What Goes Up……: modelling the Bitcoin rollercoaster ride

Dr Yang Li
Department of Computer Science and Digital Technologies
University of East London
London E16 2RD, UK
y.li@uel.ac.uk

***Abstract*— Cryptocurrencies have attracted increasing attention worldwide. Cryptocurrency assets are likely to remain a viable choice for the public in long term. In this paper, the modelling of cryptocurrency price is explored in Bitcoin bubbles prior to and during the COVID-19 pandemic. As shown here, it is necessary and possible to understand the dynamics in plausible proxy variables. A similar methodology could be deployed in other situations where market bubbles occur.**

Keywords — Cryptocurrency, Bitcoin, Dynamic, Modelling, Proxy Variable, Time Series

I. INTRODUCTION

Bitcoin has been described as "arguably the greatest bubble of recent years" [1]. The time series of Bitcoin price in Figure 1 shows a series of bubbles with earlier ones prior to 2017 being dwarfed by those in the last few years. A period of exponential growth in price reaches delusional levels and the bubble bursts as the population of susceptible buyers is exhausted. In a similar way, Isaac Newton lost what was then a fortune in the infamous South Sea Bubble of 1720 leading him to comment "I can calculate the motion of heavenly bodies but not the madness of people". The largest Bitcoin bubble yet has been brewing during the COVID-19 pandemic.

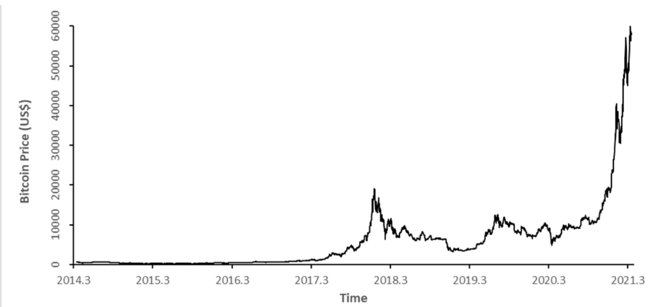

Fig. 1. Price of Bitcoin in US$ 2014 – 2021. (*data source: coindesk.com/price/bitcoin*)

Since being unveiled in 2009, cryptocurrencies (such as Bitcoin) have attracted increasing attention worldwide. Coins or assets are autonomously self-propagated through cryptocurrency mining. All transaction records are distributed through blockchain. Blockchain is a decentralized digital ledger, which is online, open to the public and resistant to modification. Apparently, cryptocurrency assets remain a viable choice for the public, even with such volatility. The power of digital currencies, which has been officially discussed by IMF and central banks, could reshape global finance. There is great potential for cryptocurrency as smart money. The economic and social impacts could be far reaching; hence the authorities will need to see if new legislation is required.

Investors and businesses are eager to understand the volatility of price in order to evaluate the risks and capture the opportunities [2][3]. The potential to model cryptocurrency behaviours has been explored prior to the first massive bubble in 2017 [4] and continue efforts are made [5]. It is complex to forecast the price trend of a cryptocurrency, because they lack intrinsic value and are thus far unregulated. The trading infrastructure is still not mature. Therefore, there are a wide range of factors that could influence cryptocurrency price, such as market confidence and trading pattern. These are mostly intangibles and would need to be measured by proxy variables.

In the following sections, the modelling of cryptocurrency price is explored in both bubbles of 2017/18 and 2020/21, that is, prior to and during the COVID-19 pandemic.

II. MARKET INTANGIBLES

To illustrate the use of market intangibles in modelling Bitcoin price, we use Google Trends. Its Search Volume Index represents the Google search popularity by keywords across various regions and languages over a specified time period. When 'Bitcoin' is the search keyword, Google Trends is clearly able to capture the 2017/18 bubble but less so the relative magnitude of the current bubble (Figure 2).

