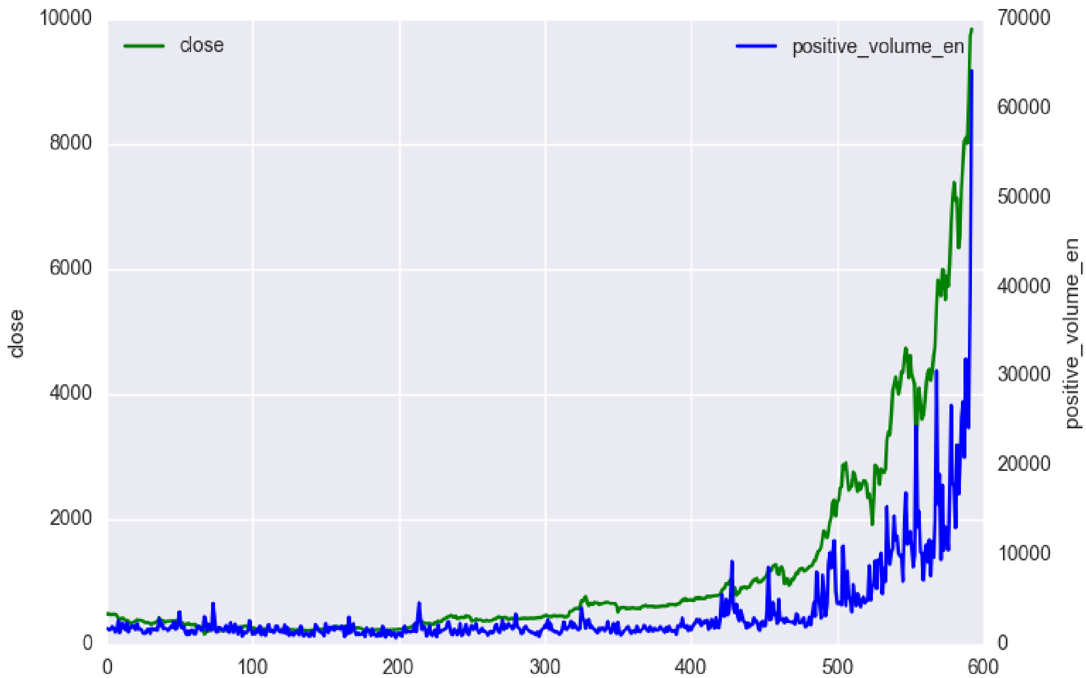

Figure 2: Daily BTC price and positively-scored market sentiment volume about BTC on Twitter and StockTwits (from September 1st, 2014 to December 1st, 2017).

4. Models & Results

We build two ordinary least squares (OLS) models to regress the price of BTC. The first model uses 1) $SMA_5(n) \times \log(SMA_5(n))$, where n is the number of unique BTC addresses obtained from [9], and 2) $SMA_5(v)$ divided by $SMA_5(t)$, where v is the estimated BTC transaction value in US dollar [10] and t is the number of confirmed BTC transactions per day [11]. The second model uses the positively-scored market sentiment volume about BTC on Twitter and StockTwits aggregated by [6] in addition to the aforementioned two variables. Fundstrat [3] uses n^2 instead of $SMA_5(n) \times \log(SMA_5(n))$, and $SMA_5(v)$ divided by $SMA_5(n)$ instead of $SMA_5(v)$ divided by $SMA_5(t)$. However, we find that the latter in both cases produce a better result. This can be attributed to the fact that $n \times \log(n)$ better explains Metcalfe's law than n^2 [7] and that there are a large number of unused BTC addresses; v therefore better captures the BTC network activity than n.

We use observations from September 1st, 2014 to September 19th, 2017 as in-sample (IS) period, and observations from September 20th, 2017 to December 1st, 2017 as out-of-sample (OOS) period. In either period, no outliers have been removed unlike the model of [3]. Table 1 below shows the regression results from the IS and OOS periods. The model with the sentiment variable exhibits better results in all the accuracy criteria.

	In-Sample			Out-of-Sample	
	Adj. R Square	Schwarz IC	Akaike IC	RMSE	MAE
2-Var OLS Model	0.919	7,780	7,767	1,839	1,737
3-Var OLS Model	0.929	7,711	7,694	1,460	1,351

Table 1: Regression results of the two models in the in-sample and out-of-sample periods.

