CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

INSTITUTE OF ECONOMIC STUDIES

Master's thesis

2020 Viktoriia Ilina

Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Is hype really that powerful? The correlation between mass and social media and cryptocurrency rates fluctuations

Author: **Viktoriia Ilina**

Study program: **Economics and Finance**

Supervisor: **Mgr. Michal Král**

Academic Year: **2020**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently; using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 5, 2021 Viktoriia Ilina

Acknowledgments

The author is grateful to Mgr. Michal Král to for his enormous patience, encouragement and feedback from the first to the last day. Besides, special thanks to Vladimir Pyrlik, M.A. for a deep insight into the world of econometrics.

Abstract

Twelve years after Satoshi Nakamoto published the paper describing the functioning mechanism and principals of cryptocurrency that maintains secure and anonymous digital transactions beyond any banks, cryptocurrencies have become a multi-billion-dollar industry comprising millions of investors, miners, developers and profiteers. However, the actual price determinants and ways to forecast future price changes remain an open question yet to discover the answer for. This study attempts to figure out whether media hype exerts that much influence upon cryptocurrencies price movements and whether it can be used as the basis for future movements prediction. Two cryptocurrencies, Bitcoin and Tezos, and 7 mass and social media factors for each of them were considered on daily basis from 08-01-2018 to 10-31-2020. To explore the interdependence between media drivers and cryptocurrencies' prices in short, medium and long timespan, this study deploys wavelet coherence approach. There was found, that price changes turn to be the supreme prior to hype, even though the growing ado may push the prices even higher. Thus, hype is failing to prove itself as a reliable cryptocurrency price predictor. Crypto investors, though, should anyways take the news background into account while building trading strategies, especially for new projects in the market.

JEL Classification	C12, G12, G14, G15, G41
Keywords	Cryptocurrency, Bitcoin, Tezos, Media, Time series analysis, Sentiment analysis, Wavelet coherence
Title	Is hype really that powerful? The correlation between mass and social media and cryptocurrency rates fluctuations

Abstrakt

Dvanáct let poté, co Satoshi Nakamoto publikoval svoji studii, popisující principy a mechanismy, díky kterým jsou digitální transakce bezpečnější a anonymnější než ve kterékoli bance, se kryptoměny staly multi-miliardovým odvětvím s miliony investory, vývojáři, minery a a spekulanty. Skutečné cenové determinanty a způsoby prognózy budoucích cenových změn však zůstávají otevřenou otázkou, na kterou je třeba ještě najít odpověď. Tato studie se se snaží zodpovědět, jak moc je cena kryptoměn ovlivněna zprávami z medií a sociálních sítí tzv. "media hype" a zda může být tato veličina použita jako prediktor pro změnu ceny. Dvě kryptoměny, Bitcoin a Tezos a 7 mediálních faktorů pro každý z nich byly zvažovány na denní bázi od 08-01-2018 do 10-31-2020. V zájmu prozkoumání vzájemné závislosti medií a ceny kryptoměn v krátkém, středním a dlouhodbém období tato studie používá přístup založený na vlnkové soudržnosti. Ukázalo se, že změna cen je to, co způsobuje "media hype", avšak tento media hype může tuto změnu umocnit. Proto "media hype" nemůžeme považovat za spolehlivý prediktor ceny kryptoměn. I přesto by měli investoři do kryptoměn vzít media a sociílní sítě v úvahu při určování investičních strategií, zvláště pro nové projekty.

Klasifikace	C12, G12, G14, G15, G41
Klíčová slova	Kryptoměna, Bitcoin, Tezos, Média, Analýza časových řad, Analýza sentimentu, Vlnková soudržnost
Název práce	Je „hype“ opravdu tak mocný? Korelace mezi masovou a sociální médii a fluktuacemi hodnoty kryptoměn

Contents

List of Tables

List of Figures

Acronyms

API Application Programming Interface

BTC Bitcoin

EOS The native cryptocurrency for the EOS system

ICO Initial coin offering

LTC Litecoin

ROI Return on investment

XMR Monero

XRP Ripple

XTZ Tezos

Master's Thesis Proposal

Author:	Viktoriia Ilina
Supervisor:	Mgr. Michal Kral
Defense Planned:	February 2021

Proposed Topic:
Is hype really that powerful? The correlation between mass and social media and cryptocurrency rates fluctuations

Motivation:
There is a ton of ado about cryptocurrencies in mass and social media around the world. Cryptocurrencies have evoked global interest among a vast variety of people from housewives to bankers and become a huge headache for governments trying to take it under control. Some are claiming that cryptocurrencies are nothing more than gold fever of century XXI and generally a deceiving bubble, while others call them the new digital gold and something to completely change the world. There is less disagreement about the underlying blockchain technology, that allows digital information to be recorded and distributed, but not edited. Blockchain technology was first outlined in 1991 by Stuart Haber and W. Scott Stornetta, two researchers looking to implement a system where document timestamps function tamper-proof, yet it took to invent the Bitcoin technology to find its first practice application. Potential appliance of blockchain, however, goes far beyond the world of cryptocurrencies, as it can be used for patient's medical data protection, maintaining electoral fairness, keeping records of property, or securing copyrights. As said ma Jack Ma, the founder and chairman of Alibaba Group Holding, "Blockchain technology could change our world more than people imagine, Bitcoin, however, could be a bubble." Warren Buffet, perhaps the best-known Bitcoin critic, argues even more offensively - "Bitcoin is probably rat poison squared". In Buffet's view, Bitcoin has no unique value at all, being nothing more but a delusion. He, however, admits, the importance of blockchain technology itself, as its functioning is completely separate from cryptocurrencies and its success has nothing to do with Bitcoin.
Only time will show whether it is something more than an alleged bubble, but still cryptocurrencies are and undeniable phenomenon of the XXI century. High volatility, the signature hallmark of cryptocurrencies and a sticking point for most arguments around them, which is utterly hindering their global implementation as a payment instrument, but at the same time attracting numerous new players to the crypto market. The history of cryptocurrencies knows enough examples of dramatic price changes being preceded by news – both those of global meaning, like the announcement of ICO ban by the government of China, which immediately dropped Bitcoin price from $5,000 to $3,000, as well as fake news, deceiving news headlines, flattering promises and influencers' tweets, as it happened in December 2017 with a tweet by Jhon McAfee regarding the "big future of altcoin with anonymized transactions", that pumped up the price of an altcoin called Verge an unbelievable 13-fold.

But is it true, that media hype is the key driver of such volatility, or are there other reasons lying under? The scientific community has not come to a general consensus, as various studies result with contradictory implication. Thus, the issue remains open to further exploration and keeps plenty of room for new research works.

Hypotheses:

1. Hypothesis #1: There is an undeniable correlation between mass and social media and cryptocurrencies rates fluctuations.
2. Hypothesis #2: Monitoring of media hype around cryptocurrencies can be used as the tool for cryptocurrency rates prediction.
3. Hypothesis #3: The influence of mass and social media is much stronger in the short run.

Methodology:

After a through exploration of literature related to cryptocurrencies price formation in general and the impact exerted by media in particular, further follows the data collection from various sources. Posts and comments extracted from Reddit themselves connote an unstructured, unorganized data set. In specie it is a random bundle of information from web users expressing their feelings, views, emotions or sharing their experience, so comes in carrying poor relevance in terms of further processing, analysis, and prediction. By this means, a preparatory sentiment analysis is required, for which VADER analyzer has been chosen within this study. Then, the time series must undergo stationarity testing, and properly transformed once found non-stationary. Pearson correlation coefficient is used to estimate the liaison between explored variables. In order to go beyond correlations and develop a better understanding with regard to the causalities, the study utilizes Granger causality analysis. Finally, for the purpose of exploring the interdependence between media drivers and prices of the cryptocurrencies in short and long timespan this study deploys wavelet coherence approach.

Expected Contribution:

The key goal of the study is to figure out whether media hype actually exerts that much of influence upon cryptocurrencies price movements and whether it can be used as the basis for future movements prediction. For this purpose, two cryptocurrencies (Bitcoin and Tezos) and 7 media factors (Google trends, Wikipedia views, number of tweets, news volume, Telegram mentions, Reddit sentiment, Reddit total number of posts and comments) for each of them have been considered on daily basis from 08-01-2018 to 10-31-2020.

Outline:

1. Introduction
2. Literature review (definition of cryptocurrency, history, advantages and disadvantages, factors influencing its price)
3. Data and methodology (detailed data collection, sentiment analysis of Reddit data, methodology of time series analysis)
4. Empirical results and discussion (including ways to overcome the existing limitations)
5. Conclusion

Core Bibliography:

1. Abraham, J., Higdon, D., Nelson, J., Ibarra, J., 2018. Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. SMU Data Science Review, 1(3), 1.
2. Betre, C., 2019. Analysis of the relationship between Prices and Transactional Drivers of Bitcoin. Senior Independent Study Theses. Paper 8364.
3. Ciaian, P., Rajcaniova, M., Kancs, D., 2016. The economics of BitCoin price formation, Applied economics, 48(19), 1799–1815.
4. Ciaian, P., Rajcaniova, M., Kancs, D., 2018. Virtual relationships: Short- and long-run evidence from BitCoin and altcoin markets. Journal of International Financial Markets, Institutions and Money, 52, 173–195.
5. ElBahrawy, A., Alessandretti, L., Baronchelli, A., 2019. Wikipedia and Cryptocurrencies: Interplay Between Collective Attention and Market Performance. Frontiers in Blockchain, 2, 12.
6. Erzurumlu, Y.O., Oygur, T., Kirik, A., 2020. One size does not fit all: external driver of the cryptocurrency world. Studies in Economics and Finance, 37(3), 545–560.
7. Kraaijeveld, O., De Smedt, J., 2020. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money. Journal of International Financial Markets, Institutions and Money, 65, C.
8. Kristoufek, L., 2013. BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. Scientific reports, 3(1), 3415.
9. Kristoufek, L., 2015. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PloS one, 10(4), e0123923.
10. Phillips, R., Gorse, D., 2018. Cryptocurrency price drivers: Wavelet coherence analysis revisited. PloS one, 13(4), e0195200.
11. Rognone, L., Hyde, S., Zhang, S., 2020. News sentiment in the cryptocurrency market: An empirical comparison with Forex. International Review of Financial Analysis, 69, 101462.
12. Salač, A., 2019. Forecasting of the cryptocurrency market through social media sentiment analysis. Bsc Dissertation, University of Twente. [online] Available at: http://essay.utwente.nl/78607/1/Salac_BA_BMS.pdf.
13. Sovbetov, Y., 2018. Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero. Journal of Economics and Financial Analysis, 2, 1-27.
14. Steinert, L., Herff, C., 2018. Predicting altcoin returns using social media. PloS one, 13(12), e0208119.
15. Valencia, F., Gómez-Espinosa, A., Valdés-Aguirre, B., 2019. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy, 21(6), 589.

1 Introduction

Twelve years after previously unknown Satoshi Nakamoto published the paper describing the functioning mechanism and principals of cryptocurrency, that maintains secure and anonymous digital transactions beyond central or commercial banks, cryptocurrencies have become a multi-billion-dollar industry comprising millions of investors, miners, developers, profiteers, and fortune seekers. By the end of December 2020, total market capitalization of cryptocurrencies has exceeded $713 billion.

There is a ton of ado about cryptocurrencies in mass and social media around the world. Cryptocurrencies have evoked global interest among a vast variety of people from housewives to bankers and become a huge headache for governments trying to take it under control. Some are claiming that cryptocurrencies are nothing more than gold fever of century XXI and generally a deceiving bubble, while others call them the new digital gold and something to completely change the world. There is less disagreement about the underlying blockchain technology, that allows digital information to be recorded and distributed, but not edited. Blockchain technology was first outlined in 1991 by Stuart Haber and W. Scott Stornetta, two researchers looking to implement a system where document timestamps function tamper-proof, yet it took to invent the Bitcoin technology to find its first practice application. Potential appliance of blockchain, however, goes far beyond the world of cryptocurrencies, as it can be used for patient's medical data protection, maintaining electoral fairness, keeping records of property, or securing copyrights. As said ma Jack Ma, the founder and chairman of Alibaba Group Holding, "Blockchain technology could change our world more than people imagine, Bitcoin, however, could be a bubble." Warren Buffet, perhaps the best-known Bitcoin critic, argues even more offensively - "Bitcoin is probably rat poison squared". In Buffet's view, Bitcoin has no unique value at all, being nothing more but a delusion. He, however, admits, the importance of blockchain technology itself, as its functioning is completely separate from cryptocurrencies and its success has nothing to do with Bitcoin.

