Journal of Advanced Research in Technology and Management Sciences
Volume: 05 Issue: 03 ISSN: 2582-3078 March 2023
Available online at: http://www.jartms.org

Smart Contract Digital Cryptocurrency Analysis of Ethereum Blockchain Transactions for Similar-web

Dhivya K[1], Dr. T. Sivakumar[2]
[1]*M. Tech,* [2]*Assistant Professor,*
Department of Computer Science,
Pondicherry University, Pondicherry, India.
[1]dhivyaakoramurthy@gmail.com
[2]tsivakumar72@gmail.com

***Abstract*— The Ethereum blockchain digital currency transactions including smart contracts are first examined using a large data analysis, which reveals intriguing features in action. The public's interest in digital currencies such as Ether, Finney, Wei and Bitcoin, as determined by the number of web searches, is then compared with Similar-web. Analysis of cryptocurrency pricing and search trends indicates the presence of major actors (and not just everyday users) who influence the marketplace when prices are lower. The pairs that provide better accurate and current estimates of Ether prices are revealed through with a cross-correlation analysis of digital cryptocurrency values and searching trends.**

***Keywords*— Ethereum, Blockchain, Similar-Web, Cryptocurrency, Smart Contract, Big Data.**

I. Introduction

Many industries and scholars from numerous disciplines are paying a growing amount of attention to blockchain as well as cryptocurrency like Ether, Finney, Wei and Bitcoin. The Blockchain's intricacy and underpinning interrelatedness, which have an impact on and draw impact from those other complicated systems like banking and social networks, are part of what attracts physicist and data analysts to it.

It makes sense to take online platforms into account whenever examining blockchains because of their decentralised nature. Decentralization was initially implemented to eliminate heterogeneity amongst systems. This would result in central banks as well as government agencies losing their control over the financial markets in the case of cryptocurrencies. Also, this would imply that the blockchain modifications are now managed by the system itself (users online). It goes without saying that the degree of decentralisation in a network determines its effectiveness. In other words, if "central" nodes are hugely disproportionate balanced, the channel is vulnerable to severe homogeneity threats [1]. A few studies have looked into the degree of central importance in decentralised network systems, and they have revealed that those who may not be as decentralised as one might assume [2].

II. Background Work

S. Haber & W. S. Stornetta proposed the first blockchain concept in 1990. They suggested an original, computation time workable method for timestamping digital information, which renders it impossible to change the date stamp after it was created [3]. The most-big consequences of this methodology is that no third-party interference would be obligated to maintain track of the date stamps. Because of its dependability, immutability, as well as transparency, blockchain technology has revolutionised numerous fields, including healthcare [4,5], public transport [6,7], digital forensics [8,9], and information security [9,10]. These features are a direct outcome of the blockchain framework: information is separated into blocks which are all linked together using cryptographic techniques. This framework precludes tampering with just about any arbitrarily defined block without affecting the others, resulting in immutability. Moreover, information kept in any node throughout the system is noticeable to all users, ensuring transparency. The decentralised data processing also prevents the two or more parties in a transfer of funds from manipulating stored data in the system in the past. In a nut - shell, the blockchain creates a smart contract that confirms the information of a money transfer without requiring the involved parties to trust one another [11].

A. Blockchain: Ethereum

The Ethereum blockchain is an instance of distributed ledger technology (DLT), which is a broad term for datasets that maintain and transmit information across a system of blocks of information. While most concepts employed in the sense of Ethereum really aren't unique towards this block chain technology, a few of them are. This section gives an overview of the specific terms employed in this paper which are connected towards the Ethereum blockchain. The Ethereum policy document includes an in-depth description of phrases and ideas (ethereum.org).

1) Gas: The price of making a transaction or computation smart contracts here on Ethereum network is measured in gas. Gas has its own changeable price, which miners calculate depending mostly on intricacy and computing power needed to process every block. The native cryptocurrency of Ethereum is called ether (ETH), and it is used to estimate gas fees. Wei is the nickname of ether's smallest number (1e-18 ETH). Often, the price of gas is specified in Gwei (1e-9 ETH).

2) Smart Contract: N. Szabo first suggested the idea of smart contracts in a 1994 unreleased article, and afterwards officially in 1997 [12]. In a nutshell, smart contracts are blockchain technology software that are self-executing, short programmes that encapsulate the user contract between the parties. Smart

contracts differ from those other contracts in that it automatically detect whether or not the provisions of the contract have been met. Furthermore, the programmes (smart contracts) are duplicated across numerous different blockchains nodes to guarantee dependability, fault tolerance, and transparency. It's noteworthy that during the past five years, the Ethereum blockchain platform has held almost 3.5 million smart contracts [13].

