finalpaper

Preprint · November 2019

DOI: 10.13140/RG.2.2.25797.19682

3 authors:

Rushikesh Mokashi
Thakur College of Engineering and Technology

Saksham Khetan
Thakur College of Engineering and Technology

Sheetal Vikram Rathi
Thakur College of Engineering and Technology

The relationship of various cryptocurrencies with Bitcoin

Rushikesh Mokashi, Saksham Khetan, Sheetal Rathi

Abstract*: *Cryptocurrencies have taken the world by storm. These virtual currencies may be unregulated by a central authority but have a strong foothold in the trading market. In circumstances like these, it becomes a discussion of substance whether one can accurately predict the price of these currencies. Our aim in this paper is to determine the relationship between Bitcoin and various popular cryptocurrencies viz. XRP, TRON, LTC, ETH. We will be using various correlation coefficients such as Pearson's and Spearman's coefficient to understand which are the currencies that are most related to Bitcoin. The data being used is collected from Binance. Individual correlation coefficients, as well as multiple regression coefficients, will be calculated and ranked accordingly. Based on these results, we will confirm them using a regression algorithm. It follows that the higher-rated cryptocurrencies will be better predictors of Bitcoin. This paper aims to be useful for financial sectors and data analysts, who use various methods for predictions and pattern analysis.

Keywords:* *Cryptocurrencies, Relationship, Bitcoin, Regression.

I. INTRODUCTION

Digital currencies have been created by mankind to ease transactions and provide more security and accessibility. Modern Cryptocurrencies use decentralized control usually through distributed ledger technology, such as the blockchain. One currency that exploded into mainstream media was the Bitcoin that grew into a billion-dollar industry. Post the Bitcoin, nearly 1500+ alternate variations of cryptocurrencies also known as 'altcoins' have been created and circulated. The market capitalization of these currencies is more than 100 billion dollars i.e. larger than the GDP of 127 countries. This paper is concerned with the various algorithms that will be used to predict the price of Bitcoin by considering the effects of Altcoins. Through experimental data, various algorithms will be applied and their accuracies would be noted. The final goal is to obtain the best algorithm and a subset of altcoins that will give the highest accuracy.

A. HISTORY

The formation of Bitcoin in 2009 denoted the introduction of the primary computerized money to accomplish broad appropriation over the globe. Notwithstanding, the idea of safe computerized money has been around since the 1980s and there have been numerous past endeavors that straightforwardly enlivened Satoshi Nakamoto's making of Bitcoin.

DIGICASH

In 1982, PC researcher David Chaum discharged the paper Blind Signatures for Untraceable Payments in which he laid out an option in contrast to the electronic exchanges hitting retail locations at the time. His paper is viewed as one of the primary propositions of computerized cash ever. He propelled an organization called Digi Cash in 1990 to popularize the thoughts in his exploration.

BIT GOLD

Viewed as an immediate forerunner to Bitcoin, Bit Gold was considered to give a progressively trustless model of executing, given the monetary properties of gold with expanded security. Bit Gold and Bitcoin are comparable because Bit Gold intended to actualize a proof-of-work style accord system where registering force is utilized to illuminate cryptographic riddles.

B-CASH

B-cash was a proposition for an "unknown, dispersed electronic money framework" made by PC engineer Wei Dai in 1998. In his exposition, Dai proposed two conventions, the first he knew was unfeasible because it required "substantial utilization of synchronous and un-jammable unknown communicate channel" yet would fill in as inspiration for the subsequent convention.