Only time will show whether it is something more than an alleged bubble, but still cryptocurrencies are and undeniable phenomenon of the XXI century. High volatility, the signature hallmark of cryptocurrencies and a sticking point for most arguments around them, which is utterly hindering their global implementation as a payment instrument, but at the same time attracting numerous new players to the crypto market. The history of cryptocurrencies knows enough examples of dramatic price changes being preceded by news – both those of global meaning, like the announcement of ICO ban by the government of China, which immediately dropped Bitcoin price from $5,000 to $3,000, as well as fake news, deceiving news headlines, flattering promises and influencers' tweets, as it happened in December 2017 with a tweet by Jhon McAfee regarding the "big future of altcoin with anonymized transactions", that pumped up the price of an altcoin called Verge an unbelievable 13-fold.

But is it true, that media hype is the key driver of such volatility, or are there other reasons lying under? The scientific community has not come to a general consensus, as various studies come up with contradictory implication. While some appeal to technical and technological aspects (Sovbetov, 2018; Hayes, 2017), others see the connection with macroeconomic and financial market variables (Smith, 2016; Wjik, 2013), and yet another group of researchers lean towards media exposure and sentiments of crypto community (Kaya, 2018; Mai et al., 2018). Thus, the issue remains open to further exploration and keeps plenty of room for new research works.

The key goal of the study is to figure out whether media hype actually exerts that much of influence upon cryptocurrencies price movements and whether it can be used as the basis for future movements prediction. The underlying hypotheses include:

• There is an undeniable correlation between mass and social media and cryptocurrencies rates fluctuations.

• Monitoring of media hype around cryptocurrencies can be used as the tool for cryptocurrency rates prediction.

• The influence of mass and social media is much stronger in the short run.

The thesis is structured as follows: Chapter 2 reviews the background and literature relevant to this research, including definition of cryptocurrency, history, advantages, disadvantages, and factors influencing its price. Chapter 3 provides a detailed description of data collection for the explored variables, including a review of the media data sources, sentiment analysis of Reddit data, and defines the research methodology. Chapter 4 comprises empirical outcomes of the carried-out research, limitations and outlines potential future extensions. The findings of this study are summarized in Chapter 5.

2 Literature review

This chapter is covering the theoretical agenda of the crypto markets and the key drivers affecting the price of cryptocurrencies. Theoretical background of this paper is based on an extract from other academic research, books, articles and approved web resources.

2.1 Cryptocurrency – definition, history, pros and cons

As defined by Lansky (2018), cryptocurrencies are type of digital currencies that applies cryptography to maintain financial transactions, keep the emission of new units under control, and verify asset transfer. As a rule, cryptocurrencies comprise a unique setup of three features – safeguarding limited anonymity, fencing off central authority penetration, and protection against double spending. Cryptocurrencies are the one-of-a-kind entity possessing this combination of features, that no other group of currencies, including fiat currencies, could present.

During last years the dimensions of technology and investment have been under a massive inflow of cryptocurrencies, blockchain applications, and related projects, initiatives or ventures. Despite a snowballing amount of new digital currencies rushing into the market and transforming it, there is just a single digital currency holding the bay tree of the broad public's acknowledgement and interest – Bitcoin (BTC). A majority of traders, investors and general crypto adherers assume Bitcoin to be the original cryptocurrency. It is, though, hard to imagine the advent of Bitcoin, not to mention hundreds of other digital currencies released afterwards, without a legacy of similar attempts reaching decades back in the past.

The original timeline begins back in 1982 with David Chaum (University of California, Berkley) publishing his study titled „Blind Signatures for Untraceable Payments" introducing the technological advancements to public and private core technology. Chaum' Blind Signature Technology intended to provide assured and complete privacy for online financial transfer users. In 1989 Chaum established DigiCash, a company settled in Amsterdam, that honed in on digital money and payment systems. The expertise of DigiCash included maintaining the governmental initiative aimed at replacing toll booths (eventually aborted) and smart cards, something in the mold of nowadays' hardware wallets. The signature of DigiCash, though, was eCash – a digital cash system, that let the company run the first-ever electronic cash transaction online back in 1994. At a moment, the system exploded the mass media, yet failed to achieve relevant extent, so the company ended up getting adjudged bankrupt. “It was hard to get enough merchants to accept it, so that you could get enough consumers to use it, or vice versa,” Chaum explained in his Forbes interview in 1999, after DigiCash had finally filed for bankruptcy. “As the Web grew, the average level of sophistication of users dropped. It was hard to explain the importance of privacy to them.” (Pitta, 1999)

Nonetheless, even though the technology failed as a business, Chaum’s legacy inspired further generations of cryptographers, hackers, and activists. So, in 1998 a computer engineer Wei Dai published his study on “B-money” introducing and dissecting the concept of a digital currency, which could be transferred along a group of untraceable digital pseudonyms. The same year, a blockchain pioneer Nick Szabo discharged Bit Gold, his own attempt to create a decentralized digital currency. Szabo’s intention was to counter the drawbacks of the traditional financial system – among other, the need for metal to produce coins, and excessive amount of required to run transaction. Even though neither of the concepts ever got officially launched, both served part of the groundwork Bitcoin raised up on.

On August 18th, 2008, Bitcoin.org was registered through https://www.anonymousspeech.com/, a platform enabling anonymous domain registration. On October 31st Satoshi Nakamoto (a web pseudonym whose identity remains undisclosed) published the paper “Bitcoin: A Peer-to-Peer Electronic Cash System” distributed via https://www.metzdowd.com/ cryptography mailing list describing the Bitcoin currency and introducing the solution against double spending, that cuts off any chance for the currency to be copied. On November 9th, the Bitcoin project appeared at https://sourceforge.net/, a public sharing platform for open-source software development and distribution. On January 3rd, 2009, Satoshi Nakamoto mined the genesis block marking the date as the one Bitcoin got set off as viable cryptocurrency. Just nine days later, on January 12th Satoshi Nakamoto and Hal Finney, a developer and cryptographic activist, ran the world’s first Bitcoin transaction in the volume of 50 BTC. Later that year, on October 5th, New Liberty Standard publicly established the value of a Bitcoin at US$1 = 1,309.03 BTC, the rate based on and equation including the cost of electricity supply consumed by a computer for Bitcoin production. A week later the first known BTC to fiat exchange took place by a Finnish developer Martti Malmi selling 5,050 BTC for $5,05 with the fiat payment running via Paypal. The first real-world purchase using Bitcoin occurred on May 22nd, 2010, in Jacksonville, Florida, where a programmer Laszlo Hanyecz set an offer at the Bitcoin Forum to pay 10,000 BTC for a US20$ worth pizza.

As decentralized encrypted currencies concept spread out, numerous rival projects hatched out. Alternatives to Bitcoin, or altcoins, are usually intended to improve the original Bitcoin protocol and offer the users some extra advantages like faster transactions, higher grade of anonymity, or empowered security.

In 2011, about two years after Bitcoin, Namecoin, the first altcoin, got introduced to the world. The core goal was to alternate the domain name system in a decentralized manner. It used a dedicated Firefox and Chrome plugin for access to ".bit" ending websites, that would automatically take the user to the proper location indicated by the registry stored on Namecoin. To confirm and retain the domain in Namecoin, users were supposed to send a transaction to the Namecoin network.

Litecoin was the next Bitcoin's heir, released in 2011 sometime later, and kept the rank-2 position among all the cryptocurrencies, right below Bitcoin, for several further years. Litecoin's key distinctive feature against Bitcoin lies in the so-called mining-puzzle.

Ripple (release date in 2012) got acknowledged for a "consensus ledger" system for the game-changing speed boost to transaction confirmation and blockchain creation period (the time goal is not formally set, but the average sticks to every few seconds). Another Ripple's distinction is its easier conversion compared to other cryptocurrencies, with an in-house currency exchange that can convert Ripple tokens into USD, Japanese Yen, Euros, and other global currencies.

Launched in 2015, Ethereum is an altcoin, that stands upon significant improvement of Bitcoin's basic architecture. The approach includes utilizing "smart contracts" that empower the performance of particular transactions, restrain the parties from going back on the agreements, and provide refund mechanisms in case of agreement violation.

Starting from 2015 the amount of new coming cryptocurrencies began to snowball, and by November 3rd, 2020, according to CoinMarketCap[1] records, reached the total number of 7,583 with a total market cap of $141 634 371 663 against Bitcoin's $254,661,588,035.

The advantages of cryptocurrencies include:

Low transaction fees. Cryptocurrencies utilize peer-to-peer transactions, which

[1] The world's most-referenced price-tracking website for cryptoassets https://coinmarketcap.com/

naturally eliminate brokerage or intermediary charges. As a result, the transfer cost decreases, that are otherwise standardized irrespective to the sender's and recipient's location. A standard transaction within Ripple protocol costs around 0.00001 XRP at on XRP token rate around $0.25 (as by October 2020). For average users, the transaction fee is so low it's almost assumed as free. Scaled to multiple monthly or daily transactions, the cost advantage especially overwhelms the attractiveness of traditional transaction systems. Plus, the advantage for travelers being able to use the same rate in every part of the world and avoiding extra exchange fees.

Instant Payments. Besides higher transaction costs, traditional systems suffer from unjustifiably stretched processing time because of procedure and bureaucracy matters. Cryptocurrencies, in contrast, provide nearly instant, 24/7 and free from holidays transactions.

Accessibility. Despite the whole globalization, transacting across regions and borders within the traditional financials systems remains complicated. With the advantage of decentralization, cryptocurrencies manage to equalize the access to everyone regardless of dissimilar boundaries. Since users only need a smartphone or a computer connected to the internet to send and receive cryptocurrencies, it now theoretically seems the one viable solution for populations previously fenced off the services of traditional banking, credit cards and other orthodox payment methods.

High level of personal data protection. Phishing and other fraud risks during traditional online transaction, like credit card payments in online stores, have been a considerable problem for a lot of users. This has been a concern for both clients exposed to phishing threat, and the sellers at risk of cyber attack and losing corporate data. Cryptocurrencies, on opposite, require no personal data reveal to run a transaction. This is possible thanks to using the two-key, private and public, approach. The public one (i.e. the BTC wallet address) is commonly accessible, while the private one is only known to the owner. Each transaction creates the evidence of being performed by the wallet owner by getting signed by interacting the private keys of both parties, and then sealed applying a mathematical function.

Risks involved by cryptocurrency transactions:

High price volatility. According to Liu and Serletis (2019), cryptocurrencies, compared to traditional asset classes, incur more than five times higher volatility, with a standard deviation of 5%. This is caused by an array of drivers. First, small market size: the crypto market remains an emerging one. At it's peak in 2018 it was valued $800 billion, which appears insignificant on the background of the US stock market valued $28 trillion, or the gold market at $7.9 trillion. The smaller the market, the higher impact may be caused by single forces operating within it. Second, no entry barriers, which results in millions of amateur traders. An average cryptocurrency holder has by far poorer education and experience, than a traditional stock market trader. This makes cryptocurrency dimension is vincible to social and psychological pathologies like "fear, uncertainty and doubt" or the "fear of missing out". In circumstances where an experienced trader would just stand by, a crypto investor shows higher chance to tush right off the bat. Third, the lack of regulation. Absence of supreme regulatory mechanism lures numerous tunned in players having enough experience to successfully manipulate trade volumes and abuse collaborative dump and pump schemes, which results in the ecosystem falling in crowd panic or euphoria, finally leading to even higher volatility

Possible use for criminal purposes. Anonymity and promise of surplus profit from the crypto market evoked pressing regulatory challenges. Cryptocurrencies have been used for illicit drugs, weapons, illegal pornography trafficking and even contract killing. They've opened new pathways to fund terrorism, doing money laundry and tax evading. Totally digital and anonymous cryptocurrencies have undoubtedly facilitated the extension of "darknet" marketplaces selling illegal wares and services. As by April 2017, there was estimated to be 27 million Bitcoin holders using it primarily for illegal deals – 37 million transactions annually worth around $76 billion, holding nearly $7 billion worth of Bitcoin put together (Foley et al., 2019). A striking example of crypto funding terrorism is the case of Al-Qassam Brigades Campaign. In early 2019 Al-Qassam Brigades used their social media account to post a fundraising call appealing the allies to donate Bitcoins for a support to their campaign. The request was later reposted on their three official websites. Emphasizing, that the donations are fully untraceable, Al-Qassam encouraged the followers to back their acts of violence, and even posted a video tutorial on making anonymous donations and using unique Bitcoin addresses for each backer. The fundraising, however, did not go successfully as promised, as the US official managed to capture 150 accounts used for the fund laundering[2], seize the Al-Qassam site network and use one to bait and decoy the donators.