B. Big Data Processing

Big data refers to information that is always expanding rapidly and has grown too big to also be analysed using traditional techniques. A network of heterogeneous network can use for the multiple processors of massive amounts of data, which is achieved by big data processing frameworks and tools like Apache Hadoop. A Hadoop cluster is essentially a group of interconnected, master-node-communicating inexpensive PCs (or virtualization software). "Moving the processing to the information" is the Hadoop concept. This requires that every node act as both a processor as well as a storage area. Those two parts, YARN (Yet Another Resource Negotiator - compute) plus HDFS (Hadoop Distributed File System - memory), were developed to cooperate in a single group [14,15]. Hadoop can scale from either a particular node to a thousand nodes by dispersing all information and running the processing concurrently. For distributed system the data amongst some of the slave node in Hadoop, it employs MapReduce. The information is broken down into a series of blocks, which are all allocated to a cluster node, which works much like the blockchain. Some other huge data framework developed on top of HDFS is Apache Spark. Spark offers quick in-memory computing by cutting down on disc reads and writes [16]. In fact, it is appropriate for both batches & stream information, providing it an flexible options available for a wide variety of large applications that process data. These activities involve graph processing, machine learning, and dynamic big data querying. Furthermore, Python, Java, and R APIs are available for Spark, which is developed in Scala.

C. Common Factors

As was already said, it can be extremely fascinating to watch how social media networks and complicated networks like the Ethereum blockchain communicate. Understanding the occurrences that lead to market conditions to shift may be possible through analysing the social issues. In light of this, Search engine Trends has been selected as the source of data for gauging interests of the public in cryptocurrencies like Ether and Bitcoin. The information displays the interest of the public as determined either by total number of gathered search queries. Data from Google Searches is openly accessible subject to certain limitations on data recurrence.

III. Data And Tool

The dump provided to a repo on Google's BigQuery were where the Ethereum dataset that was utilised for this investigation had first been assembled. This dataset is now accessible by everyone. A Hadoop cluster received the dataset, which was then saved in HDFS. The dataset was processed using Spark (in this case, primarily its Python API, PySpark), because of its outstanding speed and efficiency. From the dataset, the month total for transactions, the average amount of gas consumed, and the gas cost were retrieved.

VI. Analyses And Results

Figure 1 depicts an initial study of the data and shows an increase in transactions, especially smart contracts, in the beginning of 2019. It's fascinating how this occurrence and the recent increase in Ether as well as Bitcoin prices go hand in hand. We'll go into greater detail about any potential connections between both the values of Ether as well as Bitcoin. Another finding indicates a fairly consistent decline in gas prices, with a small spike at the start or end of every year. Moreover, as of late 2022, the typical gas consumed for transactions appeared to have stabilised. Finally, the significant correlation (pearson product moment correlation value of 0.98 between the intricacy of a smart contract as well as the required gas) between those two factors can be explained through the notion that even more sophisticated contracts would demand more gas.

