ON CORRELATION OF BITCOIN AND ALTCOINS

Mohammed Anas Hathif Sunkesula, Matouš Cejnek

Ústav přístrojové a řídicí techniky, ČVUT v Praze, Fakulta strojní, Praha

Abstrakt: V tomto projektu byla zkoumána korelace mezi kryptoměnami: Bitcoin, Monero, Ethereum a další. Korelace byla zkoumána pomocí Pearsonova a Spearmanova korelačního koeficinetu. Bylo dohledáno, že v roce 2016 byla korelace mezi sledovanýma kryptoměnama velmi nízká, skoro žádná. V roce 2017 korelace vzrostla. Následně v roce 2018 se korelace sledovaných kryptoměn výrazně zvýšila. V diskuzi tohoto článku jsou představeny a porovnány možné hypotézy proč tato korelace existuje.
Klíčová slova: Bitcoin, kryptoměny, korelace

Abstract: In this Project, a research on topic of correlation analysis of Bitcoin, Monero, Ethereum and other top cryptocurrencies is reviewed and concluded. We investigate correlation coefficients of cryptocurrencies with two different approaches: Pearson and Spearman. It has been concluded that in 2016 the correlation between the selected cryptocurrencies was very weak - almost none, but in 2017 the correlation increased and became moderate positive. In 2018, almost all of the cryptocurrencies were highly correlated.
Keywords: Bitcoin, cryptocurrency, correlation

1 Introduction

Cryptography derives from the Greek words kryptos (hidden, secret) and graphein (to write) [1]. Essentially is the method of sending a message in a concealed way to prevent other parties from reading it. The transmission of information becomes secure with the help of cryptography. Cryptography is useful for user authentication and in protecting the data from thieves.

Cryptocurrency (Herein referred to as cryptos) is a digital currency, based on principles of cryptography [1]. The main technology behind the cryptos is a blockchain [2]. A blockchain is essentially a distributed database of records, or public ledger of all transactions or digital events that have been executed and shared among participating parties. Each transaction in the public ledger is verified by consensus of a majority of the participants in the system. Once entered, information can never be erased. The blockchain contains a certain and verifiable record of every single transaction ever made. To use a basic analogy, it is easier to steal a cookie from a cookie jar, kept in a secluded place, than stealing the cookie from a cookie jar kept in a market place, being observed by thousands of people. Bitcoin (herein referred to as BTC) is the most popular example that is intrinsically tied to blockchain technology. It is also the most controversial one since it helps to enable a multibillion-dollar global market of anonymous transactions without any governmental control. Hence it has to deal with a number of regulatory issues involving national governments and financial institutions [3]. Due to the electronic nature of cryptos, they are extremely easy to use in different countries without having any technical restrictions. Though governments around the world have targeted crypto users (often under money-laundering regulations for individuals and financial regulations for would-be intermediaries), their success has been mixed. The protocols (crypto market) cannot be targeted or shut down; the best that can be done is to pinpoint prominent individual users, a drawn-out and expensive process. This is the reason that BTC has not succumbed to the same fate as the Liberty Dollar [4, 5].

Interesting feature of cryptos is lower transaction fees compared to the high fees that banks charge. [6]. Cryptos use cryptography for two major purposes. Firstly, to make sure that transactions are secured and secondly having control-govern of the monetary issuance.

BTC [1] is a digital currency and is probably one of the biggest discoveries in the financial system of the 21st century. It's production, storage, movement and all transactions with it, is made exclusively in electronic form between the participants of BTC network [7]. No country, government or bank produces or controls BTC.

As per One of the articles in CNBC [8], an analyst has suggested that the main reasons for different exchangers to have different rates on the cryptos are due to the fact that each exchange has different quantities of these cryptos. The smaller the demand in the market, the better the rates of the cryptos. According to the article, BTC's different prices were noted in one of the biggest crypto markets,

BTC Prices in different markets on 12th December 2017: Kraken - $17,212; GDAX -$17,150; Gemini - $17,039; Bitfinex - $16,957; Bitstamp - $16,979.