Cryptocurrency Crimes. According to "Spring 2020 Cryptocurrency Crime and Anti-Money Laundering Report" by CipherTrace, crypto thefts, hacks, and frauds totaled

[2] Acoording to the report of the United States Department of Justice
https://www.justice.gov/opa/pr/global-disruption-three-terror-finance-cyber-enabled-campaigns

$1.36 billion just in the first five months of 2020. This means 2020 could be the year of the highest amount stolen in crypto crimes outside 2019's $4.5 billion. By the example of 2019 and so far on, fraud and unlawful appropriation keep the biggest stake of the year's stolen crypto compared to hacks and thefts. Fraud and misappropriation count in nearly $1.3 billion – 98% of the total $1.36 billion. As a comparison, in 2018 hackers took 56%, $950 million out of total $1.7 billion. Wallets, exchanges and all the rest cryptocurrency custody services maintain continuous defense improvement, yet malicious users keep finding their way, innovating and sometimes even outrunning the current cybersecurity paragons. A lot of breachers, for instance, abuse blended attacks employing multiple techniques, such as SIM swapping, URL hijacking, phishing, simultaneously against multiple targets, this way snapping off user and admin accounts with the assistance of an impostor. For example, in May 2019, Binance, the world's largest cryptocurrency exchange platform headquartered in Malta, got held up by $40 million in crypto assets after an intricate hacker's attack using a grim mixture of viruses, phishing and several other siege weaponises.

Negative impact of mining. Crypto mining consumes excessive amounts of computational power and electricity. Krause and Tolaymat (2018) calculated, that BTC, ETH, LTC and XMR mining require even more energy to produce and equivalent market energy, than traditional metal (copper, gold, platinum) and rare earth minerals mining (except for aluminum because of its high electricity expenditure). Moreover, BTC mining alone consumed in 2017 more electricity that the entire country of Ireland. This level of consumption causes malignant social externalities – mostly due to the climate change and negative people's health impact owing to the emission from burning fossil fuels. This way, in 2018 each $1 worth of Bitcoin mined caused $0.49 and $0.37 value of health and climate pressure in the US and China correspondingly (Goodkind et al., 2020).

No system to resolve the dispute between the parties. No central authorities make it impossible to create a relevant system of that kind. In case of a dispute or accidental sending coins to a wrong recipient by mistake, the wallet owner has no way to retrieve the funds. This drawback can be abused by malicious users to rip others off their money. Lack of refund procedures makes it too easy for someone to fall into a transaction with someone promising a product or service but never intending to supply or execute it. Naturally, the existing ecosystem does no imply the existence of a deposit guarantee system.

Despite all the pros and cons, cryptocurrencies happen to be an important chapter of the social history. The phenomenon of cryptocurrency went all the way through highs and lows, losing the credit and getting back high in value, solving security issues, improving return of income and ROI. After a massive crash in early 2018 followed by a long stagnation, the crypto community and investors still tend be optimistic in their expectations about the future.

2.2 Factors influencing cryptocurrency prices

The mechanism of cryptocurrency price formation has been subject to a lot of research, correlation seeking, and prediction attempts at times reminding somewhat fortune telling. In search for profit, crypto investors are reaching out for any possible mathematical models, algorithms and analysis approaches trying to figure out the exact factors that would give the right assumption whether and when a coin’s rate is going up or down.

Most commonly assumed factors that underpin the market include:

- supply and demand;
- technical and technological aspects;
- macroeconomic and financial market variables (stock markets, exchange rates, oil price, gold price, interest rates, etc.);
- sentiments of crypto community, fear and uncertainty in the markets;
- media exposure;
- political and legal issues.

The scientific community has not come to a general consensus, and the outcomes of different research appear contradictory, so the comprehensive, scientific set of cryptocurrency price determinants is yet to be drawn up.

Thus, Buchholz et al. (2012) found Bitcoin price to follow the nature of any other currency and its changes relate to the interaction between supply and demand. As the supply determines Bitcoins scarcity and the number of tokens circulating in the market, the demand is conditioned by the amount of transaction requests to apply Bitcoin as a payment method. This statement got later confirmed by Ciaian et al. (2016), saying Bitcoin price can be explained in terms of standard currency price formation model, since the token’s price shows straight dependence on supply and demand forces. In addition, assuming Bitcoin supply exogenous, it is sensible to list the demand-side drivers among the key Bitcoin price determinants. Moreover, the bigger scale Bitcoin economy obtains, the bigger magnitude supply and demand drivers perform on price formation. In other words, the wider Bitcoin spreads, the more impact supply and demand deal to the rates changes.

Sovbetov (2018) in search for price determining factors went further and dissected not only Bitcoin, but also four altcoins - Ethereum, Dash, Litecoin, and Monero based only weekly data between 2010-2018. The study employed autoregressive distributed lag model and reveals, among other things, a straight long and short-term impact of specific crypto-market factors - market beta, trading volume, and volatility, valid for all five studied cryptocurrencies.

However, Kristoufek (2013), claimed, that for digital currencies, searching for answers within standard economic theories is inherently doomed. In his opinion, there is no fundamentals to allow to fix a “fair” price of a digital currency, and the rate changes are driven by the investors’ sentiment, their belief od disbelief in a currency’s continuous growth. Thus, future cash-flows model, purchasing power parity, uncovered interest rate parity or other standard macroeconomic theories do not work for crypto market. The price of traditional currencies, just like any standard economic goods, is exposed to supply-demand interaction and depend on the issuer’s GDP, inflation, unemployment, interest rates and other macroeconomic variables. For digital currencies instead, the supply function is either fixed by pristine emission limit or sticks to an intentionally designed algorithm (like Bitcoin mining). As no “orthodox” economy lies under digital currencies and possessing tokens itself does not generate any profit as there is no interest rates, the demand is only driven by a rush for speculative profit. In other words, the market exists in a different dimension, disobeys the standard rules and is dominate by surplus profit seekers, speculators, trend chasers, short-term investors, as thus the player’s sentiment should be considered in the first place.

In line with Kristoufek, Bouoiyour and Selmi’s (2015) study, that, through the prism of Bitcoin attractiveness to investors, figures out the extent to which speculation, considering it’s prior to traditional forces like supply and demand, impacts Bitcoin price formation. By their estimation, around 20% of Bitcoin price is shaped by investors’ attitude, while 70% of 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.

In a later study (2015) Kristoufek moves on and dissects various price change drivers scoping fundamental, technical, and speculative sources, looking to define the time and scale, or frequency, related behavior oh the interconnection. Kristoufek then reveals, that contrary to the assumption about Bitcoin's merely speculative nature, its long-term price formation shows straight dependance on classic fundamental factors – usage in trade, money supply and price level, and follows the monetary economics and the quantity theory of money.

Hayes (2017), in his turn, looked deeper into the technical aspect of price formation matter, ending up with a consideration, that more than 84% of relative value formation can be explained by three variables: computational power (a term for mining difficulty), rate of coin production, and relative complexity of mining algorithm. This means, relative rates of production can be identified as prevailing determinant for mining effort. With a given hash power capacity, the more increases the difficulty, the less units the process yields, and the higher relative cost of production grows. The same happens when reducing block reward or implementing more rigorous algorithms – again, the yield on units decreases. In other words, the concept assumes, that relative production costs on the margin perform as a direct driver of the cryptocurrencies' value formation.

Li and Wang (2017) suggest, that in the short term, Bitcoin's exchange rate moves along in accordance with changes in economic fundamentals and market conditions. At the same time, long-term Bitcoin rate relies more on economic fundamentals, and, after Mt. Gox closed, appears more rigid against technological drivers. The researchers also distinguished Bitcoin's price determination to get more exposed to mining technology, while the impact of mining difficulty was fading.

In a study „What can be expected from the Bitcoin" van Wijk proves, that in the long run, common economic indicators, such as the Dow Jones index, EUR-USD relationship and oil price also bring in some significant contribution to Bitcoin rate formation. As opposite, Ciaian et al. (2016) calls these conclusions false claiming, that Dow Jones, exchange rates and oils prices are only relevant in the short term. According to him, once market forces of supply and demand or investment attractiveness are included into the model, the impact of global macro-financial development gets insignificant.

According to an analysis by Wang, Xue, and Liu (2016) Bitcoin price is shows negative relation with the stock price index and the oil price, but instead a positive relation with daily trading volume. This is mainly owing to investors willing more to make pull profit from the market once stock index goes up, so other investment targets lose in value. Market doldrums flowing from economic recession push investors to look more at hedging products, and on this basement Bitcoin price would ascend. Highly dependable on investors' behavior and somehow indicating their assumptions regarding inflation, oil price may indirectly affect Bitcoin price as well. To certain extent, the daily volume of Bitcoin trading may be taken as a mirror for investors' engagement into Bitcoin – the higher engagement pushes the price up, and vice versa.

On another side, Sovbetov (2018) suggests, that SP500 index shows weak positive long-run impact on Bitcoin, Ethereum, and Litecoin, turning the sign to negative once considered in the short term. The only exception is for Bitcoin that generates an estimate of -0.20 at 10% significance level.

Poyser (2017) uses Bayesian Structural Time Series Approach to explore Bitcoin's market price correlation with a set of internal and external drivers. Poyser ends up finding Bitcoin to perform negative association with neutral player's sentiment, gold price and CNY to USD ratio. On the other hand, a positive relation appears with stock market index and USD to EUR ratio. In summary, Poyser defines Bitcoin as an entity of mixed properties retaining speculative aspect, safe haven, and potential capital flights instrument features.

Kristoufek (2015), though, insists on the opposite, saying there's "No sign that the Bitcoin is a safe haven, which is in fact expected considering the present behavior and (in)stability of prices". His method relies on using a kind of safe haven testing by examining the Bitcoin's relationship with the Financial Stress Index (FSI) and CHF gold price. FSI turned out to show only one time period of relevant correlation with Bitcoin price – the exact period of Cypriot crisis with most simultaneous movements sticking to an around 30 days scale. Within this interval of time, growth of FSI pushes Bitcoin up, yet beyond Cypriot crisis there wasn't detected any other long-term time sections of statistically significant and reliable correlations. In the view of gold price, the viable relationship does not appear, except for two around 60-days islands. This, however, most likely comes form gold price dynamics itself. The first island coincides with the gold's price skyrocketing around September 2011, and the second one – with its continuous decrease. Apart from proving Bitcoin not been closely connected to gold's price metamorphoses, Kristoufek also questioned whether gold could still be assumed a safe haven as it used to be.

Rodrigo Hakim das Neves (2020) in his turn, claims, that on top of global crises Bitcoin tends to go up in price, as it becomes a desired alternative investment. Thus, financial markets could still use Bitcoin as a safe haven and take advantage of its built-in properties to help investors or authorities obtain alternative gear for monetary transactions matters.