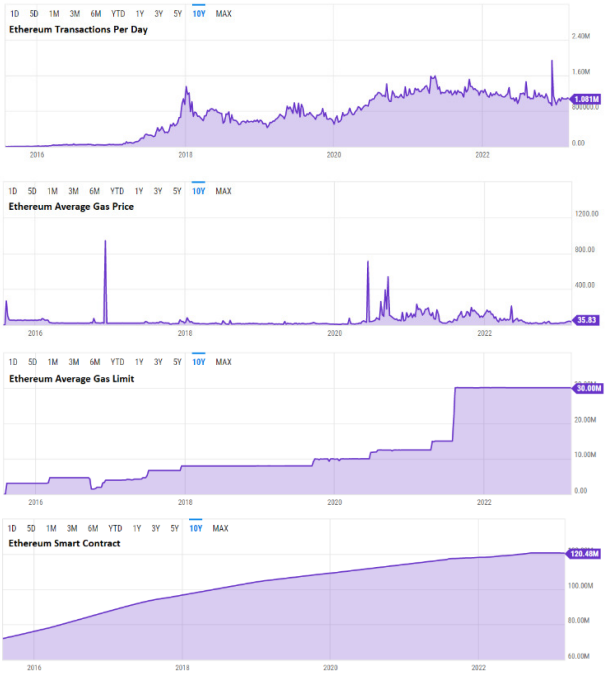

Fig. 1. Ethereum transactions and gas over a period of time

The data from Google Searches is now included in the study as an exogenous characteristic. The very first logical assumption may be that as even more individuals look up Ethereum internet, more of them are going to likely invest in and use the system for operations. Hence, fig 2 shows the amount of transactions as well as the level of interest in Ethereum among the general population. The frequency of internet searches also isn't represented on the y axis, which solely represents the amount of transactions. Everyday Web Analytics data were gathered at 180-day periods, scaled, concatenation, and then normalised on a scale of 100. They have indeed been adjusted yet again to suit the y value of a figure for better visibility. The overall results demonstrate show, after the initial sharp surge in Web Analytics information in the middle of 2021, the transaction volume did not undergo the very same profound shift, potentially indicating a lack of confidence in cryptocurrencies and blockchain technology. In any case, the platform's transaction volume peaked in the beginning of 2019 at

the same time that internet search activity peaked. A simple conclusion may be that a rise in the amount of transactions made correlates to a rise in the best interests of the country in a cryptocurrency. But, using this knowledge to anticipate how these things will impact the value of ETH may be a more alluring goal.

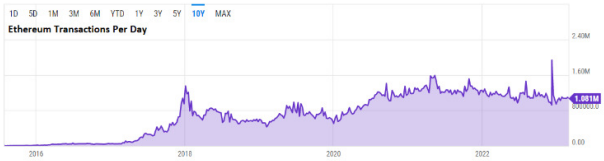

Fig. 2. Number of transactions in similar-web

The value of ETH and Bitcoin (BTC), as well as their each volumes, are plotted in fig 3 using the same analogies as above. The increase in ETH/USD in 2022 is thought to have been tied to the increase in BTC/USD, which was brought on by Tether (USDT). Griffin as well as Shams' analysis of the Bitcoin blockchain revealed that Tether-based transactions were timed to occur after the decline in BTC/USD, which in turn sparked the sharp increase in Bitcoin ’s price [17].

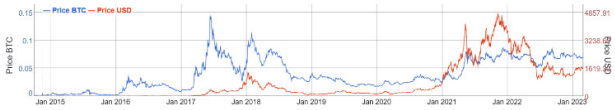

Fig. 3. Cryptocurrency prices in similar-web

A comparison of the cryptocurrency pair initially reveals a similar pattern. Early in 2021, as demand in the cryptocurrency industry increased, the exchange rates for ETH and BTC both soared. The prices of cryptocurrencies, especially ETH, would not decline as anticipated in line with public interest after the comparable abrupt decline in search terms during 2021. This could be a sign that more than just regular consumers are getting involved because whales (big, strong businesses) may be waiting to make significant investments until after the downturn. This presumption is in line with the findings of Griffin and Shams [17], who concluded that demand as well as supply aren't the sole factors affecting the price of cryptocurrencies. According to our investigation, a significant player on Bitfinex who invested a significant amount of Tether in Bitcoin produced the pricing distortion.

When choosing what commodities to invest in, a statistical study of the return on digital currencies can be helpful in light of the factors influencing the shift in crypto values. Figure 4 shows the exponential return of ETH/USD and BTC/USD displayed over time in order to create a time series with stationary characteristics that represents changes in cryptocurrency prices.

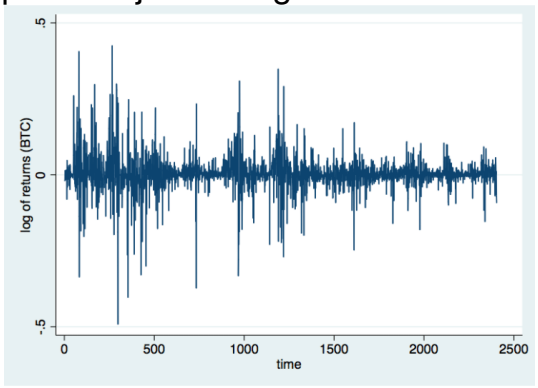

Fig. 4. Cryptocurrency log returns over a period of time

Given that both appear to reflect a gaussian distribution, it may be concluded from analyzing the probability density function (PDF) between Ether and Bitcoin's log returns (fig. 5). Nevertheless, Bitcoin's log returns Pattern looks to be smaller and much more pronounced than Ether's, making it a better option for investors who are risk averse.