Also one of the other reasons for this factor is that the prices for these currencies is not fixed so most of the other markets would be pricing the BTCs as per the flow of the demand. For example if one big market sets a price, then the other lower markets would try to adjust their prices accordingly but at the end of the day, it depends on how much supply each market has, that causes these change in the prices.

Another challenge is the inefficient way to manage a balance in different exchange markets. Traders who have a vast investment, practice the act of arbitrage, where the crypto is bought at a lower price and is sold in a different market for a higher price. Sometimes in order to make a profit, huge amount of transactions have to be made, which in the other hand consumes a lot of time, which can result in the change of the exchange rates too. So it can be quite risky to get into the act of arbitrage [8].

According to the data of BTC distribution, There are many addresses that have their own distinctive prices of the BTCs they are selling. This causes a lot of differences in the exchange rates of these cryptos in the market. Seeing this opportunity, consumers also try to practice the act of arbitrage, But this act involves a lot of time and risk. It is believed currently there are 23.7 million BTC addresses. But the active users are very less compared to this figure approx 3 to 6 million. This is found out through the survey by Cambridge center for alternative finance. Their justification to this figure is based on only the active users and not all those that hold an account for cryptos [9].

In this paper we study correlation in order to decide which cryptos one can invest in. For instance if a person chooses two cryptos of opposite correlation nature, the person may not loose much, as from either of the cryptos the person may gain some profit. Correlations help decide investors which cryptos to invest in and what is their behaviour going to be in near future in order to avoid huge looses.

2 Correlation of cryptos

In mathematics, specifically in statistics, we are interested in measuring and analyzing the degree of relationship between pairs of random variables. This statistical relationship can be examined through a statistical technique called correlation, that shows whether or not and how strong the variables are related to each other. The research of how variables are correlated is called correlation analysis. Correlation refers to the relationship between mean values and it helps us to understand two valuable things. First, we can measure and analyze the degree of relationship between a pair of variables and secondly it shows us a forecasting relationship that can be utilized in practice. However, the existence of the correlation is not adequate to deduce the existence of a causal relationship.

Another similarity measure is mutual information, [10] It is a distance between two probability distributions. Correlation is a linear distance between two random variables. The mutual information is between any two probabilities defined for a set of symbols, while there cannot be a correlation between symbols that cannot naturally be mapped into a R raised to N space. On the other hand, the mutual information does not make assumptions about some properties of the variables. If the data is smooth, correlation may tell more about them; for instance if their relationship is monotonic. If there are some prior information, then among both the methods we can be able to switch from one to another [11]. The equation for mutual information follows:

$$I(A;B) = \sum_{b \epsilon B} \sum_{a \epsilon A} p(a,b) * log(\frac{p(a,b)}{p(a)p(b)}), \tag{1}$$

where A is the feature data and B is the output data [12]. The general rule is that the greater the mutual information, the more informative the feature. The price data of the other cryptos apart from BTC is collected, Once we collect this data, using Pearson correlation coefficient [13], we check how correlated the cryptos are. In the correlation method, If the correlation coefficient is +1 then both the cryptos are moving in the same direction, If the correlation coefficient is -1 then both the cryptos are moving in the opposite direction to each other. If it is 0, then there is no correlation between the two cryptos.

In Fig 1 we can see the daily percentage change price of BTC and Monero (hereafter referred as XMR)

It is widely understood that the use of larger samples in applications of factor analysis tends to provide results such that sample factor loadings are more precise estimates of population loadings and are also more stable, or less variable, across repeated sampling.

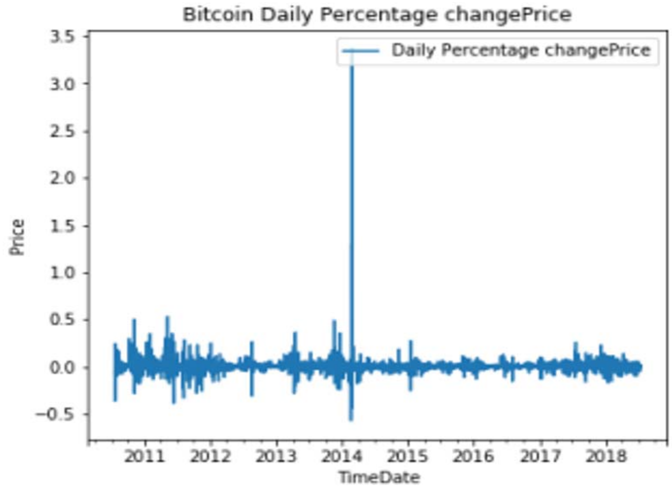

Fig. 1: Daily percentage change price of BTC and XMR [1]

Fig. 2: Prices of cryptos [1]

Summary of hypotheses provides a basis for the following hypotheses about effects of sample size in factor analysis:

1. As N increases, sampling error will be reduced, and sample factor analysis solutions will be more stable and will more accurately recover the true population structure.