Smith (2016) on the other hand, suggests considering and treating Bitcoin rather as digital gold. In his perception, Bitcoin is somewhat mirroring the relationship between gold and conventional nominal exchange rates. Despite high volatility of Bitcoin's nominal price and its ephemeral correlation with other nominal exchange rates, Bitcoin's relative price shows significant, relevant mutuality towards conventional market exchange rates. Relative Bitcoin prices, ran by arbitrage momentum, quickly shapeshift to wager for the parity with market exchange rates. This is, however, a one-way street correlation, as market exchange rates do not demonstrate any reaction to Bitcoin price changes. In fact, nearly half the changes in relative Bitcoin price can be justified by this relationship. Evidently, floating nominal exchange rates determine a majority of the Bitcoin market price volatility – same as for traditional markets.

Liu and Tsyvinski (2018) tried to figure out if there was a resemblance between cryptocurrency pricing and stocks, and, by their sample, risk factors that drive movements in stock prices do not seem to apply to cryptocurrencies. Traditional macroeconomic factors, exchange rates variations or general commodity prices, that apply for conventional assets show at best marginal influence upon cryptocurrencies. Liu and Tsyvinski conclude, that only crypto-market specific drivers (profit momentum, instantaneous attractiveness) are relevantly useful to trace the returns on cryptocurrencies' variations. These conclusions fall in with Polasik et al. (2015) who claimed any similarity between Bitcoin returns and conventional currencies or global macroeconomic drivers statistically insignificant and insufficient.

Besides, in a working paper entitled "Cryptocurrencies as an Asset Class: An Empirical Assessment" by Bianchi dissecting trading patterns of 14 largest cryptocurrencies (including Bitcoin) from April 2016 to September 2017 on a weekly basis, no correlation was revealed with any economic driver to be referred to by investors when making their decisions. The utter conclusion was, apparently, that it's past returns, sentiment, and hype in the crypto community, which entirely determine whether the price would drop or leap.

All this makes quite enough sense. In the age of instant access to almost any information from any spot on the world's map, with almost every citizen online 24/7, the impact of mass media, forums and opinion shapers is tremendous. A couple of web posts from journalists or influencers may blow a wave able to define the howling success or a fiasco of a single product, a start-up, or an entire corporation – with no exception for cryptocurrencies. Those last ones, actually, appear the most exposed to the ado and hype pumped by media. In the first place, this vulnerability flows from lack of governmental regulation, so the estimation of a cryptocurrency's value is shaped based whether the vibes around it in the news, blogs and forums are positive or negative. Cryptocurrencies are projects run by groups of people sowed across the globe, grounded by no relevant law or physical resource, so the movements of their prices are determined by public moods, hype, popularity, or mainstream.

This concept finds confirmation in several studies:

- Bitcoin price depends how engaged the investor get about the cryptocurrency. In the long run the relationship appears obvious, yet prices keep skyrocketing on periods of rush, and fall even deeper in the phases of rapid decline. (Kristoufek, 2015)

- on top of speculative nature of crypto markets, public moods, sentiment, and newsbreaks bring in undeniable, strong impact on the Bitcoin price formation. (Kaya,2018)

- there is plenty of evidence among both long and short-term attractiveness factors revealing hopping mass excitement and interest in Bitcoin to be in most cases a consequence of its anterior price bounce. Conversely, as demonstrated by the case of Bitcoin's collapse, a currency's image misfortune draws the price down. (Neves, 2020)

- the attractiveness of cryptocurrencies is crucial for most subject tokens (Bitcoin, Etherium, Monero, Litecoin) except for Dash, but strictly in the long-term matter. (Sovbetov,2018)

- the sentiment around social media is a viable indicator of for Bitcoins upcoming price movements assumptions. However, different publications exert uneven influence. (Mai et al., 2018)

- phases decadent attitude – the so-called Fear, Uncertainty and Doubt (FUD) periods – are considerable drivers of market uncertainty, which diminish the assumed value of cryptocurrencies and therefore drag their prices down. (Civitarese & Menders, 2018)

A curious research was presented by Viglione (2015), that utilizes an alternative approach to price drivers to reveal the inverse nature of Bitcoin price hops relation to economic freedom. According to Viglione, in low economic freedom ecosystems investors are forces to pay higher premium above global prices that already comprise trading volumes, bid-ask spreads and other microstructure differences. Apparently, investors in countries with more rigorous capital control and higher taxation are willing to pay more for the opportunity to diversify financial assets internationally at lower cost, given by Bitcoin, compared to investors from environments with softer taxation policies.

The problem of regulation is another driver that has a significant impact on the price formation of cryptocurrencies.

Despite a common assumption about cryptocurrencies' functioning beyond the reach of authorities, new information or announcements regarding even potential regulatory measures turn to significantly affect transaction volumes and users' behavior. The impact varies depending on what kind of regulations are about to come. The most painful outcome is carried through general prohibitions regarding cryptocurrencies or the changes in the way securities law treats them. Then follow the news on empowering policies against money laundering policies, black market or those concerning cryptocurrency interflow with other markets under regulation. In addition, news about authorities plotting to set up specific framework aimed at cryptocurrencies and initial coin offerings come along with notable market rise ups. In view of these conclusions, the functioning of cryptocurrency markets seem to undergo the influence of regulated financial institutions, making cryptocurrencies exposed to national regulatory organs' sticky fingers. (Auer & Claessens, 2018)

The same line follows „Taming the blockchain beast? Regulatory implications for the cryptocurrency Market" by Shanaev et al.(2020). It reveals economically and statistically significant disturbance to the crypto market exerted by news and announcements about stiffening of regulatory policies, anti-laundering measures, exchange and issuing restrictions, or even launching government-approved cryptocurrencies. Policy liberalization or rejection of state-backed blockchain payment projects, on the contrary, results in digital coins prices starting to move up. In general, Shanaev's key takeaways suggest, that, at least at the current phase of cryptocurrency and blockchain social evolution, redundant administrative control is destructive, while authorities' commitment to let the development of cryptocurrencies run its own course developing in their natural "sandbox" ecosystem, in the only way to let it develop properly. In other words, crypto markets would have a much easier job getting over high volatility and token price instability issues, once governments agree to refuse overregulating and let the industry move forward freely. In addition, investors might take some advantage from this paper, as it offers a comprehensive overview on how to properly adjust the strategy utilizing the most out of analyzing news and crypto world sentiment.

Thereby, even despite quite a substantial volume of existing research concerning cryptocurrency price formation, the matter remains open to further exploration and keeps plenty of room for new research works.

3 Data and methodology

This chapter defines the research methodology and elaborates data collection for the variables under consideration, including a review of media data sources and Reddit data sentiment analysis.

3.1 Data collection

3.1.1 Historical prices of cryptocurrencies

Since Bitcoin was the first, most recognized and most valued cryptocurrency, previous academic research on intercorrelation between cryptocurrencies and market sentiment has been focused on this coin almost solely (e.g. Kristoufek, 2015; Ciaian et al., 2016; Mai et al., 2015). As other currencies evolved over the years, academists started to take them into account alongside Bitcoin, too. Ethereum, the world's first, best known altcoin of highest market cap, became Bitcoin's first heir in view of researcher's consideration (e.g. Wooley et al., 2019; Abraham et al., 2018). Although, nowadays more and more minor altcoins become subject to analysis, like ZClassic (Li et al., 2019), BitCoinDark, Voxels and PureVidz (Steinert & Herff, 2018) or Qora and MintCoin (Ciaian et al., 2018).

This work is no exception for that emerging trend. The research is built around Bitcoin as the core crypto market unit, and Tezos, which used to be omitted by earlier researchers.

Tezos is a decentralized, open-source blockchain network that can execute peer-to-peer transactions and serve as a platform for deploying smart contracts[3]. The initial token for Tezos blockchain is known as tez or tezzie, abbreviated as XTZ.

One of Tezos's key competitive advantages consists in its on-chain governance model, which allows the blockchain to incorporate changes automatically and avoid hard fork (dividing into two cryptos). Otherwise, as it happened to Ethereum and Ethereum classic, hard fork may cause contention or even enforce a cryptocurrency's split into two counterparts. Another crucial feature of Tezos consists in the Proof of Stake (PoS) consensus mechanism. PoS allows users to confirm transactions (and earn a payoff out

[3] According to information from the official website of the project https://tezos.com/

of it) without mining effort, that makes an extra burden with its expensive hardware and technical proficiency requirements.

Tezos was first introduced in a whitepaper issued in 2014, though managed to draw the public's attention yet in 2017 by sourcing $232 million during the ICO, thus becoming a contemporary most appreciated ICO in history, setting up a milestone for the entire crypto world to refer to. In September 2020, Société Générale-Forge, a spin-off technological start-up of the French investment bank Société Générale-Forge, used Tezos blockchain as the core for its experimental project on a Central Bank Digital Currency (CBDC)[4]. The company got authorized by the French central bank, the Banque de France, to elaborate digital euro.

By October 2020, XTZ reached $1.65 billion of market capitalization, which earned it the 17th place in the CoinMarketCap[5] most valuable cryptocurrency ranking.

In terms of the research, historical data scopes the time period between August 1st 2018 and October 31st 2020.

An important feature of the cryptocurrency market is the fact, that prices across various exchanges may significantly differ. Of the main reasons for this is that a standard or global price does not exist at any point in time. The price is neither bound to the USD or other conventional currencies, nor to a particular region, country, or marketplace. Following the general rule for all types of commodities, supply and demand vary across different time periods and markets, affects cryptocurrency prices fluctuations as well. Liquidity is another important matter. For instance, larger exchanges, like Binance and BitForex, generate higher trading volumes of Bitcoin compared to smaller ones. As a result, the difference in supply leads to different price levels across the exchanges. Finally, the fees. Most marketplaces charge their users with some sort of commissions for provided services and transaction, which also contributes to the inaccuracy in valuating the final trading price.

BitInfoCharts[6] estimates average cryptocurrencies' prices by gathering data from a diverse pool of sources, also reaching for peer-to-peer platforms. By this means, this

[4] https://blockchain.news/news/tezos-blockchain-chosen-french-digital-euro-societe-generale-forge/

[5] https://coinmarketcap.com/

[6] https://bitinfocharts.com/

platform appears as a relevant resource for gathering relatively accurate general value figures with no attachment to any particular exchange.

3.1.2 Google Trends data

Google Trends is a tool, that offers access to a bulk pool of Google search engine incoming search requests since 2004. The data set offered by Google Trends is anonymized (no identity behind the request is revealed), categorized by the request subject, and aggregated (grouped by sheared feature). This allows to sort out the volume and rate of interest about particular subjects by the scale of the entire globe or down to a single town. Google Trends, though, does not offer absolute search values, but instead generalizes search data for the purpose of easier comparison between terms. The search volume index is calculated by dividing each data point by the total search requests within a geographic region and time period[7]. The numbers are then scaled between 0 and 100 on a search term's proportion to all searches on all topics. The level of granularity of the returned data is bound to the designated historical time interval: daily search volumes are returned for queries under 90 days and weekly search volumes for queries longer than 90 days.

Exploring the correlation between cryptocurrencies' price rates and google trends on time periods that exceed 3 months, some researchers, like Kristoufek (2013), apply weekly data spans, or transform daily data by combining it with weekly data for long time (Phillips & Gorse, 2018; Garcia, 2014; Abraham et al., 2018). For the purposes of this paper, it was decided to use daily data, yet retrieved not by direct reconstructing, but instead extracted using the Python's pytrends package (Pseudo API for Google Trends) which offers a simple interface for automized downloading Google Trends reports.

This work considers search terms by the English names of each cryptocurrency ('Bitcoin', 'Tezos'), omitting the abbreviations. Besides, it only considers global data with no regional binding.

3.1.3 Reddit data

Reddit is a public network for news aggregation, content sharing and rating, and open discussion. According to Alexa Internet[8], as by October 2020, Reddit ranks as 17th in the world and 7th in the US most visited website, with 40.7% of its total community

[7] FAQ About Google Trends Data by Google support
https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052/

[8] Alexa Top 500 Global Sites https://www.alexa.com/topsites

settled in the United States. As follows from the company's official report released at the end of 2019, Reddit counts more than 430 million active users monthly.