II. LITERATURE REVIEW

A. Predictive Analysis

Predictive analytics refers to the methodology of deriving useful knowledge from data to give a rough forecast of what might happen in the future. Cryptocurrencies are known for their extreme volatility as people constantly trade in real-time. This makes the domain highly risk dominated and not a very safe investment. While many traders might not consider the impact that big data has on cryptocurrency pricing, the potential outcomes of finding some patterns are very promising. Sentiments play a major role in driving the value of a cryptocurrency. A positive word of mouth will skyrocket its prices while a negative buzz will plummet the currency to the bottom. So, we understand that structured data, as well as indirect unstructured information, can affect the price of a particular currency. Using analytics on existing data, one can benefit from cryptocurrency [1]. Blockchain is the framework that records digital currency exchanges, which happen on a distributed (P2P) organize. The blockchain is a record of all the digital currency exchanges that have ever happened on that arrange. Predictive analytics can clue companies into market trends, including:

- History of coin costs and exchanging volumes.
- Which factors are influencing the cost of a coin, for example, past interest, new showcase guidelines?

- Regardless of whether value developments will be a present moment or long haul.
- If highs will rapidly move to lows and the other way around or if highs will keep on climbing or lows will keep on plunging.

A.I. can be utilized to examine information about cryptographic money exchanging, blockchain exchange information, just as different factors, to give clients information about the anticipated development of explicit digital currencies. Furnished with this data, more purchasers will have the option to enter this thriving universe of speculations with certainty. A major possibility is that fluctuations in the pricing of one currency can affect the price of another currency drastically. That forms the major crux of our topic which we intend to find results for by experimentation.

B. Coefficients Used

The estimation of correlation coefficients is far-reaching in research as it measures the outcomes obtained and the connections between these yields. We usually test the null hypothesis of zero correlation and/or confidence intervals for the correlation are computed. There are a few distinct strategies for this purpose; we think about the performance of various techniques. The two coefficients we will consider here are - Pearson's correlation coefficient and Spearman's correlation coefficient [2].

Spearman's correlation coefficient:

Spearman's correlation is used to measures the strength between two variables which are ranked along with the association direction. To use Spearman’s coefficient, we need two ordinal variables, interval or ratio. Related variables get their strength measured by this value. Similar ranked observations give a higher Spearman’s correlation coefficient value.

Pearson's correlation coefficient:

Pearson’s correlation coefficient provides a measure of the strength of the linear association between two variables. It has a value taking a range from +1 right up to −1. It is often calculated by taking the covariance of the two given variables. This result is further divided by the product of their standard deviations giving the final value. A value of 1 implies that a linear equation has all data points lying on a line where we have Y increasing as X also increases. -1 gives a line where Y decreases but X increases. Finally, 0 implies that there exists no correlation whatsoever between the variables. It is an extremely useful parameter in our aim to extract a relationship between various cryptocurrencies as this gives information about the magnitude of the association between currencies and their fluctuating prices. Pearson’s correlation coefficient is independent of the unit of measurement giving us a pure number. We make use of a scatterplot to graphically determine the spread of values and if there exists any form of association [5]. Pearson correlations are suitable only for metric variables whereas, for ordinal variables, we make use of the Spearman correlation.

Multiple Correlation and Regression

In measurements, the coefficient with respect to the multiple correlation is an extent of how well a given variable can be foreseen using a straight limit of a variety of components. It is the relationship between the variable's characteristics and the best conjectures that can be enlisted straightforwardly from the prescient factors. Multiple Regression serves to predicts a single variable from the weighted direct entire of numerous factors while multiple correlation evaluates the nature of this relationship.

Linear Regression

Linear regression is frequently acknowledged as the direct way to deal with demonstrating the connection between a dependent variable and at least one independent variable. When one independent variable is in picture, we call it simple linear regression. However, in case of more than one independent variable, the procedure is called multiple linear regression. The center thought is to acquire a line that best fits the information. The best fit line is the one for which all-out expectation blunder are as little as could be expected under the circumstances. Error is the separation between the point right up to the regression line [7].