Figure 3 and 4 respectively show the regression charts of the 2-var and 3-var OLS models. Actual daily BTC price is shown in the dashed red line, while the IS regression in blue and the OOS regression in green. It is visually recognizable that the model with the sentiment variable performed better in the OOS period. Figure 5 and 6 show the quantile-quantile probability plot of the residuals of the two models. They both indicate the presence of extreme values in the residuals and hence the distribution is heavy tailed (i.e., the residuals are not normally distributed).

The estimate of an OLS regression model is still a reasonable estimator with the presence of non-normal residuals. However, the Gauss-Markov theorem requires that, in order for the OLS estimate to be the best linear unbiased estimator (BLUE), the residuals must be uncorrelated. The correlograms of the two models shown in Figure 7 and 8 indicate that the residuals of both models are actually autocorrelated, which violates the Gauss-Markov theorem's assumptions, suggesting the estimates are inefficient. Removing the autocorrelation in the residuals is outside the scope of this paper; however, we believe that the 3-var OLS model accurately explains the price of BTC despite the correlated residuals (hence the inefficient estimate).

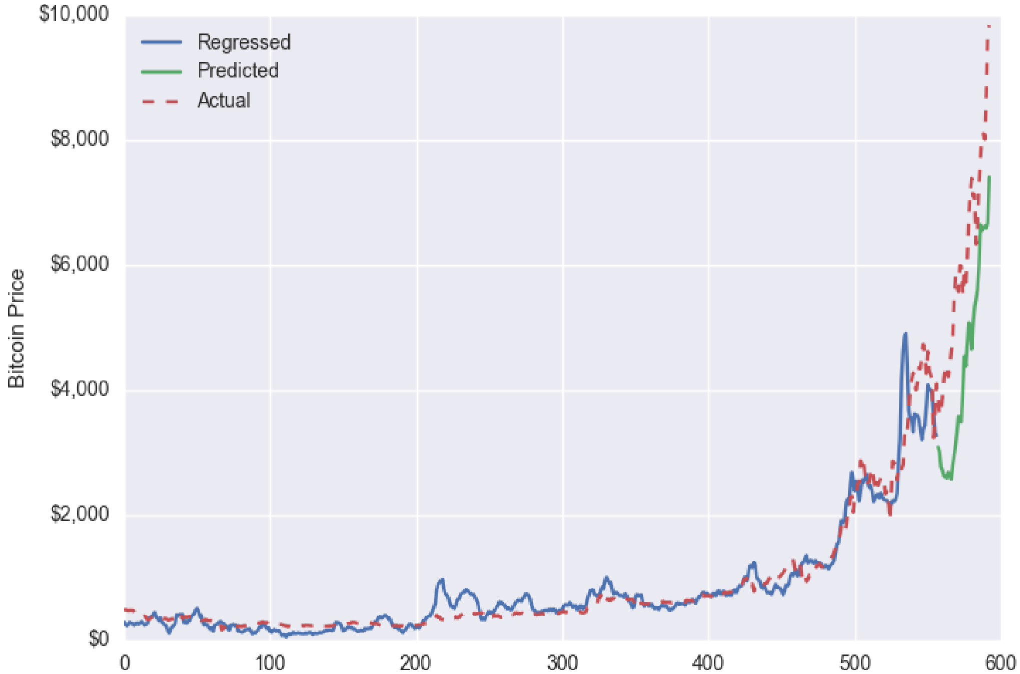

Figure 3: In-sample and out-of-sample regression result of the 2-var OLS model.