Reddit's registered users are enabled to share content in the form links, text, image of video, offering other users to rate it positively or negatively by "upvoting" or "downvoting", comment the posts themselves or the comments left by other users unfolding discussion threads within a single post domain. Posts within Reddit's ecosystem are sorted by subject into so-called "subreddits", boards created by users. Since the very foundation of Reddit in June 2005, the amount of appearing subreddits has been sustainably growing, and has by now reached the number of over 2,2 million. Subreddits may as well refer to general topics like tech, gaming, or environment or specific, niche subjects (down to a particular product or event), letting the users find desired interaction on their personal points of interest. The internal hierarchy of Reedit implies particular submissions to move up the subreddit's feed as they gather more upvotes, and even appear on the platform's frontpage once the number of upvotes reaches a designated hurdle. Recently, using Reddit as a source to extract data from for purposes of cryptocurrencies studies has become a common trend (e.g. Wooley et al., 2019; Bremmer, 2018; Phillips, 2019; Salač, 2019). This happened for an array of reasons:

- Post length. While Twitter limits the posts to 280 symbols, Reddit allows to 40,000 characters for posts and 10000 for comments, which allows the present the thoughts much clearer and therefore avoid ambiguity while conducting sentiment analysis.

- Content format. Text is much easier to analyze, than images or videos. While Twitter and Facebook allow to post visual content without any copy (actually, images on Twitter generate 150% engagement than text messages), Reddit requires every post to contain a textual part, this way unintendedly aiding data mining and processing.

- Content hygiene. Reddit is seen to be less exposed to cryptocurrency-related spam and flooding thanks to strict moderation cutting off off-topic discussion, which naturally leads to higher proportion of quality content.

- Topic-based structure. Most cryptocurrencies have their own subreddit, so data miners have part of the job initially done by the platform.

- Data availability. Compared to social media networks like Twitter or Facebook, Reddit is relatively open to data acquisition.

This study scopes Reddit posts and comments form 2 subreddits: r/bitcoin (1,730,472 members by, as by 31.10.2020) and r/tezos (30,115 backers correspondingly) withing time period between 1.08.2018 and 31.10.2020.

The study also required to involve https://pushshift.io/ website for data collection, since Reddit had limited the time frames for gathering post and comment data via its API. Pushshift is a platform meant to supply researchers with proper access to social media data collection, archiving and analyses, and contains Reddit data reaching back to 2015. In addition, Pushshift includes Reddit's historical data from the moment of inception and provides real-rime dataset updates. Utilizing Pushshift allowed to obtain topic IDs within the designated time period, and then export json files containing post titles, bodies and related comments by direct requesting Reddit API.

3.1.4 Wikipedia views

Wikipedia is the world's 13th most visited website (as for October 2020) and one of the most popular collaborative knowledge web bases[9]. Wikipedia views, however, are scarcely considered within cryptocurrencies-related studies (e.g. Kristoufek, 2013; Dickerson, 2018; Phillips & Gorse, 2018; ElBahrawy et al., 2019). Nonetheless, the numbers of daily views on Wikipedia pages can be a useful indicator to estimate the overall attention of the Internet community (Yoshida et al., 2015), and monitoring these numbers could provide relevant tracking for new users looking for knowledge on cryptocurrencies, and extra insights on other online drivers, focusing on users with less expertise in the first place (Phillips & Gorse, 2018). Thus, this kind of data also seems relevant to be included into this study.

Since Wikipedia does not enable direct access to analytical data, external sources of data extraction are required. Currently, the approach to collecting daily data for Wikipedia views depends on the explored time period. The most applicable approach for period after 1.07.2015 is using the official mwviews Python library, that connects to Wikipedia's page through API. Earlier data can be extracted by one-month packets from the http://stats.grok.de/ website. In terms of this work using data from 2018-2020, the mwviews packet would be ample. The views data has been gathered for two English pages: 'Bitcoin' and 'Tezos'.

3.1.5 Twitter data

Twitter is one of the most recognizable brands in todays tech and media environments. As of the second quarter of 2020 Twitter counted 186 million monetizable daily active users, in contrast to 139 million at the same period of 2019. Above that, Twitter notes

[9] As stated in the already mentioned rating Alexa https://www.alexa.com/topsites

about 500 million guest or not logged-in users visiting the platform every month[10]. The brand Twitter stands for an online social network and microblogging platform built around users sharing text-based messages or visual content units called "tweets". In 2017 the initial 140-digits text post limit got extended to 280 characters. The short format of the tweets became Twitters hallmark, promoting informal collaboration and rapid information exchange. For example, companies find good use of Twitter for publicizing corporate product or CSR news and announcements, interact with the customers, and monitor the brand's image in the eyes of the community. For easier navigation Twitter uses metatags starting with # symbol called hashtags.

In this means, Twitter appears to contain plenty of real-time information about community trends. No wonder then, that Twitter is commonly used as a source of social indicators for the purpose of predicting the movements of financial assets, now as well including the cryptocurrency markets. In terms of that researches may use both straight text processing following up with sentiment analysis (e.g. Stenqvist & Lönnö, 2017; Li et al., 2019; Valencia et al., 2019), and analyzing the amounts of those texts (e.g. Matta et al., 2015; Abraham et al., 2018; Kraaijeveld & Smedt, 2020).

This study only considers as inputs the number of tweets per day collected directly from BitInfoCharts.

3.1.6 Telegram data

Telegram is a cloud-based instant messaging service, that allows users to share text and voice messages, multimedia content, and make voice or video calls. In April 2020 Telegram's dev team officially reported the app hitting 400 million active users and being among top 10 most downloaded applications in the second half of the year (as by Sensor Tower data)[11]. Besides global acknowledgement and sustainable growth of active user base, Telegram also took over the space of messaging services for cryptocurrency enthusiasts. One of the basic reasons for this is Telegram's highest anonymity compared to other popular messengers. Users take advantage of its end-to-end encryption option, which fences their privacy off the reach of government, regulators or other institutions looking to get access to people's private information exchange. In the light dingy policies and actions of numerous governments across

[10] As reported by Statista - a German company specializing in market and consumer data. https://www.statista.com/topics/737/twitter/

[11] https://telegram.org/blog/400-million/ru?setln=en/

multiple regions, users engaged in trading activities have been struggling for assured privacy they've finally found in Telegram.

Another reason of Telegram's popularity is how easy groups and channels can be created and operated. A Telegram group can gather up to 100.000 users able to send messages quickly and safely. Telegram became home to a legion of free or paid channels and chats sharing news and alerts within their closed communities. The admission charged by some closed Telegram groups and channels reaches even $5,000. Assuming the fee as an investment promising capital multiplication, plenty of users are willing to pay for that access.

Thus, some recent studies already include Telegram along with other sources (e.g. Smuts, 2019; Hamrick et al., 2019; Mirtaheri et al., 2019).

This work is using the indicator of daily amount of references to 'Bitcoin'/'BTC' and 'Tezos'/'XTZ' cryptocurrencies across open chats and communities on daily basis regardless of geographical location. The data sourced from a web-portal Telegram Analytics[12] - a project with statistical data on more than 470 000 Telegram channels and 2.5 billion posts in eight languages, through direct requests to the right holders.

3.1.7 News

In the mediocre opinion, cryptocurrency market is assumed to depend on the news background like no other market, being extremely vulnerable to its drivers. However, existing research on that issue show controversial findings. Thereby, Lamon et al. (2017) and Yao et al.(2019) suggested Bitcoin price to show reaction to news articles. Inan(2018) on the contrary, concluded, that American news articles are unable to clearly explain or define the price changes, so that the prices are unpredictable by nature. Rognone et al.(2020) lean to a mediate opinion, saying there exists a sort of Bitcoin users' optimism, meaning that only positive news seen to boost Bitcoin returns, with the investors tending to ignore negative ones. By this means, considering news as another price change driver appears applicable in terms of the study.

The amount of news articles containing references to Bitcoin and Tezos was received from Omenics[13], a cryptocurrency data analytics platform designated for aggregating, analyzing the information, and therefore mapping out price and sentiment trends. "Investors in cryptocurrencies are overwhelmed by information," says Omenics' co-

[12] https://tgstat.com/

[13] https://omenics.com/

founder Pierre Alexandre Picard, “we want to leverage the abundance of data in the crypto market to track and, one-day, be able to anticipate market changes.” The platform aggregates data from more than 100 crypto news sites, both showing nominal quantity of news by days, and displaying the articles straight within the portal simply by clicking on the desired date.

3.2 Sentiment analysis of Reddit data

Posts and comments themselves connote an unstructured, unorganized data set. In specie it is a random bundle of information from web users expressing their feeling, views, emotions or sharing their experience, so is of poor relevance in terms of further processing, analysis, and prediction.

Sentiment analysis (or opinion mining) is designated to solve this challenge, as one of the most applied field in Natural Language Processing (NLP) and transforms bulk, unstructured text batches into structured, quantitative essence of the sentimental frame of mind contained within the text mass (Ma, 2020).

Hereby, preprocessing data is necessary in terms of defining the efficiency of the following phases of sentiment analysis. Json files mentioned in section 2.1.2 were converted using Power Query into two .xlsx files (one for each cryptocurrency) with 2 sheets. Sheet 1 contained the information on topics, and sheet 2 – for comments. Columns “created_utc” (containing information on the post/comment creation date), “title” (post headline), “selfetext” (the post’s body) and “body” (comment text) were applied in further analysis. Irrelevant columns were eliminated. Post headline and body were merged into one column “Text”. Next, the data from sheet 2 got merged into sheet one (column “body” renamed as “Text”), along with creating an additional column containing the information about the type of the text (post or comment). This way, the data was finally generated and converted into three columns titled “Date”, “Type” and “Text”. The date column also got converted from unix into an easier readable datetime format. In the resulting file, empty rows got erased, as well as comments presented as “[removed]”, meaning, that the comment had been deleted by a moderator, the AutoModerator, or a spam filter, and the “[deleted]” ones, meaning the had been removed by the user himself, as those do not carry any meaning and we cannot identify what was in there, but the score received from them might affect the final result. Comments left beyond the explored time period were also removed from the data set.

In the outcome, two data sets got shaped – one on Tezos containing 38,489 rows (10,318 posts and 28,171 comments) and the other one on Bitcoin of 188,500 rows (20,934 posts and 167,566 comments).

For the actual sentiment analysis of gathered data it was chosen to utilize VADER. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool. This tool is deliberately harnessed to sentiments expressed in social media, which is generally full of colloquial and informal expressions, emotional abbreviated expressions and finally emojis common for social media posts (Hutto & Gilbert, 2014). VADER software is fully open sourced under the [MIT License].

The lexicon approach implies an algorithm building up a dictionary, which is a comprehensive word index of sentiment features. All the features were rated by 10 independent human graders in terms of polarity and intensity to a scale between “-4: Extremely Negative” and “+4 Extremely Positive”, and then marked in the dictionary by average score. The word “rejoice”, for example, is rated +2.0, the word “tragedy” is alleged highly negative and rated -3.4, and the “:(“ frowning emotion rated -2.2. In total, VADER’s lexicon dictionary counts around 7,500 sentiment features. Any word not found in the base scores 0, which means “neutral”.

Some originally neutral structures, like “not” or “but”, may invert the sentiment’s polarity, or alternate the entire intensity of a sentence, as it happens with the words “very” or “extremely”. VADER’s developers also implement a dedicated set of heuristic rules, that maintain the matters of contrastive conjunctions, adverbs, punctuation, capitalization, and so on.

For sentimental scoring of a text in the whole, VADER examined the text for defined familiar features, adjusted the intensity to the rules, calculated the score of detected features and generalized the final score to (-1, 1). Typical threshold values presented in Table 3.1.

Table 3.1: Standardized thresholds values for VADER sentiment scores

Compound score	Sentiment
greater than or equal to 0.05	positive
from -0.05 to 0.05	neutral
less than or equal to -0.05	negative

Source: own work based on Hutto & Gilbert (2014).

Apart from the sentence’s compound score, VADER also outputs the percent ration of positive, negative, and neutral features.

The advantages of VADER over approaches include:

- Showing high performance on social media type text, additionally generalizing to multiple domains with alacrity
- Requiring no practice data, as developed from a generalizable, valence-based touchstone sentiment lexicon curated by human power
- Performing high computational speed appropriate for cloud computing with streaming data
- Showing little vulnerability to the compromise between speed and performance.