2. Quality of factor analysis solutions will improve as communalities increase. In addition, as communalities increase, the influence of sample size on quality of solutions will decline. When communalities are all high, sample size will have relatively little impact on quality of solutions, meaning that accurate recovery of population solutions may be obtained using a fairly small sample. However, when communalities are low, the role of sample size becomes much more important and will have a greater impact on quality of solutions.

3. Quality of factor analysis solutions will improve as over determination of factors improves. This effect will be reduced as communalities increase and may also interact with sample size [14].

The relationship between BTC-XMR looks like linear for most data points, but for very positive returns seems to be confusing. Because the scatter plot between BTC-XMR doesn't give us clear results of their correlation, it is better to calculate the Pearson and Spearman correlation coefficient. A few top cryptos are selected from July 2016 - July 2018 to have a better opinion how more cryptos are correlated with each other. Before finding the correlation of those cryptos, it would be good to see their prices in a graph.

	BTC	XMR
ADF Statistic	-1.29466	-1.67071
p-value	0.631611	0.446265
Critical Values		
1%	-1.67071	-3.436
5%	-2.898	-2.864
10%	-2.586	-2.568

Fig. 3: Augmented Dickey-Fuller Statistics [1]

2.1 Why Is It Useful to Find the Correlation Between the Cryptos?

Portfolio management is about how you invest in the product's development resources – project prioritization and resource allocation across development projects. There are four goals in portfolio management – maximizing the value of the portfolio, seeking the right balance of projects, ensuring that portfolio is strategically aligned, and making sure there aren't too many projects for limited resources. And there are many tools – some quantitative, others graphical, some strategic – designed to help chose the right portfolio of projects. In short, Portfolio management is fundamental to new product success. But it's not as easy as it first seems. Not only we must seek to maximize the value of the portfolio, but the development projects in the portfolio must be appropriately balanced, there must be the right numbers of projects, and finally, the portfolio must be strategically aligned. No one portfolio model can deliver on all four goals, and so best-practice businesses tend to use multiple methods to select their projects. [15]

There are three important reasons why we need to find the Pearson correlation coefficient between the cryptos.

- Help us to invest in two different cryptos that don't cancel each other out. Assume two cryptos, BTC and Ethereum where their correlation coefficient r approximates -1. That means that the price of one cryptocurrency, for example BTC, will rise while the other will decrease, as we have a negative correlation. If the investor wants to invest on both of them, then any potential profit will be cancelled out.

- Hedging is an attempt of securing against negative events and it is very similar with the first reason. The difference is that someone can hedge his investment by choosing to invest in a crypto that moves opposite from his initial investment to make up any losses. The disadvantage of this choice is that we will have smaller profits when the initial crypto that we selected moves strong.

- Instead an investor putting all his money into one crypto, he can diversify the risk of his choice by purchasing other cryptos that have different correlation coefficient. For example, someone can look on different cryptos that have constantly correlation close to 0. Diversification plays an important role in modern portfolio theory which says that we can diversify away risk, by reducing the correlation between the returns of the selected assets in the portfolio. Despite the fact that the Pearson correlation coefficient has some advantages, many analysts mention some of it's drawbacks. First of all, Pearson's method is only valid when we have 100% linear dependencies and this is not so often observed. However, we can surpass that obstacle by using Spearman's correlation coefficient. Moreover correlation between cryptos changes very often because it is a new area and we are still in a period of precariousness. A common mistake that many do when they are trying to find the correlation is that they use "raw data". The correct correlation can be found from daily returns (i.e percentage changes) for each crypto. Returns are simpler to work with and give a better approximation to the investor for its return. They provide better results when the price series is non-stationary, but most important, returns will captivate changes in the strength of the network effects of different cryptos [1]. Usually lot of people use "raw data" (non-stationary) without making them returns concluding to false correlation and results.