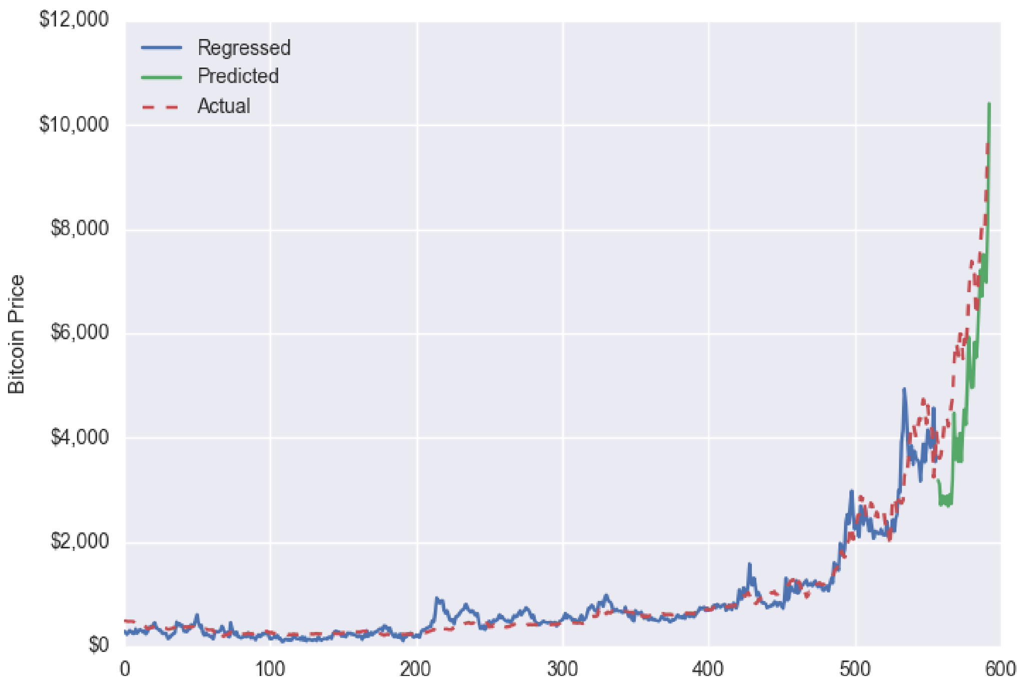

Figure 4: In-sample and out-of-sample regression result of the 3-var OLS model.

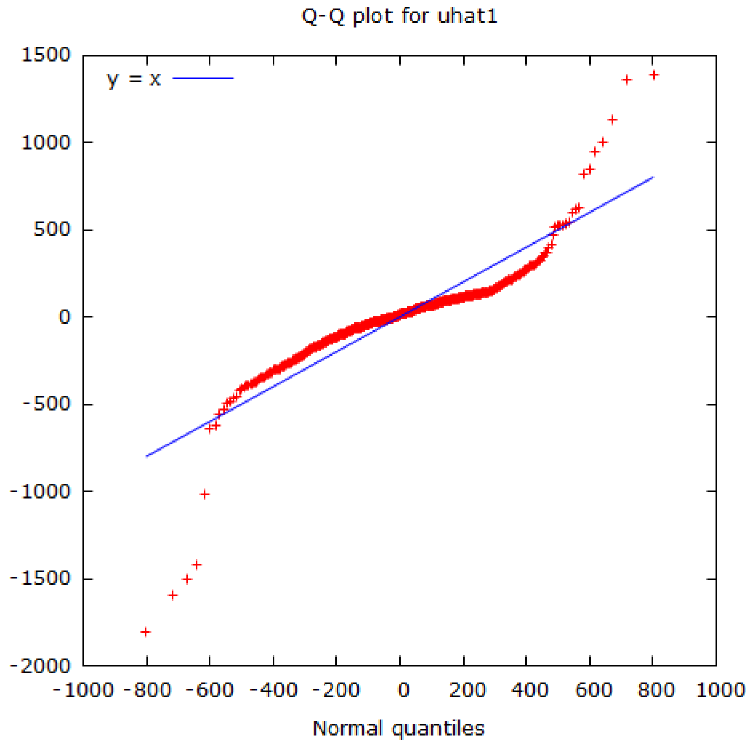

Figure 5: Q-Q probability plot of the residuals in the 2-var OLS model.