All these advantages incline researchers to widely apply this method, for cryptocurrency analysis as well (e.g. Perry-Carrera, 2018; Kraaijeveld & Smedt, 2020; Valencia et al., 2019; Steinert & Herff, 2018; Stenqvist & Lönnö, 2017; Salač, 2019; Wooley et al., 2019).

This study runs the analysis through applying VADER in the "nltk.sentiment" Python library to the gathered data sets. After scoring each post and comment, resulting compound scores were placed in a separate column with all values labelled in accordance with typical threshold values. The outcome sample is presented in Table 3.2.

Table 3.2: Sentiment analysis sample

Date	Type	Text	Neg	Neu	Pos	Compound	Label
2018-09-02	comment	market is too down! :-(	0.411	0.589	0.000	-0.4199	Neg
2018-09-08	comment	This is amazing. Please bring schnorr signatur...	0.000	0.547	0.453	0.8442	Pos
2018-08-07	post	Selling a property for 20k in BTC. Wish me luck!	0.000	0.539	0.461	0.7177	Pos
2020-09-17	post	has bitcoins improved the world in anyway?	0.000	0.659	0.341	0.4767	Pos
2020-09-10	post	Belarus Nonprofit Helps Protestors With Bitcoi...	0.000	0.526	0.474	0.5423	Pos
2020-09-10	post	Where is my money going?	0.000	1.000	0.000	0.0000	Neu

Source: own work in Jupyter Notebook.

On completing the sentiment classification, each sentiment distribution was evaluated for each of the cryptocurrencies. Ass seen in Table 3.3, for both currencies about a half of processed posts and comments were positive, and around one-third – neutral. Tezos, though, seems to slope more positively.

Afterwards, the outcoming dataset was grouped by date, and the calculated compound scores of the days put into a separate column. Additionally, the total of posts and comments for each date was also separated to a new column

"Reddit_total_posts_plus_comments".

Table 3.3: Sentiment distribution of collected posts and comments

	Bitcoin		Tezos	
	Value	Percent	Value	Percent
Positive	81,750	43,37%	20,568	53,44%
Neutral	64,537	34,24%	11,771	30,58%
Negative	42,213	22,39%	6,150	15,98%

Finally, the columns containing daily average sentiment and total of posts/comments were moved to corresponding compiled data set for further analysis.

3.3 Methodology

3.3.1 Stationarity and unit root tests

Stationarity is among crucial concepts in time series analysis. Stationarity stands for statistical properties of a time series – autocorrelation, expectation, variance – staying unchanged over time (Shumway & Stoffer, 2017). In reality however, time series often contain trends, random walks, periodic fluctuations, etc., or even a combination of those, which essentially makes them non-stationary by mature. This occurrence is particularly typical for financial data. So, for financial models using non-stationary time series data ends up in receiving invalid outcomes contorting the picture and leading to false forecasting. By this means, most empirical time series studies nowadays require stationarity testing in the first place. Stationarity of a time series can be validated in several ways, such as summary statistics, looking at plots or statistical tests. Of those three, using statistical tests has proven the best in terms of verifying the data's stationarity. Most often used are KPSS stationarity test, DF, ADF and PP unit root test, or Breitung nonparametric unit root test (Zuo, 2019).

Sampling on the paper by Kristoufek (2013), this study runs stationarity testing using the Augmented Dickey-Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. The null hypothesis of ADF implies that a unit root against the alternative of no unit root. The null hypothesis of KPPS, in its turn, consists in stationarity against an alternative of a unit root. Using two tests in a tandem allows to reveal, whether the explored series is indeed stationary.

If the stationarity test comes out negative, the series can be transformed into stationary using several data transformation methods. According to Salles et al.(2019), (i) mapping-based and (ii) splitting-based are two main classes by which most researched transformation methods are looking to handle time series non-stationarity. Mapping-based methods usually derive new mapped representations of time series through one of three operations. (i) First basic mathematical transformation, that uses standard mathematical manipulation (percentage changes transform, moving average smoothing, logarithmic or box-cox transform). (ii) Second, detrending, which consists in removing a deterministic trend. Finally, (iii) differencing, which also implies trend removal, but does not involve the simple, fractional or season model estimation.

Splitting-based transformation also rely on methods capable of deriving new representation off the rime series data, applying various techniques to split a time series into a set of sub-series called component series. Component series then act as simplified, or even stationary, inputs for time series prediction methods for separated analysis and prediction. Among the examples of such transformations there are pattern-based (time series pattern mapping), moving average-based (KZ filter, KZA algorithm), time-frequency domain (EMD, VMD, wavelet transform, wavelet package transform, Hilbert-Huang transform) and frequency domain (KZFT algorithm, Fourier transform).

In the last few decades researchers have been tending to reach for splitting-based transformation methods, especially in bundle with computer intelligence techniques more and more frequently, although it until now remains underexplored.

Picking up the right transformation for adopted data model and the general problem, however, turns not that simple. In the first place, it requires thorough analysis of their features and potential advantages. Initial data assumptions (including linearity, seasonality, and non-stationarity varieties) and intrinsic properties, such as computational algorithm or mathematical transformation, should be among the considered features.

3.3.2 Pearson correlation

Pearson r correlation is a correlation statistic most commonly used for measuring the scale of interrelation for two linearly related continuous variables (Pearson, 1930).

Assume, that x and y are the quantitative measures of two random variables on the same sample of n. The formula for computing the sample Pearson’s correlation coefficient r is expressed as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$$

Are, correspondingly, the sample means of variable x and y. Otherwise stated, under an assumption, that x and y sample variances are positive, the linear correlation coefficient r might be expressed as the ratio of the sample covariance of the two variables to the outcome of their respective standard deviations s_x and s_y,

$$r = \frac{\mathrm{Cov}(x, y)}{s_x s_y}$$

Consequently, the correlation coefficient appears as a scaled instance of covariance. The range of the sample correlation measurement r is set between -1 and +1. In case of positive linear correlation between x and y (i.e., higher levels of both variables associate with one another), results $r > 0$, and results $r < 0$ occurs for negative linear correlation between x and y (i.e., higher levels of one variable entail the counterpart's lower levels). The value $r = 0$ stands for absent association, neither positive, nor negative. The direction of the association is reflected by the sign of the linear coefficient, and the intensity of the association is defined by the correlations' magnitude. Variables show perfect linear positive correlation at the point where the correlation coefficient equals +1. In that case, once one variable goes up, the second one follows it proportionally and in the same direction. Apparently, the coefficient of -1 stands for perfectly negative, or inverse, correlation, meaning that the variables are moving in contrary directions – while one variable goes up, the second one decreases proportionally. Zero correlation coefficient signals of no relationship existing between the variables. Besides, once two random variables X and Y are normally distributed, the population Pearson's correlation coefficient is represented as

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

Here σx and σy are the corresponding population standard deviations of X and Y. Table 3.4 represents the way to interpret the correlation coefficient's size (strength).

Evaluation of the received coefficients' significance involves the p-value, which stands for a probability of randomly finding these correlation measures just by chance. The smaller the p-value, the more confidence it gives about the virtual presence of a relationship instead of a randomly rolled result. For common acceptance, the significance level is set at 5%, which means just five percent probability of getting a randomly rolled result from the sample. Although, the existence of the correlation itself is yet insufficient to subtract the presence of a casual relationship.

Table 3.4: Interpretation of Pearson's correlation coefficient

Size of Correlation	Interpretation
1.0 (or -1.0)	Perfect correlation
From 0.7 to 1.0 (or from -0.7 to -1.0)	High positive (negative) correlation
From 0.3 to 0.7 (or from -0.3 to -0.7)	Moderate positive (negative) correlation
From 0 to 0.3 (or from 0 to -0.3)	Low positive (negative) correlation
0	No correlation

Source: own work based on Statistic Solutions (2019).

3.3.3 Granger causality

Granger causality analysis applies for better understanding in terms of the causalities while going beyond the correlations. Granger (1969) approach looks to test whether Y variable is caused by X variable through checking out how much of previous values of Y reflect upon the current Y, and then adding lagged values of X to see if it makes the explanation more comprehensive. If the coefficients on lagged X's turn out statistically significant or if X shows its use for predicting further Y values, then Y is considered Granger-caused by X. The causation often acts in two-way manner with X- and Y-Granger causing Y and X correspondingly. Notably, Y's Granger-causation by X, though, does not stand for Y being affected by X. Granger causality itself is incapable of indicating the causality in its common meaning, yet is meant to just measure precedence and information content.

Foe Granger causality analysis Eviews uses the following bivariate regressions:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_l y_{t-l} + \beta_1 x_{t-1} + \dots + \beta_l x_{-l} + \varepsilon_t$$

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \dots + \alpha_l x_{t-l} + \beta_1 y_{t-1} + \dots + \beta_l y_{-l} + u_t$$

for all possible (x, y) pairs of series in the group. Correlation between the rest of the regressions if not considered. It is also worth keeping in mind, that Granger causality requires only statistically stationary time series to apply to.

The null hypothesis implies, that X show no Granger-cause on Y in the first regression, and that Y either does not Granger-cause X in the second one.

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_l = 0$$

As the regression equations are set up, the F-test runs by the formula, saying:

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$

RSS_R here stands for the residual sum of squares for the model consisting some limitations (such as containing only lagged values of Y or even other variables, and omitting lagged values of X, for example). RSS_{UR} stands for the residual sum of squares for the model that does not comprise limitations; m is the based on the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) number of lags for the X-variable; k stands for the quantity of parameters computed in the model of no limitations.

The rejection of the null hypothesis is said to fail if the p-value is above 0.05, meaning neither X Granger-causing the Y, nor Y causing X. In contrary, p-value under 0.05 implies a successful rejection of the null hypothesis, meaning one variable actually Granger-causing the other one.

3.3.4 Wavelet coherence analysis

Until lately, wavelet analysis has been finding wide application rather in physics and engineering, than economics and finance. However, as wavelet analysis has been discovered to carry properties applicable for modelling various financial and economical occurrences, the trend started to change. The range of phenomena for which wavelet analysis finds application, also includes the matters of cryptocurrency markets price formation (e.g. Kristoufek, 2015; Phillips & Gorse, 2018; Kang et al., 2019; Betre, 2019; Goodell & Goutte, 2020; Erzurumlu et al., 2020).

The term wavelet stands for wave-resembling functions meant to transform signals into a representation comprised of time and frequency domain elements. In graphical representation wavelets appear as wavy oscillations, whose amplitude comes off from zero, grows up, and then finally bounces back to zero. Alternatively, wavelets can be considered as a bandpass filter to attach to the explored time series. Such filter would solely give way to elements of the time series within a limited range of frequencies to various extents bound to the wavelet's energy spectrum.

As by Terrence and Compo (1998), wavelets take the form:

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right)$$

The u parameter defines the wavelet's location. The s scale parameter stands for the wavelet's width, describing the extent to which the wavelet is stretched while keeping the same wavy lineament. For larger s values, the wavelet's width increases, meaning that more of the subject time series is scoped. This, however, reduces the observation's granularity, which means the time series is viewed from a higher level. Low scales enable the analysis of higher frequency, short-term dynamics of the explored time series. High scales, conversely, permit long-term (lower frequency) dynamics analysis. If at a specific temporal location and scale the time series follows a similar pattern to the wavelet, such occurrence generates a large transform value. Applying the wavelet function is continuously (as it's done in this study), defines as continuous wavelet transform. The continuous wavelet transform is expressed as

$$W_x(u,s) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{\mathrm{s}}}\Psi^*\left(\frac{\mathrm{t}-\mathrm{u}}{\mathrm{s}}\right)dt$$

where Ψ * is the complex conjugate of Ψ. To avoid information loss, the initial series can be rearranged from the continuous wavelet transforms for designated frequencies. A wide range of complex-valued wavelets enabling multivariate analysis has already found application in similar studies (Kristoufek, 2015; Phillips & Gorse, 2018). This study involves Morlet wavelet for its well-balanced relationship of time and frequency localization.