  In time series analysis the first thing that someone has to do before proceeding with any method, is to check whether the data that is being dealt with is stationary or not. If a series is non-stationary then we can easily conclude to inaccurate results. To check whether or not the time series are stationary or not, Augmented Dickey Fuller (Herein referred as ADF) test which is implemented in Python is used. If the ADF statistic is less than the critical value, then the null hypothesis is rejected implying that the series is stationary and we don't have to make further transformations. If the null hypothesis can't be rejected then the series is not stationary. In fig 3 we see the results of the ADF test. [1]

  As we can see, the ADF test is above the 5% critical value and the p-value> 0.05. As a result, the null hypothesis of the ADF test cannot be rejected and the series is not stationary. To make the data stationary,

	BTC	XMR
ADF Statistic	-5.777333	-6.887006
p-value	0.000001	0.00002
Critical Values		
1%	-3.514	-3.436
5%	-2.898	-2.864
10%	-2.586	-2.568

Fig. 4: Augmented Dickey-Fuller Statistics [1]

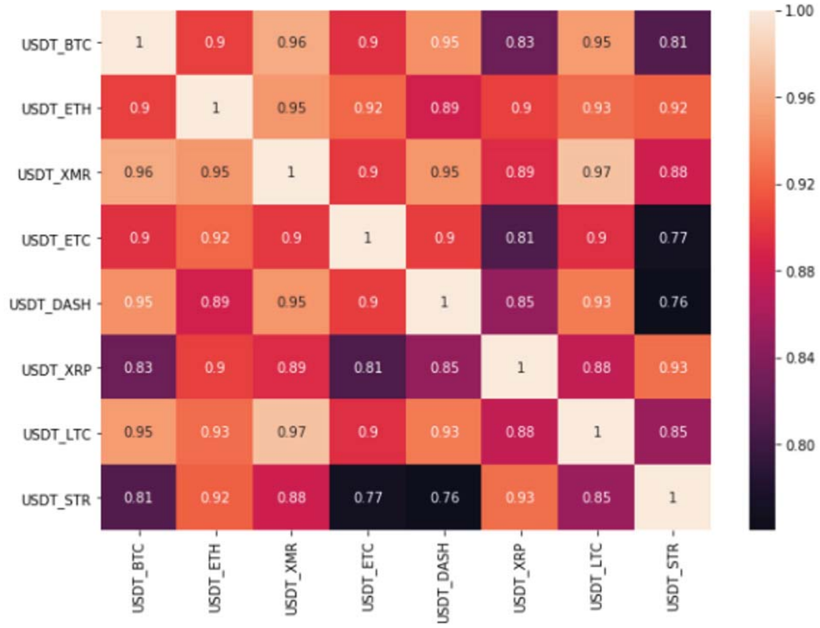

Fig. 5: Pearson correlation of "raw" data from July 2016-July 2018 [1]

we use first differences, i.e: [1]

$$y[t] = p[t] - p[t-1], \tag{2}$$

As we can see in fig 4, after using first differences our data became stationary. The test statistic is less than the critical value at 5% and the p-value < 0.05, so it can be concluded with 99% confidence that the data is stationary. Now that we have our data stationary, we can look at the correlation of BTC and XMR.[1]

In a crypto time series analysis, the correlation requires stationary data which means constant variance over time and be mean reverting. Stationarity in correlation is achieved by converting the daily close prices into daily returns (percentage price changes). Then it will be mean-reverting, as the returns move above and below a constant mean and the variance will be constant. In fig 1 we can see the daily percentage change price of BTC and XMR. [1]

In Fig. 5 we can see the Pearson correlation of "raw data" and after we will show the Pearson Correlation coefficient of daily returns.

The difference between Spearman and Pearson is that Spearman correlation ranks the datapoints and after is running a Pearson correlation. In Spearman correlation the two variables are compared based on how much their rankings move together [1]. It's main purpose is to find co-motions in datasets that are non-linear. In a perfect positive correlation, Spearman has a value of +1, whereas Pearson is very close to +1. Now we will see, in figure 5 and 6, how the cryptos are correlated every year from July 2016 - July 2018.