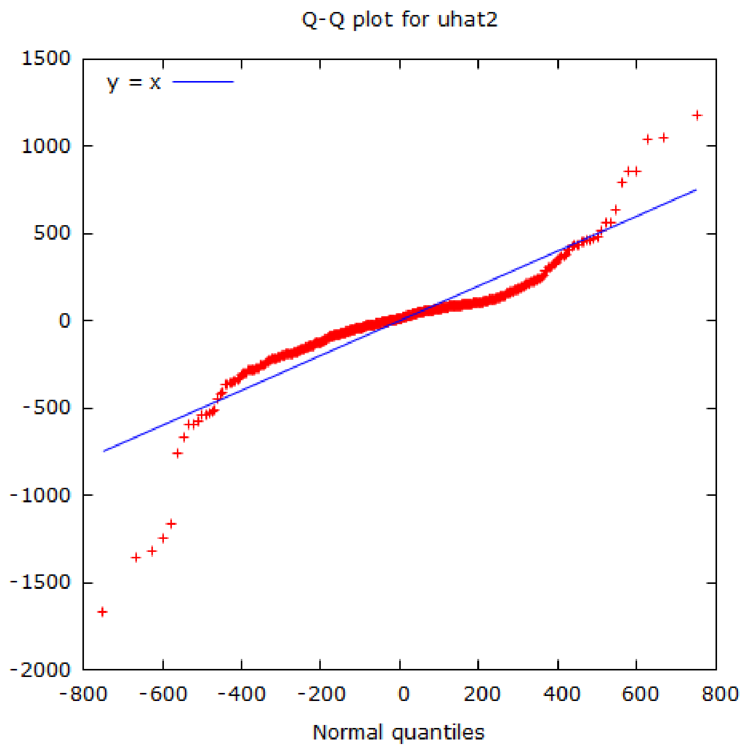

Figure 6: Q-Q probability plot of the residuals in the 3-var OLS model.

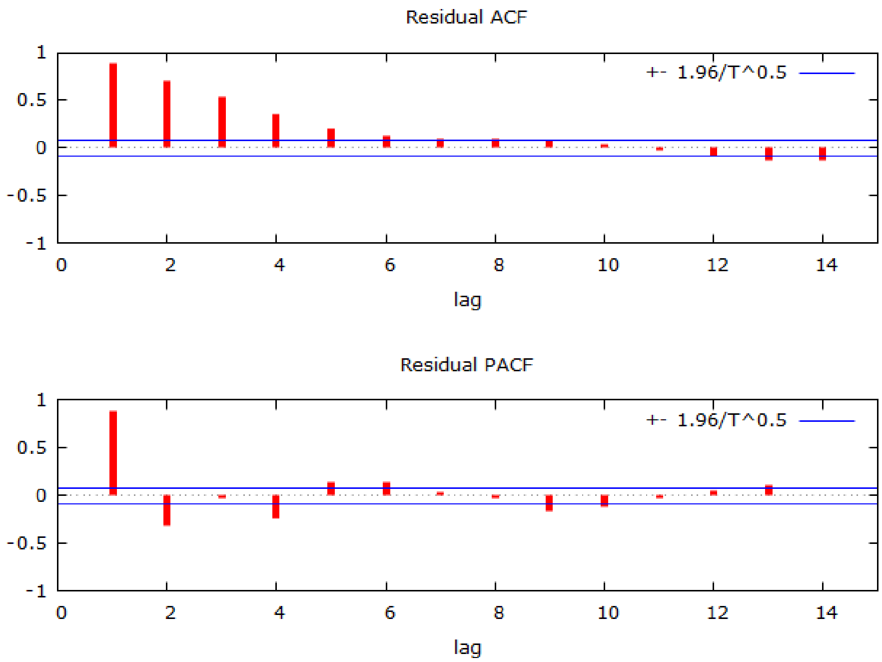

Figure 7: Correlogram of the 2-var OLS model.

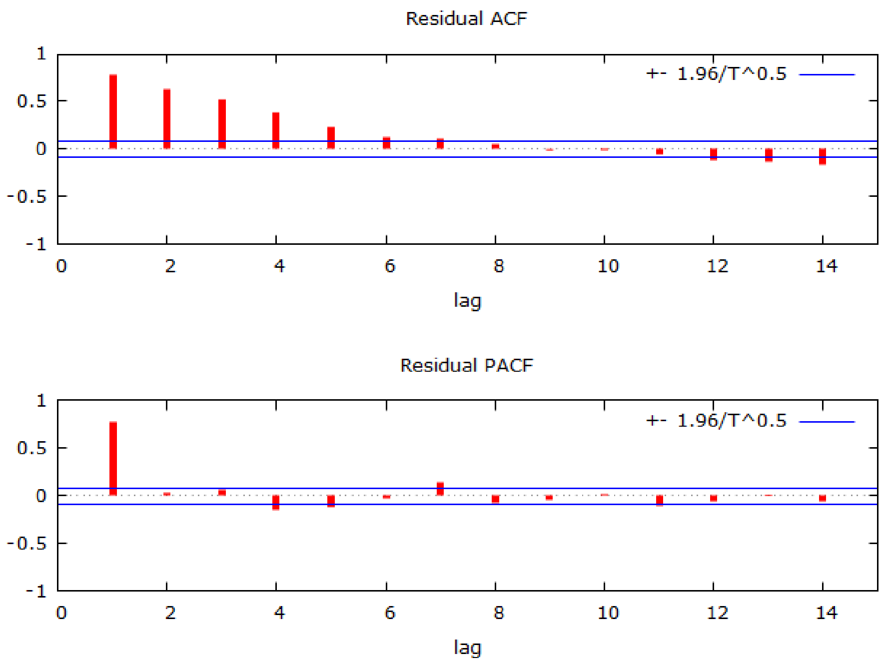

Figure 8: Correlogram of the 3-var OLS model.