Continuous wavelet transforms come in handy while dissembling and examining the composite waveforms of an explored time series. Examining two time series and looking to figure out the locations of similar correlations with a particular wavelet, may successfully involve another wavelet transform known as cross wavelet transform. For two continuous wavelet transforms, $W_x(u, s)$ and $W_y(u, s)$, this is expressed as:

$$\mathrm{W_{x,y}(u,s) = W_x(u,s)W_y^*(u,s)}$$

where Wx(u, s) and Wy(u, s) are, correspondingly, continuous wavelet transforms of series x(t) and y(t), and the * symbol stands for the complex conjugate. Continuous wavelet transform reveals local covariance between time series at each scale, which means highlighting the areas of the time–frequency domain, within which time series perform high coinciding power. Wavelet coherence is capable of spotting the time series' co-movement in the time-frequency field. Wavelet coherence of time series, as defined by Torrence and Webster (1999), can be expressed as:

$$\mathrm{R^2(u,s) = \frac{\left|S(s^{-1}W_{x,y}(u,s))\right|^2}{S(s^{-1}|W_x(u,s)|^2)S(s^{-1}\left|W_y(u,s)\right|^2)}}$$

Where S is taken as a time and scale smoothing operator, with $0 \leq R^2(u,s) \leq 1$. The value of the wavelet squared coherence $R^2(u,s)$ is placed between 0 and 1, where high value indicates strong co-movement, and low value denotes weak co-movement between the time series. The wavelet squared coherence, though, contrary to standard correlation coefficient, can only be expressed in positive values. Graphically presented wavelet squared coherence allows to single out areas of the time series' co-movement in time-frequency domain. By this means, distinguishing whether the correlation is positive or negative appears impossible. Thereby, the phase difference of Terrence and Compo (1998) can be utilized to both reveal the co-movements' positive and negative nature, and casual relationship between the time series as well. The wavelet coherence phase difference is defined as below:

$$\Phi_{xy}(u,s) = \tan^{-1}\left(\frac{\Im\{S(s^{-1}W^{xy}(u,s))\}}{\Re\{S(s^{-1}W^{xy}(u,s))\}}\right)$$

where,$\Im$ and $\Re$ are, correspondingly, the imaginary and real parts of the smoothed cross-wavelet transform. Black arrows the wavelet coherence map mean to indicate the phase. If the phase-difference is zero, it means the rime series are moving conjointly. When the series are in-phase or positively correlated, the arrows point to the right. And conversely, left-oriented arrows signalize of the series being out of phase and correlated negatively. Arrow pointing upwards indicates the firs time series lead over the second by $\pi/2$. Respectively, second time series leading the first by $\pi/2$ is indicated by an arrow pointing downwards. Most commonly, though, the arrows appear in combined positions.

4 Results and discussion

This chapter comprises empirical outcomes of the carried-out research, limitations and outlines potential future extensions.

4.1 Empirical results

Pre-processing and sentiment analysis of the gathered data resulted in 2 generated data sets (one for Bitcoin and one for Tezos). The explored period includes 823 days from 1st august 2018 to 31st October 2020. Each cryptocurrency received 6584 related data points. Descriptive statistics for the examined variables are presented in Table 4.1.

Table 4.1: Descriptive statistics of dataset

Variables		Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Probability
Bitcoin	Price	7851.626	8018.000	13738.00	3256.000	2512.383	-0.155197	2.115656	30.12207	0.000000
	Google trends	27.95181	26.19923	100.00	15.14878	9.073119	2.216068	12.32380	3654.707	0.000000
	Wikipedia views	9603.168	8745.000	28791.00	4781.000	3221.472	1.645091	7.682926	1123.227	0.000000
	Tweets	25203.00	23685.00	87570.00	0.000000	9111.396	1.649598	9.210512	1695.899	0.000000
	News volume	302.3062	286.0000	955.0000	106.0000	111.0598	1.277004	5.733553	479.9211	0.000000
	Telegram mentions	19471.15	8325.000	106715.0	1352.000	23932.04	1.710769	4.817080	514.6735	0.000000
	Reddit sentiment	0.128293	0.129102	0.272101	-0.053126	0.047126	-0.034544	3.423198	6.305204	0.042741
	Posts + comments	229.0401	183.0000	654.0000	38.00000	146.1703	0.451993	1.927644	67.45640	0.000000
Tezos	Price	1.579547	1.352000	4.206000	0.359000	0.855394	0.700095	2.818223	68.36301	0.000000
	Google trends	13.36980	10.83513	100.0000	0.000000	9.930002	2.470876	13.86528	4885.716	0.000000
	Wikipedia views	133.4143	117.0000	1016.000	0.000000	134.9605	1.643339	8.365485	1357.630	0.000000
	Tweets	451.3366	378.0000	3367.000	0.000000	327.6805	2.914936	18.12909	9014.480	0.000000
	News volume	7.470231	6.000000	36.00000	0.000000	5.926457	1.498263	5.733235	564.0890	0.000000
	Telegram mentions	180.5930	77.00000	6097.000	2.000000	470.7554	9.021642	93.83648	294113.7	0.000000
	Reddit sentiment	0.250698	0.252167	0.585507	-0.171912	0.092148	0.016938	3.794173	21.66750	0.000020
	Posts + comments	46.76671	42.00000	150.0000	2.000000	22.56710	0.883292	3.648723	121.4494	0.000000

As follows from the descriptive statistics in Table 4.1, variables perform high standard deviation relative to the mean. This signalizes of the data being spread out and including outliers. Besides, the majority of the variables shows off leptokurtic characteristics. This means, the data is heavy tailed or overloaded with outliers. The except is for Bitcoin and Tezos prices, and Bitcoin subreddit posts and comments total, which are as well imperfect and show platykurtic characteristics. There should also be mentioned a positive skewness of most of the variables (except for Bitcoin price and Reddit sentiment for both currencies) at the spot of longer or fatter tail on the right side of the distribution. As a result, the Jarque-Bera test outcome shows high and significant values, disproving the assumption of all the explored variables' normality.

Visual evaluation of the explored time series descries noticeable trends, changing levels, tips of seasonality, changing variance, etc., which indicates of non-stationarity. For example, Figure 4.1 presents the prices graphs within the explored period, which clearly expose the trend element. Media data graphs, along of their size, are presented in the appendix A.

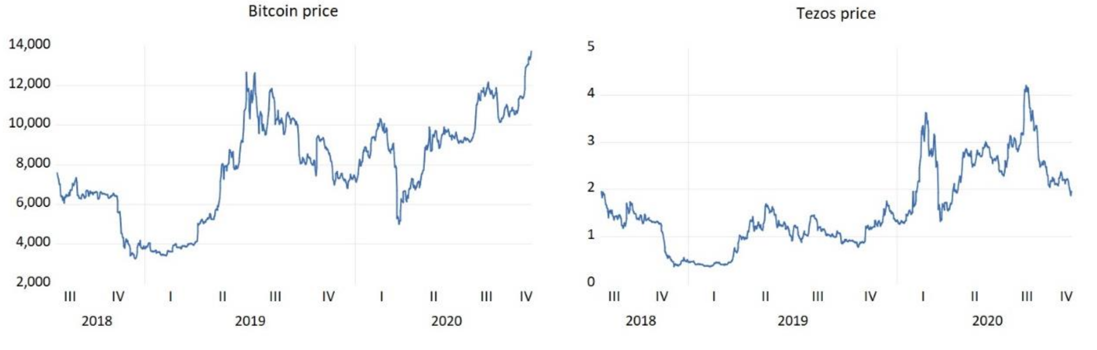


Figure 4.1: Cryptocurrency raw prices

The autocorrelation function (ACF) is utilized to attest the non-stationarity indication. Figure 4.2 presents the ACF for the time series shown in Figure 4.1.

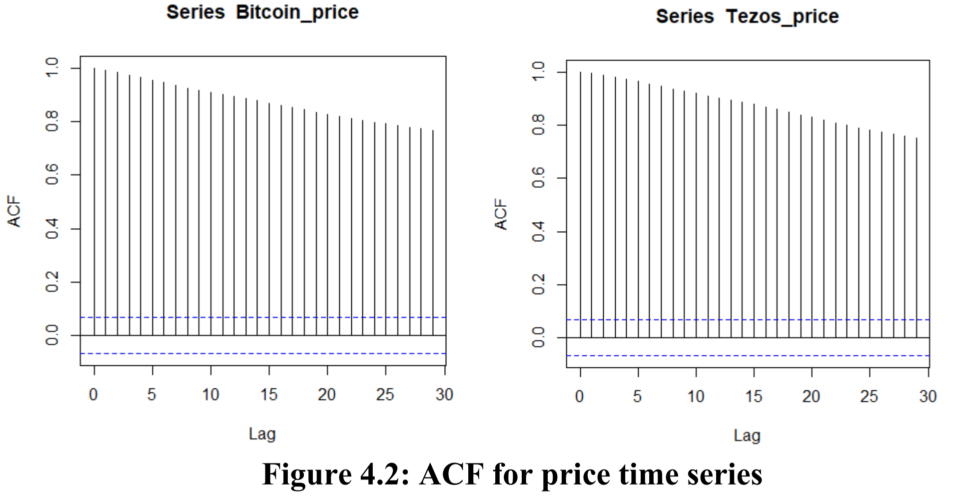


Figure 4.2: ACF for price time series

As seen in the figures above, for both cryptocurrencies, the autocorrelation functions are decreasing very slowly, and hold notably higher above the significance range shown as the blue dotted lines, which is the cue of non-stationary series.

Though graph and ACF plots exploration are giving univocal witness of this study's raw data non-stationarity, for solid assurance there are applied an augmented Dickey-Fuller test (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS), as mentioned in the methodology. Testing both stationarity hypothesis and the unit root hypothesis allows to distinguish seemingly stationary series, series of unit root, and series that do not have informative enough data or tests to tell whether they are integrated or stationary.

The results of the tests are displayed in Table 4.2.

Table 4.2: Stationarity and unit-root tests for raw data

Variables		ADF	P-value	KPSS	P-value
Bitcoin	Price	-1.9971	0.5795	6.0842	<0.01
	Google trends	-4.9654	<0.01	1.2053	<0.01
	Wikipedia views	-4.5969	<0.01	2.0582	<0.01
	Tweets	-3.3704	0.05855	3.0576	<0.01
	News volume	-3.6891	0.02454	7.2743	<0.01
	Telegram mentions	-0.73352	0.9675	7.8338	<0.01
	Reddit sentiment	-8.6312	<0.01	0.51446	0.03841
	Post + comments	-2.8531	0.2172	2.1894	<0.01
Tezos	Price	-3.1543	0.0958	7.2747	<0.01
	Google trends	-4.9919	<0.01	4.2738	<0.01
	Wikipedia views	-3.7173	0.02313	6.0831	<0.01
	Tweets	-5.1397	<0.01	1.1531	<0.01
	News volume	-5.2166	<0.01	2.3964	<0.01
	Telegram mentions	-6.2754	<0.01	2.0187	<0.01
	Reddit sentiment	-8.2374	<0.01	0.344	>0.1
	Post + comments	-6.0327	<0.01	2.783	<0.01

Table 4.2 shows, that all the time series, except for Tezos Reddit sentiment, are not stationary. Whereas, using non-stationary time series may end up obtaining fictitious results that indicate the nonexistent relationship between two variables. To avoid the misconception and receive the relevant outcome, non-stationary data has to be in the first place transformed from into stationary.

Even though in such cases researchers most commonly apply log returns instead of raw data (e.g. Phillips & Gorse, 2018; Kang et al., 2019; Kraaijeveld, & Smedt, 2020), the current study utilizes first differencing methodology instead. For a range of reasons.

First, the raw date contains a large amount of zeroes and negative numbers, which disables straight log-transformation, or those figures would otherwise have to be eliminated from the study, which might potentially compromise the accuracy of the research, as those cases are numerous, or there should be added a small constant value to each value of variable, and then run a log transformation. However, the natural logarithm of of 0.00001 is -5, while one of 0.01 is -2. So, the logarithms appear significantly different, despite both original values stand close to zero. In this view, the matter of choosing the relevant number and the distortion (or no distortion) of the received results comes out too questionable.