2.2 Why are the cryptos correlated?

As per one of the research paper, The correlation of the cryptos (other cryptos to BTC) was higher in the year 2018, As high as 0.87 or higher [16]. In order to understand how correlated the cryptos are, we will require

	USDT_BTC	USDT_ETH	USDT_XMR	USDT_ETC	USDT_DASH	USDT_XRP	USDT_LTC	USDT_STR
USDT_BTC	1.000000	0.904083	0.959776	0.896917	0.945015	0.825950	0.951504	0.812522
USDT_ETH	0.904083	1.000000	0.948842	0.924132	0.885633	0.903664	0.926248	0.920089
USDT_XMR	0.959776	0.948842	1.000000	0.899866	0.948350	0.893156	0.974611	0.882038
USDT_ETC	0.896917	0.924132	0.899866	1.000000	0.903181	0.809784	0.895109	0.769262
USDT_DASH	0.945015	0.885633	0.948350	0.903181	1.000000	0.849962	0.934257	0.760369
USDT_XRP	0.825950	0.903664	0.893156	0.809784	0.849962	1.000000	0.875075	0.927506
USDT_LTC	0.951504	0.926248	0.974611	0.895109	0.934257	0.875075	1.000000	0.852919
USDT_STR	0.812522	0.920089	0.882038	0.769262	0.760369	0.927506	0.852919	1.000000

Fig. 6: Table pearson correlation of "raw data from July 2016- July 2018 [1]

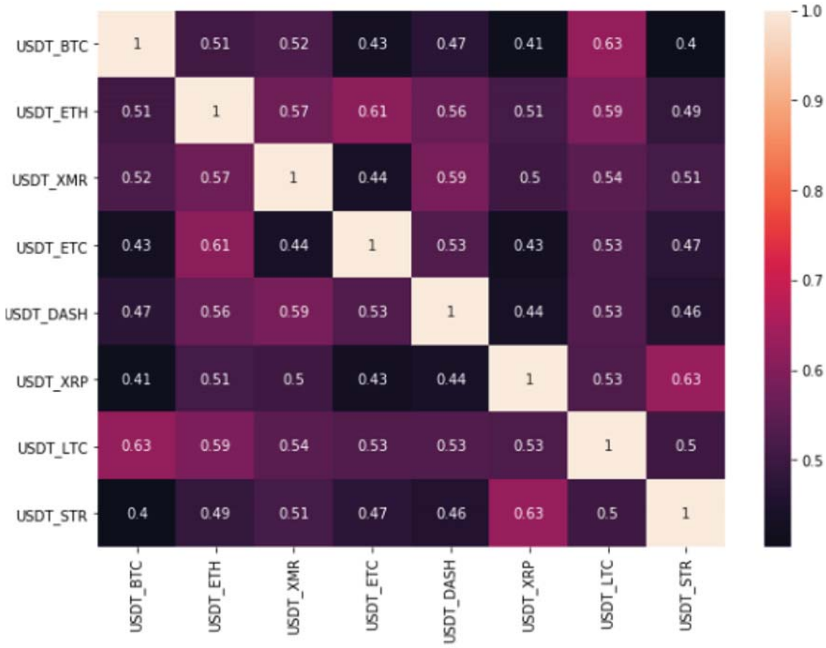

Fig. 7: Spearman correlation from July 2016-July 2018 [1]

information about: a) The price data of the crypto we would like to compare with other cryptos; b) The entire price data of the other cryptos.