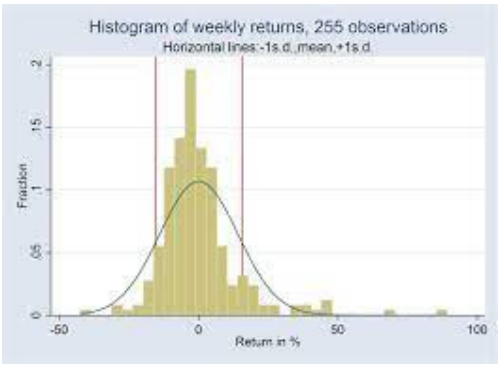

Fig. 5. Cryptocurrency log returns histogram

Table 1 provides a more detailed and quantified breakdown of the log returns. It's interesting to note that Ether's log return reached a level of 80%, which is about 3 times greater than Bitcoin's. Notwithstanding, Ether's minimal log yield had fallen to -81%, which is almost 1.5 times smaller than Bitcoin's.

TABLE-1: Cryptocurrency Log Returns

	ETH/Wei	BTC/USD	%Reth	%Rbtc
Mean	347.30	6455.57	4.38	1.25
Std Dev	384.43	6688.89	7.90	5.02
Minimum	1.43	303.18	-91.26	-31.59
Maximum	1945.82	47984.18	90.40	31.45

Lastly, overall cross-correlation between the four consecutive periods is computed to determine the time it takes for adjustments to be mirrored in cryptocurrency cost. The days that pass till the second or third time series has the highest cross-correlation with the first is indicated by the time lag, or "lag max," that corresponds to the greatest cross-correlation. A heatmap representing the highest cross-correlation between every pair of the 4 time periods is shown in Figure 6. The pairs (ETH/USD, BTC/USD), (ETH/USD, GoogleBTC)and (GoogleETH, GoogleBTC) had the greatest cross-correlation scores, indicating that they could be future opportunities.

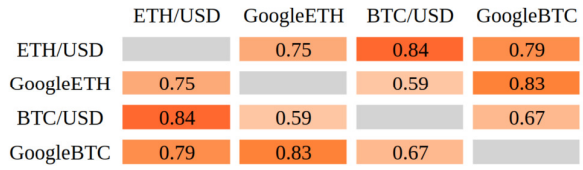

Fig 6. Max cross-correlations heatmap

Figure 7 displays the modifications in cross-correlations in relation to the time lag. It can be seen that the (ETH/USD, GoogleBTC) pair has the most lag time. Simply put, this shows how an growth in the quantity of Bitcoin-related google searches is correlated with just an improvement in the ETH/USD about a couple of months later. The (ETH/USD, GoogleBTC),(ETH/USD, GoogleETH) and (ETH/USD, BTC/USD) pairs rank highest for timely forecasts according to time lags.

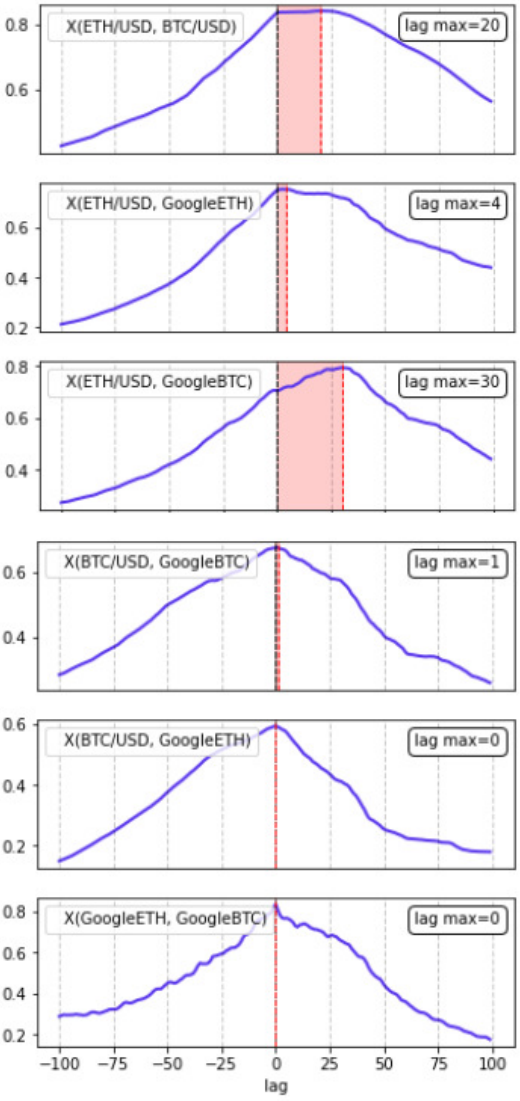

Fig. 7. Cross-correlations time log period

It should be emphasised that the degree of cross-correlations should also be taken into account when using this information in subsequent analysis and predictions.