We have Y as a variable dependent on X which is our independent variable. The regression line is:

$$Y = B0 + B1X$$

Here B0 is a constant that is determined depending on the condition and requirement and B1 is known as the regression coefficient. In the case of a more complex, multi-variable environment consisting of several variables, the linear equation might look like this.

$$f\ (x, y, z) = w1x + w2y + w3z$$

Here w_w represents the coefficients. We will use these values to make our model learn and get trained.

.

III. METHODOLOGY

In this paper, we aim to find the various relationships that exist between several cryptocurrencies. By considering data of a pre-specified period, we aim to show the hidden pattern that exists between currencies where a rise/fall in one currency is accompanied by a similar variation in some other currency. We will be using Linear Regression to ultimately train and test our model.

A. Data Set

To compare multiple currencies and their prices, we make use of existing historical data from the website Binance. Using the Python programming language, we download data about cryptocurrencies from this website. The dataset used in this paper is from 8th October 2018 – 8th October 2019, with data taken in hourly intervals. The major advantage of this is that a web API (in this case Binance) is a specially framed URL which returns machine-readable data instead of returning a webpage. This data set gives us the freedom to select data over a specific period and a custom time interval that we can select [4]. The 5 currencies that we will be considering are – Bitcoin (BTC)[9], Ethereum (ETH), Ripple (XRP), Tron (TRON), and Litecoin (LTC).

B. Preprocessing the Data

Now that we have obtained the set of values for 1 year with an hourly interval, we have with us a table comprising the date and the value of the currency on that particular date. We repeat this procedure for all the five mentioned currencies and get 5 distinct tables. As the penultimate of preprocessing, we apply the inner join and combine all the 5 sets to give us one big table. This table contains the date under consideration and values of all currencies on that day. The data type selected is float to represent accurate measures.

C. Correlation and Test Data

Based on our defined interval, we obtain 24 values daily for a period of 1 year. To ensure the efficiency of our model, we use the first 6000 values as data for deriving the Pearson's correlation coefficient for both 1vs1 as well as 1vs2 relation between currencies. Values post the initial 6000 values are used to verify our hypothesis using the Linear Regression method.

D. Pearson's Coefficients Values

We have already defined what the Pearson's correlation coefficient is and how can one find the value for the same. Using our refined data set and the predefined formula, we calculate the value for the currencies and obtain the following values.

Pearson's correlation BTC ETH: 0.968
Pearson's correlation BTC XRP: 0.646
Pearson's correlation BTC TRON: 0.582
Pearson's correlation BTC LTC: 0.810

	BTC	ETH	XRP	TRON	LTC
BTC	1.000000	0.968475	0.645665	0.581843	0.809867
ETH	0.968475	1.000000	0.668756	0.663495	0.807356
XRP	0.645665	0.668756	1.000000	0.131875	0.199106
TRON	0.581843	0.663495	0.131875	1.000000	0.738602
LTC	0.809867	0.807356	0.199106	0.738602	1.000000

E. Spearman's Coefficient Values

As mentioned earlier, Spearman's coefficient of correlation is also a reliable measure of correlation. Thus, to verify the pattern obtained by Pearson's coefficient, we compute the value of Spearman's coefficient and obtain the following results:

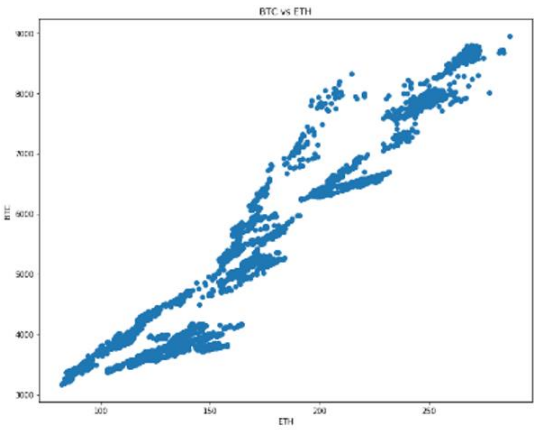
Spearman's correlation BTC ETH: 0.944
Spearman's correlation BTC XRP: 0.655
Spearman's correlation BTC TRON: 0.409
Spearman's correlation BTC LTC: 0.800

Thus, we observe that the findings correspond accurately to the relations derived by Pearson's coefficient of correlation.