The Google Flu Trends tool [6] was a model developed for predicting flu epidemics. However, the dynamic of Google search keywords, the influence of ranking algorithms and metrics eventually led to the abandonment of this initially successful predictive service after a few years. Fong studied the spurious correlations between Google search keywords and stock market trends [7]. Such correlations often misguide stock market prediction. As real correlations could become spurious correlations, the flaw may not be the techniques used in Google Trends, but rather the underestimation of the dynamic of keyword searches on Google.

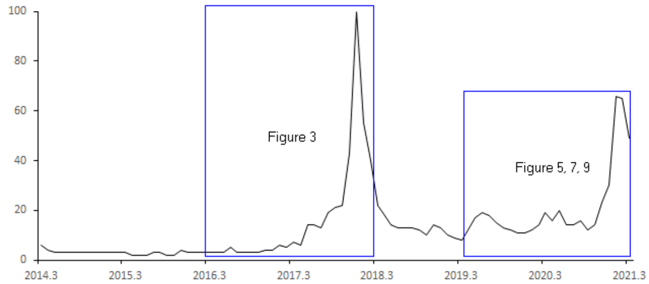

Fig. 2. Google Trends for 'Bitcoin' during 2014 – 20121. (*data source: trends.google.com*)

III. Capturing the Variables

Daily Bitcoin prices are extracted through R package coindeskr which accesses the 'CoinDesk' Bitcoin Price Index API. Daily data of Bitcoin transactions were downloaded from blockchain.com available in CSV format every other day. Major transaction variables retrieved are as follows:

- A Total number of Bitcoins in circulation,
- Trading volume on bitcoin exchanges (USD),
- Count of unique trading addresses on the blockchain,
- Number of transactions on the blockchain,
- Count of Blockchain wallets,
- Miner's revenue (USD)

Daily Google Trends data were retrieved by running R package transform for the keywords Bitcoin, blockchain and cryptocurrency.

News headlines have been acquired from coindesk.com by employing web scraping techniques using R (cran.r-project.org) and Selenium WebDriver (selenium.dev). Sentiment analysis was carried out on these headlines using lexicons of positive and negative words after preparation as a corpus using the R package tm. Headlines were classified as positive, neutral or negative.

IV. Modelling Bitcoin Price for the 2017/18 Bubble

Daily Bitcoin

This analysis focuses on the 2017/18 bubble. Correlations were tested between the Bitcoin price and the variables discussed above from March 2016 to March 2017 when the market started to change. Three variables that are well correlated with the Bitcoin price are identified as the most pertinent proxy variables, which are 'trading volume', 'miner's revenue' and Google Trends of 'Bitcoin' (see Table 1). The proxy variables are then used in modelling from March 2017 to March 2018 (a volatile period for the market). A time series model of AutoRegressive Integrated Moving Average (ARIMA) has been used to see how effective the model could be with proxy variables over the period forward. ARIMA is a class of models with the identified proxy variables as regressors to fit a time series based on its own lagged values, its own lagged errors, and the elimination of non-stationarity.

TABLE I. Correlation between candidate proxy variables and Bitcoin price during 2017/18

Candidate Proxy Variable	Correlation Coefficient
Total number of Bitcoins in circulation	0.614
Trading volume on bitcoin exchanges (USD)	0.828
Count of unique trading addresses	0.727
Number of transactions	0.501
Count of Blockchain wallets	0.819
Miner's revenue (USD)	0.983
Google Trends for 'Bitcoin'	0.936
Total number of news headlines	0.488

The result of the modelling is given in Figure 3. The modelled Bitcoin price change corresponds well to the real trend of the Bitcoin price during the time-period of March 2017 to March 2018 except at the highest peak prices. The black and red lines show the real trend of the Bitcoin price, while the blue line is the modelled trend. The darker blue shaded area shown in the figure is at the 95% confident level, and the lighter shaded area is at the 80% confident level.