The main reasons for the cryptos to be correlated to BTC are:

- Liquidity: BTC is tradable against almost all altcoins, volume represents the significant percentage of the industry, due to these factors it becomes feasible for trade to taking place in cryptos which means a rise in the top cryptos. Hence the correlation with the BTC and top cryptos rises [17]

- Dominance: Market capitalization accounts for nearly half of the crypto assets the cryptos are moderate correlated. The results showed that in 2016, there was a very weak correlation for almost all the cryptos and they were quite independent from each other. Only the pair of Litecoin (herein referred to as LTC) - BTC standed out with a strong positive correlation of 0.75, meaning that they were moving in the same direction. But that changed over the next years. In 2017 the correlation became moderate positive for most of the cryptos. The highest correlation for that year was between Ethereum (herein referred to as ETH)- Ethereum classic (herein referred to as ETC) and ETH - XMR with 0.6 and 0.55 respectively. ETC - BTC and Ripple (herein referred to as XRP) - BTC had the lowest value with 0.11 and 0.13 respectively. In 2018, almost all the cryptos that we selected were highly positive correlated with each other. This analysis give us valuable and interesting information about cryptos.

  In study [18] relays this in a study and shows that, while not always, investor sentiment can predict stock prices. Considering these facts, investor sentiment could be a significant factor in the price movement of crypto prices,

  This study is interesting because it shows several factors which may influence price and gives an overview of what are some of the more important variables. The research showed that 20% of BTC's price is driven

	USDT_BTC	USDT_ETH	USDT_XMR	USDT_ETC	USDT_DASH	USDT_XRP	USDT_LTC	USDT_STR
USDT_BTC	1.000000	0.507222	0.520892	0.429775	0.472280	0.413657	0.626738	0.403122
USDT_ETH	0.507222	1.000000	0.570340	0.613165	0.562456	0.512416	0.592494	0.485525
USDT_XMR	0.520892	0.570340	1.000000	0.440158	0.585619	0.502613	0.543955	0.514640
USDT_ETC	0.429775	0.613165	0.440158	1.000000	0.529056	0.428267	0.530878	0.474469
USDT_DASH	0.472280	0.562456	0.585619	0.529056	1.000000	0.438555	0.532482	0.460311
USDT_XRP	0.413657	0.512416	0.502613	0.428267	0.438555	1.000000	0.526234	0.630520
USDT_LTC	0.626738	0.592494	0.543955	0.530878	0.532482	0.526234	1.000000	0.501635
USDT_STR	0.403122	0.485525	0.514640	0.474469	0.460311	0.630520	0.501635	1.000000

July 2018

Fig. 8: Table of Spearman correlation from July 2016 [1]

by investors' attractiveness to BTC as determined by the volume of Google search queries. This was the major determinant of price in the study. the results indicate that positive sentiment (conveyed through the variable: 'attractiveness to BTC') affects BTC's price, the authors showed that the remaining 70% of BTC's price movements is explained by 'its own innovative shocks' which is an ambiguous explanation, effectively relying on using the residual as the signal of systemic unexplained component of price formation.

If BTC acts as a safe haven or a hedge for currencies. Therefore, stock market indices have an expectation to be negatively correlated with the price of BTC. Considering the interpretations of fear in regards to asset prices and the contradicting arguments for and against BTC being a hedge against the stock market, it seems appropriate to isolate the interpretation of a correlation between cryptos and fear as being 'cryptos are a hedge against the stock market in times of fear' and not just 'cryptos are a hedge against the stock market' over all stock markets environments [18].

- Similar behaviour: One important reason why some cryptos are correlated with bitcoin is, they belong to the same family meaning bitcoin and the other cryptos have similar function and prospects for a long time. (eg: LTC and BTC share similar behavior) [19]

- Limited or no options for trading or investing other currencies than BTC/fiat, BTC/ETH. Investors must first buy BTC or ETH before they can exchange one of these for the desired crypto. This pushes the dominance of BTC over the market, resulting in a "follow the leader" behaviour. At Pool of Stake we think that any investor should take notice of the correlation between cryptos. This can be used as an useful tool to determine macro trends inside this crazy and volatile market. Some of the advantages of studying the correlations between pairs are: a) Avoiding investing in two coins that cancel each other out. You invest in two different cryptos but the correlation coefficient tends to -1. This means that if the price of one is going up, the other will go down, therefore cancelling out the potential profit of your acquisition. b) Diversify your portfolio. Instead of investing 100% of your portfolio in only one coin, you should find another closely correlated coin and split some of the investment to diversify the portfolio but still take advantage of the pair's inertia. If one asset is in danger of losing its value, you still have 50% of your account allocated to safer position. [19]

3 Conclusion

In this project, we focused on a new area, the area of cryptocurrencies specifically on Bitcoin, Ethereum, Litecoin and Monero. The goal of this project is to study the correlation of BTC with different cryptos. We focus on correlation of cryptos from July 2016 to July 2018. After we talked about time series and how to make the data stationary with the help of the Augmented Dickey-Fuller test (ADF) [20]. Additionally, we discussed the Pearson and Spearman correlation coefficient of the history data presented in other research papers. The main reasons for the cryptos to be correlated to BTC are: liquidity, dominance, similar behaviour, limited or no options for trading or investing other currencies than BTC/fiat, BTC/ETH.

Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. *SGS18/177/OHK2/3T/12*.

Literature

[1] Angelos Kalaitzis. Bitcoin-monero analysis: Pearson and spearman correlation coefficients of cryptocurrencies, 2018.

[2] Melanie Swan. *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc.", 2015.

[3] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, Vignesh Kalyanaraman, et al. Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2(6-10):71, 2016.

[4] Cameron Harwick. Cryptocurrency and the problem of intermediation. *The Independent Review*, 20(4):569–588, 2016.

[5] Lawrence H White. The troubling suppression of competition from alternative monies: The cases of the liberty dollar and e-gold. *Cato J.*, 34:281, 2014.

[6] Colin LeMahieu. Raiblocks: A feeless distributed cryptocurrency network. *URL https://raiblocks. net/media/RaiBlocks_Whitepaper__English. pdf*, 2017.

[7] Satoshi Nakamoto et al. Bitcoin: A peer-to-peer electronic cash system. 2008.

[8] Bob Pisani and Todd Haselton. Here's why bitcoins prices are different on each exchange-, 2017.

[9] Alex Morris. U today- how many people own bitcoin and how they use it., 2018.

[10] Joy A Thomas and TM Cover. Elements of information theory. *John Wiley & Sons, Inc., New York. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, MPH (2009),"Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," Journal of the Royal Society Interface*, 6:187–202, 1991.

[11] Pau Vilimelis Aceituno (https://stats.stackexchange.com/users/135502/pau-vilimelis aceituno). Mutual information versus correlation. Cross Validated. URL:https://stats.stackexchange.com/q/241312 (version: 2016-10-20).

[12] Alex Greaves and Benjamin Au. Using the bitcoin transaction graph to predict the price of bitcoin. *No Data*, 2015.

[13] David Freedman, Robert Pisani, and Roger Purves. Statistics (international student edition). *Pisani, R. Purves, 4th edn. WW Norton & Company, New York*, 2007.

[14] Robert C MacCallum, Keith F Widaman, Shaobo Zhang, and Sehee Hong. Sample size in factor analysis. *Psychological methods*, 4(1):84, 1999.

[15] Robert Gravlin Cooper, Scott J Edgett, and Elko J Kleinschmidt. Portfolio management for new products. 2001.

[16] Binance Research. Are cryptoassets highly correlated?, 2019.

[17] Adam Hayes. Cryptocurrency value formation: an empirical analysis leading to a cost of production model for valuing bitcoin. *Telematics and Informatics, Forthcoming*, 2016.

[18] Constantin Gurdgiev, Daniel O'Loughlin, and Bartosz Chlebowski. Behavioral basis of cryptocurrencies markets: Examining effects of public sentiment, fear and uncertainty on price formation. *Forthcoming: Journal of Financial Transformation*, 49, 2019.

[19] Pool of Stake. Cryptocurrency correlation, 2018.

[20] Yin-Wong Cheung and Kon S Lai. Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280, 1995.

Selected article from
Tento dokument byl publikován ve sborníku

Nové metody a postupy v oblasti přístrojové techniky, automatického řízení a informatiky 2019
New Methods and Practices in the Instrumentation, Automatic Control and Informatics 2019
27. 5. – 29. 5. 2019, Zvíkovské Podhradí

ISBN 978-80-01-06617-1

Web page of the original document:
http://iat.fs.cvut.cz/nmp/2019.pdf

Obsah čísla/individual articles:
http://iat.fs.cvut.cz/nmp/2019/

Ústav přístrojové a řídicí techniky, FS ČVUT v Praze, Technická 4, Praha 6