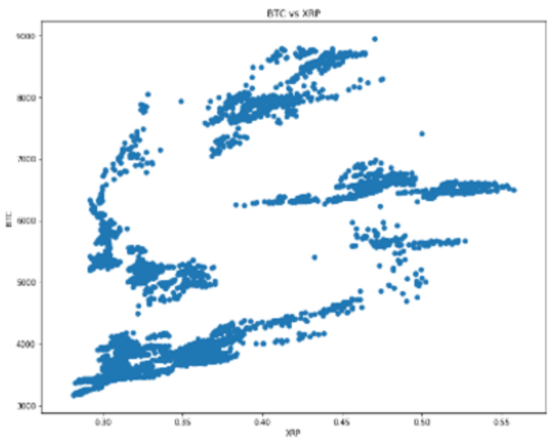
F. Plotting Graphs

To determine which currencies are closely related to each other, we have derived the Pearson's coefficient above. A value close to 1 shows a high dependency rate between the corresponding currencies. To visualize any association between currencies better, we plot the graphs.

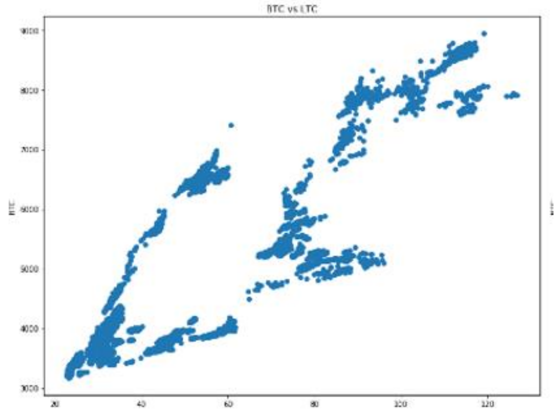
BTC vs ETH



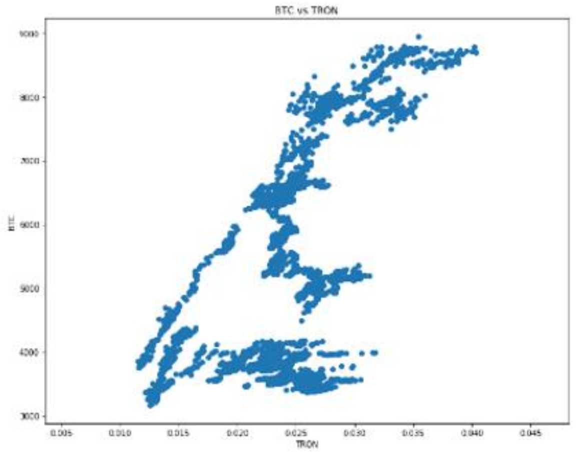
BTC vs XRP



BTC vs LTC



BTC vs TRON



G. Multiple Correlation

We have calculated the value of coefficients above by considering a 1vs1 relation between currencies. However, in the real world, we often have more than just 2 variables and it is equally important to calculate the relations existing between them as well. Let us consider a case where we have 3 variables namely *x*, *y*, and *z*. In such a case, we define multiple correlation coefficient as:

$$R_{z.xy} = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xz}r_{yz}r_{xy}}{1 - r_{xy}^2}}$$

Where 'r' for two variables can be defined as –

$r = cov(x,y)/s_x s_y$

x and y are the independent variables with z being the dependent variable [6]. The results obtained show a definite relation between all the possible combinations and help derive a pattern showcasing currency changes as a pair i.e. 1vs2 correlation.