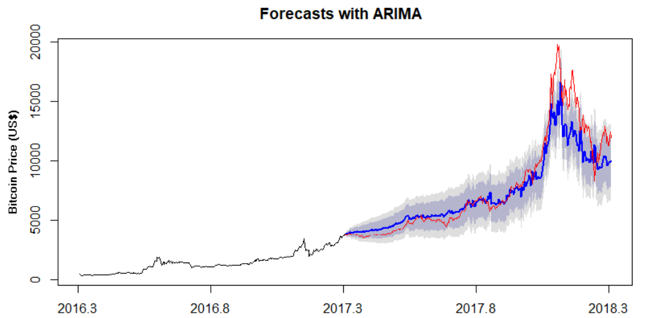

Fig. 3. The modelled Bitcoin price trend during 2016 -2018 by using ARIMA (*blue line with confidence intervals*), and the real price trend (*red line*).

V. Modelling Bitcoin Price for the 2020/21 Covid Age

2020 was a dramatic year for cryptocurrency market and the drama has been extended to 2021. In early 2020, the market was in turbulence because of the uncertainty and opportunism raised by Covid-19 [8]. From the middle of the year, a surge in Bitcoin price started. Covid-19 seems to have heightened the need for digital financial services and consequently has a positive impact on the cryptocurrency market efficiency. Three new models are constructed using 2020/21 data, despite the market still undergoing change due to ongoing Covid-19 pandemic.

The trial modelling for 2020/21 data is carried out in this study following the same principle, which has been applied for 2017/18 data earlier in this paper. Ten candidate proxy variables are prepared as the predictors in the modelling, which are listed as follows. Correlations between the Bitcoin price and the candidate variables were volatilising during this period.

TABLE II. CORRELATION BETWEEN CANDIDATE PROXY VARIABLES AND BITCOIN PRICE DURING 2020/21

Candidate Proxy Variable	Correlation Coefficient
Trading volume on bitcoin exchanges (USD)	0.732
Count of unique trading addresses	0.647
Number of transactions	0.127
Count of Blockchain wallets	0.926
Miner's revenue (USD)	0.978
Google Trends for 'covid'	0.291
Google Trends for 'Bitcoin'	0.893
Google Trends for 'cryptocurrency'	0.949
Google Trends for 'blockchain'	0.934
Death number of Covid-19 worldwide	0.706

XGBoost (Extreme Gradient Boosting framework) is used to identify the feature importance. XGBoost is an ensemble machine learning technique to boost the classifiers. The Bitcoin price trend is then modelled by ARIMA with the selected variables. As the worldwide outbreak of pandemic started in March 2020, the modelling is carried out from April 2020. In the three months period of April to June, 'trading volume', 'wallets' and Google Trends of 'Covid' are identified by XGBoost as relative more important to the Bitcoin price (Figure 4). However, the inconsistent correlations with Bitcoin price might reflect the opportunistic trading and lack of confidence in the market because of the initial Covid-19 outbreak (correlation coefficient is 0.19 for 'trading volume', 0.86 for 'wallets', -0.88 for Google Trends of 'Covid'). The Bitcoin price is then forecasted by ARIMA with these three variables in the next three months period of July to September (Figure 5). The RMSE is 813 and the MAE is 733 for the modelled prices.

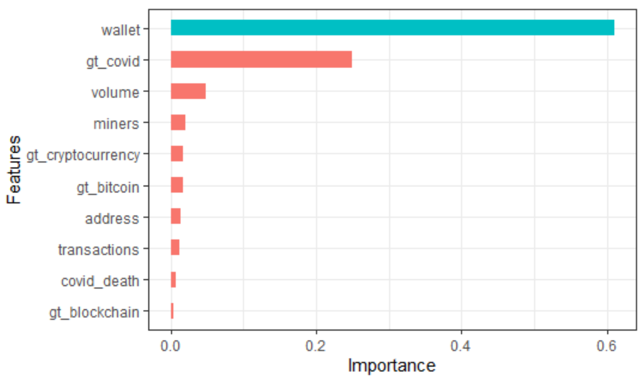

Fig. 4. The feature importance during April to June 2020 by using XGBoost.