In contrast to certain other pairs, the (ETH/USD, BTC/USD) and (ETH/USD, GoogleBTC) combinations offer both precise and on-time predictions. Moreover, it is very important to remember how although the aforementioned arguments emphasise on cross-correlation, this idea shouldn't be confused for causality. It is necessary to use additional techniques, including such Bayesian knowledge in understanding and a thorough knowledge of economics and Blockchain, to investigate the potential causal relations.

V. Conclusions

The results in this article imply that taking into account the interactions among social as well as financial complicated networks improves asset values forecasts. Although just two resources were examined in this study, it might have been advantageous to include more elements in the assessment. This method also makes it easier to spot marketplace scams and cryptocurrency pricing bubbles that are the result of non-public actors. In order to get more information, future analysis can include data from other social sources of data.

References

1. Grilli R, Tedeschi G, Gallegati M. Business fluctuations in a behavioral switching model: Gridlock effects and credit crunch phenomena in financial networks. Journal of Economic Dynamics and Control. 2020. p. 103863. doi:10.1016/j.jedc.2020.103863
2. Demythifying the belief in cryptocurrencies decentralized aspects. A study of cryptocurrencies time cross-correlations with common currencies, commodities and financial indices. Physica A: Statistical Mechanics and its Applications. 2020;556: 124759.
3. DIMACS (GROUP), Haber S, Scott Stornetta W. How to Time-stamp a Digital Document. 1990. 4. Angraal S, Krumholz HM, Schulz WL. Blockchain Technology: Applications in Health Care. Circ Cardiovasc Qual Outcomes. 2017;10. doi:10.1161/CIRCOUTCOMES.117.003800
5. Du X, Chen B, Ma M, Zhang Y. Research on the Application of Blockchain in Smart Healthcare: Constructing a Hierarchical Framework. J Healthc Eng. 2021;2021: 6698122.
6. Subramanian N, Chaudhuri A, Kayıkcı Y. Blockchain Applications and Future Opportunities in Transportation. Blockchain and Supply Chain Logistics. 2020. pp. 39–48. doi:10.1007/978-3-030-47531-4_5
7. Niu B, Mu Z, Cao B, Gao J. Should multinational firms implement blockchain to provide quality verification? Transportation Research Part E: Logistics and Transportation Review. 2021. p. 102121. doi:10.1016/j.tre.2020.102121
8. Kumar M. Applications of Blockchain in Digital Forensics and Forensics Readiness. Blockchain for Cybersecurity and Privacy. 2020. pp. 339–364. doi:10.1201/9780429324932-20
9. Adebayo A, Rawat DB, Njilla L, Kamhoua CA. Blockchain - enabled Information Sharing Framework for Cybersecurity. Blockchain for Distributed Systems Security. 2019. pp. 143–158. doi:10.1002/9781119519621.ch7
10. Sukhija N, Sample J-G, Bautista E. Advancing the Cybersecurity of Electronic Voting Machines Using Blockchain Technology. Essentials of Blockchain Technology. 2019. pp. 235–256. doi:10.1201/9780429674457-11
11. Werbach K. The Blockchain and the New Architecture of Trust. MIT Press; 2018.
12. Szabo N. Formalizing and securing relationships on public networks. First monday. 1997. 13. Giuseppe Antonio Pierro, Roberto Tonelli, Michele Marchesi. An Organized Repository of Ethereum Smart Contracts' Source Codes and Metrics. Future Internet.
2020;12: 197.
14. Vavilapalli VK, Murthy AC, Douglas C, Agarwal S, Konar M, Evans R, Graves T, Lowe J, Shah H, Seth S, Saha B. Apache hadoop yarn: Yet another resource negotiator. Proceedings of the 4th annual Symposium on Cloud Computing. 2013. pp. 1–16.
15. Borthakur, D. The hadoop distributed file system: Architecture and design. In: Hadoop Project Website. 2007.

16. Zaharia M, Xin RS, Wendell P, Das T, Armbrust M, Dave A, Meng X, Rosen J, Venkataraman S, Franklin MJ, Ghodsi A. Apache spark: a unified engine for big data processing. Commun ACM. 2016;28: 56–65.
17. Griffin JM, Shams A. Is Bitcoin Really Untethered? The Journal of Finance. 2020;75: 1913–1964.