Conclusion

A simple model that accurately explains the price of BTC was presented. Specifically, we demonstrated that by combining the concept of Metcalfe's law the variation of the BTC price can be accurately explained by using a market sentiment volume about BTC that is positively scored on social networks. The assumption is that the price of BTC is predominantly a result of the demand-supply balance of buyers and sellers, which is greatly influenced by market sentiment. It is likely that market sentiment is also highly affected by the price itself, forming a circular and bidirectional reflexivity. In such a relationship, cause and effect influence one another and together create a feedback loop, wherein neither can be clearly assigned as sole cause or sole effect. As a result, asset values are not driven only by the economic fundamentals but often by biased assumptions and actions of participants spreading across the market. This leads to markets having boom-and-bust disequilibrium, which the price of BTC seems to be experiencing.

References

[1] Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008.

[2] BitcoinWiki, "Controlled supply." https://en.bitcoin.it/wiki/Controlled_supply

[3] Business Insider, "Analyst says 94% of bitcoin's price movement over the past 4 years can be explained by one equation." http://www.businessinsider.com/bitcoin-price-movement-explained-by-one-equation-fundstrat-tom-lee-metcalf-law-network-effect-2017-10

[4] Twitter. https://twitter.com/

[5] StockTwits. https://stocktwits.com/

[6] Decryptz. https://decryptz.com/

[7] Hirshland, "Guest Blogger Bob Metcalfe: Metcalfe's Law Recurses Down the Long Tail of Social Networks." https://vcmike.wordpress.com/2006/08/18/metcalfe-social-networks/

[8] BitcoinWiki, "Address reuse." https://en.bitcoin.it/wiki/Address_reuse

[9] Blockchain.info, "Number Of Unique Addresses Used." https://blockchain.info/charts/n-unique-addresses

[10] Blockchain.info, "Estimated USD Transaction Value." https://blockchain.info/charts/estimated-transaction-volume-usd

[11] Blockchain.info, "Confirmed Transactions Per Day." https://blockchain.info/charts/n-transactions

Appendix A

Regression details of the 2-var OLS model in the in-sample period (09/01/2014–09/20/2017).

Dep. Variable:	y	R-squared:	0.920
Model:	OLS	Adj. R-squared:	0.919
Method:	Least Squares	F-statistic:	3173.
Date:	Thu, 14 Dec 2017	Prob (F-statistic):	3.97e-304
Time:	11:50:41	Log-Likelihood:	-3880.5
No. Observations:	557	AIC:	7767.
Df Residuals:	554	BIC:	7780.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-509.1976	30.406	-16.747	0.000	-568.922	-449.473
x1	0.0001	7.39e-06	14.268	0.000	9.09e-05	0.000
x2	0.8462	0.015	56.891	0.000	0.817	0.875

Omnibus:	183.775	Durbin-Watson:	0.216
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4735.097
Skew:	-0.843	Prob(JB):	0.00
Kurtosis:	17.184	Cond. No.	1.37e+07

Appendix B

Regression details of the 3-var OLS model in the in-sample period (09/01/2014–09/20/2017).

Dep. Variable:	y	R-squared:	0.930
Model:	OLS	Adj. R-squared:	0.929
Method:	Least Squares	F-statistic:	2442.
Date:	Thu, 14 Dec 2017	Prob (F-statistic):	1.84e-318
Time:	12:07:50	Log-Likelihood:	-3843.1
No. Observations:	557	AIC:	7694.
Df Residuals:	553	BIC:	7711.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-495.2378	28.498	-17.378	0.000	-551.216	-439.260
x1	9.749e-05	6.97e-06	13.990	0.000	8.38e-05	0.000
x2	0.6772	0.024	28.794	0.000	0.631	0.723
x3	0.0689	0.008	8.919	0.000	0.054	0.084

Omnibus:	192.000	Durbin-Watson:	0.427
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3262.986
Skew:	-1.045	Prob(JB):	0.00
Kurtosis:	14.672	Cond. No.	1.37e+07