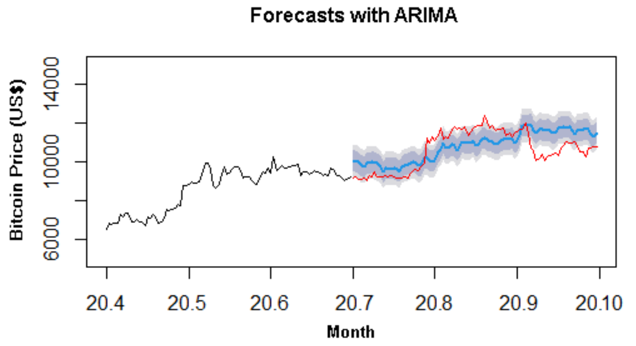

Fig. 5. The modelled Bitcoin price trend during July to September 2020 by using ARIMA, and the real price.

This procedure is repeated from July where the market kept changing abruptly. From July to September, 'miner's revenue', 'wallets' and Google Trends of 'Cryptocurrency' are identified as relative more important (Figure 6). The correlations turn consistent, when the cryptocurrency is regarded as an alternative investment during pandemic (correlation coefficient is 0.77 for 'miner's revenue', 0.52 for 'wallets', 0.66 for Google Trends of 'Cryptocurrency'). Accordingly, the Bitcoin price is forecasted with the above three variables from October to December. The RMSE is 2368 and the MAE is 1990 for the modelled prices.

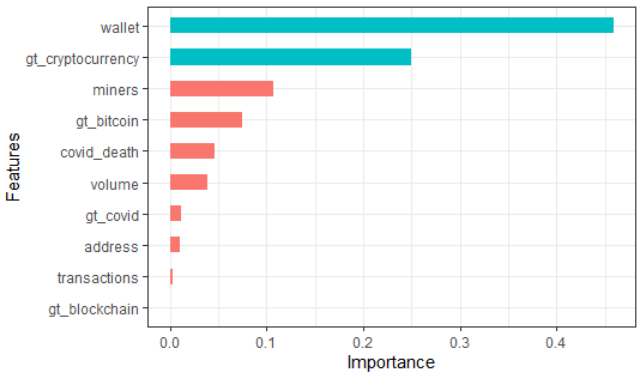

Fig. 6. The feature importance during July to September 2020 by using XGBoost.

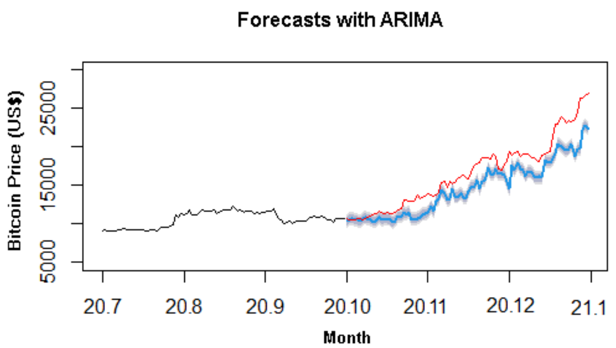

Fig. 7. The modelled Bitcoin price trend during October to December 2020 by using ARIMA, and the real price.

From October to December, 'miner's revenue', Google Trends of 'Bitcoin' and 'Covid death' are identified as relative more important by XGBoost (Figure 8). The correlations are strong, which shows the bubble scenario (correlation coefficient is 0.94 for 'miner's revenue', 0.94 for Google

Trends of ‘Bitcoin’, 0.88 for ‘Covid death’). The Bitcoin price is forecasted with the above three variables from January to February (Figure 9). The RMSE is 8362 and the MAE is 6614 for the modelled prices.

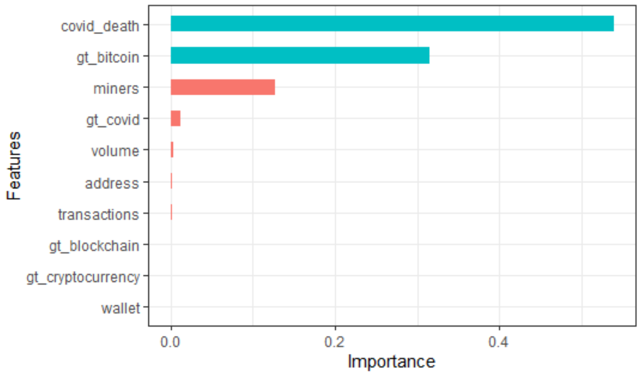

Fig. 8. The feature importance during October to December 2020 by using XGBoost.