BTC vs (ETH and XRP) is: 0.968
BTC vs (ETH and TRON) is: 0.972
BTC vs (ETH and LTC) is: 0.97
BTC vs (XRP and TRON) is: 0.817
BTC vs (XRP and LTC) is: 0.949
BTC vs (TRON and LTC) is: 0.81

IV. TESTING

A. Testing Using Linear Regression

From the multiple correlation performed above, we know that BTC vs. ETH and XRP will give the best accuracy. As mentioned earlier, entries post the 6000 were to be used as test data for Linear Regression.

The variables used for our Linear Regression model are:
Test Size = 0.25
Random state = 0

To determine the accuracy of our experiment and how well the model fits the given data, we make use of two parameters namely: R-Squared and root mean squared error (RMSE).

R-squared is a factual proportion of how close the information is to the regression line. It is otherwise called the coefficient of assurance, or the coefficient of different determination for various regression. R-squared is characterized as the level of the reaction variable variety clarified by a direct model. On the other hand, root mean squared error is a quadratic scoring value that additionally measures the normal extent of the blunder. It's the square base of the normal squared contrasts among expectation and genuine perception.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}$$

The top 3 correlation coefficients obtained are:
1. BTC vs (ETH and TRON) is: 0.972
2. BTC vs (ETH and LTC) is: 0.97
3. BTC vs ETH is: 0.968

Corresponding Linear Regression R2 and RMSE values obtained are –

Currencies	Correlation	R2	RMSE
BTC vs (ETH and TRON)	0.972	0.45	704128.58
BTC vs (ETH and LTC)	0.970	0.43	726126.62
BTC vs ETH	0.968	0.42	732144.82

Thus, a decreasing pattern of R2 and an increasing pattern of RMSE is observed.

This confirms our correlation strategy and yields expected results

V. RESULT AND DISCUSSION

Linear regression verifies the Pearson's coefficient values obtained. The r2 and RMSE values follow patterns as expected of from our covariance results. Since linear regression is the most basic algorithm for predictive analysis, we feel that the results of our study will be even better supported through more complex algorithms such as RNN and LSTM.

VI. CONCLUSIONS

Thus, we have established the relation that Ethereum and TRON are the best predictors of Bitcoin price fluctuations. Conversely any fluctuations in Bitcoin can cause similar patterns in Ethereum and TRON respectively. The causes of this pattern are unknown and might be subject to further analysis. This paper was limited to 1v1 and 1v2 analysis of cryptocurrencies. A larger subset of cryptocurrencies could potentially yield better results and thus could be the starting point for further analysis. Also, this study covers the most popular cryptocurrencies in the market right now, hence further expanding the pool of cryptocurrencies could potentially yield better results.

REFERENCES

1. Laura Alessandrettia, Abeer ElBahrawy, Luca Maria Aiello, and Andrea Baronchelli. Machine Learning the Cryptocurrency Market. (2018) SSRN
2. Richard Taylor, EdD, RDCS. Interpretation of the Correlation Coefficient: A Basic Review. (1990) Journal of Diagnostic Medical Sonography
3. de Winter, Joost C. F, Gosling, Samuel D, Potter, Jeff. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. (2016). American Psychological Association.
4. Binance.com. (2019) Binance. [Online]-Available: https://www.binance.com/
5. Karl Pearson. Notes on the History of Correlation. (1920) Biometrika
6. Neter, John, Michael H. Kutner, Christopher Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*. (1996) McGraw Hill, Boston.
7. David A. Freedman. Statistical Models: Theory and Practice. (2009) Cambridge University Press.
8. Lehman, Ann. Jmp For Basic Univariate and Multivariate Statistics: A Step-by-step Guide. (2005) Cary, NC: SAS Press.
9. https://bitcoin.org
10. https://www.sciencedirect.com/science/article/abs/pii/S027553191530026X
11. https://www.researchgate.net/publication/323772682_Relationship_between_lower_limb_anthropometry_and_temporo-spatial_parameters_in_gait_of_young_adults