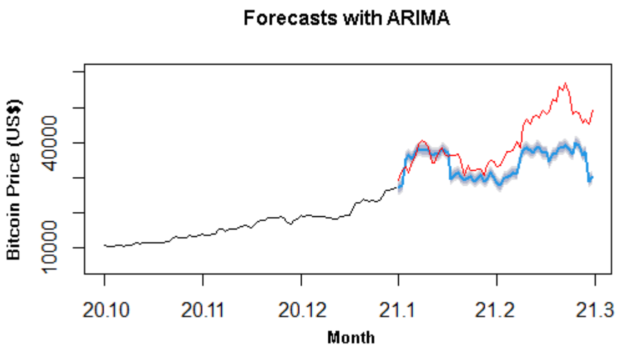

Fig. 9. The modelled Bitcoin price trend during January to February 2021 by using ARIMA, and the real price.

The result in Figures 5 and 7 show the modelled price trend has a reasonably good match with the real price trend. In Figure 9 however, as in Figure 3, the modelling underestimates peak prices in the bubble. Shorter-term modelling is more able to stay on track, forecasting further ahead is difficult. What is evident from this study is that as Bitcoin bubble grew during the 2020-21 Pandemic, the relative importance of the proxy variables changed. Thus early in the bubble technical aspects such as ‘count of blockchain wallets’ seems to drive sentiment. As the pandemic progresses, the variable ‘count of Covid-19 deaths worldwide’ rises in importance until it tops variable importance.

VI. Conclusion

The cryptocurrency market is complex and its currencies lack intrinsic values. Since 2017 the Bitcoin market has seen turbulence and bubbles that exponentially rise and then crash. While many phenomena in cryptocurrency market still needs to be explained and understood, as shown here it is possible to develop data-driven models of cryptocurrency price trend, at least in short term, but also to understand the dynamics in plausible proxy variables to see how drivers of sentiment in the market can change even in the short term in fuelling a bubble. When there is a change in state, such as the current pandemic, then it is clear that the choice of proxy variables needs to be adjusted. A similar methodology could be deployed in other situations where market bubbles occur.

References

[1] A. Kucharski. “The Rules of Contagion”, Profile Books, London, 2020.

[2] A. Kamsky. “A Comprehensive Guide To Why Is Bitcoin Volatile?”, CNN, Available at https://www.ccn.com/education/why-is-bitcoin-volatile/ 2023.

[3] J. Wood. “Putting Crypto Volatility in Context: What We Can Learn From the History of Bitcoin Crashes”, CoinDesk, Available at https://www.coindesk.com/learn/putting-crypto-volatility-in-context-what-we-can-learn-from-the-history-of-bitcoin-crashes/ 2023.

[4] M. J. Amjad and D. Shah. “Trading Bitcoin and Online Time Series Prediction”, Proceedings of Machine Learning Research, 55: 1-15, 2016.

[5] L. Bergsli, A. F. Lind, P. Molnar and M. Palasik. “Forecasting volatility of Bitcoin”, Research in International Business and Finance, Vol. 59, 2022, 101540.

[6] D. Lazer, R. Kennedy, G. King and A. Vespignani. “The Parable of Google Flu: Traps in Big Data Analysis”, Science, Vol. 343, Issue 6176, 1203-1205, 2014.

[7] W. M. Fong. “Bursting the big data bubble”, Significance, June 2017, 20-23.

[8] E. Demir, M. H. Bilgin, G. Karabulut et al. “The relationship between cryptocurrencies and COVID-19 pandemic”, Eurasian Econ Rev, 10, 349–360, 2020.