Second, checking log returns for stationarity with both previous approaches failed, and most of the series turned out to be non-stationary at this level. Thus, such transformation would not solve the problem, so other approaches are necessary. An attempt to apply percentage change instead of log returns neither solved the challenge of non-stationarity. Apparently, differencing may be useful in terms of stabilizing the mean of time series through eliminating the changes in time series level, and so excluding – or at least mitigating – trend and seasonality. This approach to cryptocurrency data transformation in order to achieve stationarity has already proven itself for the studies by Sovbetov (2018) and Alahmari (2019).

Table 4.3 illustrates the stationary nature of the first differences for each explored time series.

On achieving stationarity, Pearson correlation calculation was applied to estimate the liaison between explored variables. Table 4.4. shows the Pearson correlation between a number of cryptocurrency prices and related media factors across the whole time period. Should be mentioned, that this correlation analysis only considers same day correlations.

Table 4.3: Stationarity and unit-root tests for differencing variables

Variables		ADF	P-value	KPSS	P-value
Bitcoin	Price	-9.1001	<0.01	0.17118	>0.1
	Google trends	-10.498	<0.01	0.01015	>0.1
	Wikipedia views	-11.381	<0.01	0.014588	>0.1
	Tweets	-12.1	<0.01	0.07261	>0.1
	News volume	-10.544	<0.01	0.026944	>0.1
	Telegram mentions	-10.46	<0.01	0.25854	>0.1
	Reddit sentiment	-15.399	<0.01	0.013755	>0.1
	Post + comments	-12.074	<0.01	0.018866	>0.1
Tezos	Price	-8.8981	<0.01	0.097255	>0.1
	Google trends	-10.579	<0.01	0.0093944	>0.1
	Wikipedia views	-11.506	<0.01	0.014219	>0.1
	Tweets	-12.164	<0.01	0.0086021	>0.1
	News volume	-11.701	<0.01	0.01238	>0.1
	Telegram mentions	-12.531	<0.01	0.0054724	>0.1
	Reddit sentiment	-15.52	<0.01	0.0048624	>0.1
	Post + comments	-11.88	<0.01	0.012735	>0.1

Despite Bitcoin and cryptocurrency prices are famous for their close bonds with media and general publicity, the majority those correlations are, as seen in Table 4.4, weak, except for Google Trends and Wikipedia views for Tezos, which can be, at a pinch, said to perform moderate positive correlation. Peculiarly, the coefficient indicators for Tezos turn to be significantly higher that those of Bitcoin. Moreover, 4 out of 7 indicators for Bitcoin are not only close to zero, but also statistically insignificant (with significance level set at 5%). The exact nature of that occurrence is hard to tell – presumably, this is due to shorter history of Tezos as a project and a cryptocurrency, its smaller community and market cap, which in total makes it more vulnerable to info pumping and dumping showing faster and more sizeable effect.

The weak correlation might be explained by several reasons. First, if the direction of the correlation within the time period is lacking on consistency, the overall correlation is more likely to shrink across the whole timespan. The source of inconsistency can be illustrated by two hypothetical circumstances. For one instance, unexpected positive news regarding a cryptocurrency might push both prices and number of daily posts up. This would lead to a positive correlation within the period. For the other, some part of the cryptocurrency's ecosystem might be backdoored. The number of daily posts would rather, again, increase, as the community would spread the news further and engage more users into discussion – yet the price would rather go down along with negative news getting louder. For this period, the correlation would come out negative. Second, the correlation might be instable through the time, being strong at some partial periods and weak at other periods, which is also reflecting in the weak overall correlation. Another reason may consist in unstable users' activity across the web. Engagement across different networks and platforms varies by days of the week. For example, as reported by Popsters (2020), in 2019 the lowest activity across all social networks was detected on Tuesdays and Fridays, while the peak fell on Sundays. Also, the activity traditionally runs down to minimum on holidays (Christmas, Easter and so on). Besides, technical issues like server crashes lead to the same condition. Low social media activity then does not necessarily coincide with price drop periods, so the correlation on longer time periods finally thins out.

Table 4.4: Pearson correlation coefficient between cryptocurrency prices and social and mass media features

Variables	Bitcoin		Tezos	
	Pearson coefficient	P-value	Pearson coefficient	P-value
Google trends	0.1013	<0.01	0.3166	<0.01
Wikipedia views	0.0190	0.5851	0.3030	<0.01
Tweets	0.0865	0.01312	0.1952	<0.01
News volume	0.0162	0.6431	0.2750	<0.01
Telegram mentions	0.0232	0.5053	0.1622	<0.01
Reddit sentiment	-0.0269	0.441	-0.0106	<0.01
Posts + comments	0.0693	0.04701	0.1522	<0.01

Even though the correlation may help detect the correlation between time series, it still does not reveal whether a time series is leading the other one. However, this can be investigated through Grander causality.

In keeping with the current study, applying Granger causality may show four possible outcomes:

- Unidirectional Granger causality from the publicity factor (social and mass media) down to the price of the cryptocurrency - media changes coming before price changes.
- Unidirectional Granger causality from the cryptocurrency price down to publicity factor – price changes coming before changes in media.
- Bidirectional, or positive feedback, causality. Changes in both variables may precede each other.
- No Granger causality of any kind (unidirectional or bidirectional).

Granger causality test outcomes are substantially dependent on lag condition. Too little number of lags may allow finding residual autocorrelation ending up in a skewed test. If the number is too high, it may lead to incorrectly rejecting the null caused by erroneous correlation. Although optimum lag length can be determined through various methods, the most common one is applying information criteria – AIC and BIC (Bruns & Stern, 2019). For the use of the current study, the lag for each pair was chosen in regard to minimal Akaike information criterion defined by Eviews software.

Table 4.5 represents the outcome of pairwise Granger causality test for Bitcoin.

Table 4.5: Pairwise Granger causality test for Bitcoin

Null Hypothesis	F-statistic	P-value
Google trends does not Granger cause Bitcoin price	2.27903	0.0451
Bitcoin price does not Granger cause Google trends	2.82278	0.0155
Wikipedia views does not Granger cause Bitcoin price	0.79122	0.5769
Bitcoin price does not Granger cause Wikipedia views	1.25617	0.2754
Tweets does not Granger cause Bitcoin price	1.63918	0.0328
Bitcoin price does not Granger cause Tweets	1.38837	0.1105
News volume does not Granger cause Bitcoin price	0.47601	0.9782
Bitcoin price does not Granger cause News volume	0.79129	0.7325
Telegram mentions does not Granger cause Bitcoin price	0.81022	0.6585
Bitcoin price does not Granger cause Telegram mentions	0.60754	0.8603
Reddit sentiment does not Granger cause Bitcoin price	0.65388	0.7826
Bitcoin price does not Granger cause Reddit sentiment	1.10342	0.3552
Reddit total posts and comments does not Granger cause Bitcoin price	1.26473	0.2586
Bitcoin price does not Granger cause Reddit total posts and comments	1.22897	0.2787

As seen in the table above, Granger causality is not confirmed for all the variables but Google Trends and tweets. Notably, the number of tweets tagged #Bitcoin is likely to precede changes in the price of the cryptocurrency in a one-way fashion, while in case

Google Trends bidirectional causality appears – price changes and search queries show influence upon each other, which signalizes of Bitcoin's possible exposure to price bubbles. These findings go in line with the implications drawn by Kristoufek (2013).

The outcomes of Granger test, similarly to Pearson correlation coefficient, differ for the two cryptocurrencies. The result for the test on Tezos is presented in Table 4.6.

Table 4.6: Pairwise Granger causality test for Tezos

Null Hypothesis	F-statistic	P-value
Google trends does not Granger cause Tezos price	2.10555	0.0068
Tezos price does not Granger cause Google trends	2.83038	0.0002
Wikipedia views does not Granger cause Tezos price	1.73555	0.0196
Tezos price does not Granger cause Tezos views	2.73231	4.E-05
Tweets does not Granger cause Tezos price	2.59259	0.0060
Tezos price does not Granger cause Tweets	1.72117	0.0803
News volume does not Granger cause Tezos price	2.29191	0.0004
Tezos price does not Granger cause News volume	1.89260	0.0055
Telegram mentions does not Granger cause Tezos price	0.97381	0.49591
Tezos price does not Granger cause Telegram mentions	3.54014	1.E-07
Reddit sentiment does not Granger cause Tezos price	0.65893	0.7915
Tezos price does not Granger cause Reddit sentiment	1.18262	0.2910
Reddit total posts and comments does not Granger cause Tezos price	0.87812	0.6453
Tezos price does not Granger cause Reddit total posts and comments	1.69324	0.0159

Thus, complete absence of any kind of Granger causality has been detected solely for the Tezos / Reddit sentiment pair. Meanwhile, there is a correlation between the price and the number of Reddit posts and comments. That is, price changes evoke movements in the users' activity in a corresponding subreddit. An identical condition is observed in Telegram, which also reveals unidirectional Ganger causality from the cryptocurrency price down to Telegram mentions. Further, though, the situation starts to coincide with Bitcoin – the number of tweets tagged #Tezos seem to set off price changes, and Goggle Trends showing bidirectional causality. However, the bidirectional causality in case of Tezos does not stick to Google Trends only, but also appears in the price / Wikipedia views and price / news volume pairs. This inclines even more towards acknowledging its bubble behavior and the speculative nature, as increasing interest may push the price up, then the growing price attract even more media interest, and finally pushing the price even higher. In this manner, from December 2018 Tezos price skyrocketed by uncanny 1100% from \$0,31 to \$3,75 within just 14 months, which can definitely not by explained by fundamental factors.

In terms of exploring the interdependence between media drivers and prices of the cryptocurrencies, this study deploys wavelet coherence approach. Wavelet coherence technique unscrambles signals into the frequency dimension and allows to figure out the correlation between two sets of time series data over time in various frames. By this means, wavelet analysis turns to be an ultimate solution for proper understanding the long and short-term effects, assuring more credible results compared to those provided by traditional time domain causality approaches for time series variables affected by nonlinearity and erratic behavior.

For ultimate clarity, short, medium, and long term should be defined explicitly. In term of the current study, short term stands up to 8 days, medium – 8-16 and 16-32 days, and long term refers to 32-63, 64-128 and 128 to 256-day period.

The present study aims at the outcomes of the wavelet coherence analysis to reveal and dissect the magnitude of the sway one variables bear over other. The wavelet location in time is designated by the x-axis, and the y-axis is the wavelet period in days. Morlet wavelet is deployed for the wavelet power spectrum. Contours are for wavelet-squared coherencies of 0.0, 0.2, 0.4, 0.6, 0.8, and 1. The region under affect or edge effects is displayed by the white cone influence. Regions with 5% level of significance are indicated in the plot by the black outline. The estimation is run utilizing the Monte Carlo simulation with phase-randomized surrogate series. The color scale for the coherence lies in between blue and red hues, which correspondingly indicate low, close to zero, and high, close to one, coherence.

The direction at which the arrows are oriented indicates two parameters – the time series, that is leading the relationship on the current spot, and the correlation. Left-looking arrows signalize of negative correlation between the two time series at the spot, whish stands for anti-phase. Arrows looking rightwards indicate positive correlation – in-phase. If the first time series is the leading one, the arrows are looking towards left-up or right-down. Arrows oriented left-down or right-up mean, that the leading time series is the second one.

Information form adjacent data is used at each point. For finite data series the start and end areas will be missing out on some data, especially at bigger period bands. Using a cone-shaped field of influence to define the difference in outcome reliability, is an acknowledged standard. Areas of less reliable results beyond the cone of influence are distinguished with the pale hues. Computation resulting in the cone-shaped field requires the more data, the higher period bands are.

For ultimate representativeness, received wavelet coherence scalograms between different cryptocurrency and factor combinations are presented in Appendix B. Scalograms for different currencies are placed in separate columns, and scalograms for different factors are placed in separate rows., Ascending the column gives the view on a cryptocurrency's relationship with different factors, and tracking along a row allows to compare any interrelation between the factor and different currencies.

For both cryptocurrencies, the short-term relationship looks unstable and chaotic. The correlation appears and disappears, in one moment media factors precede the price, in another they are already falling back. For Tezos, however, those fleeting are, as a rule, positively correlated (signed by right-oriented arrows), that is, increases of on-line activity go along with growths in price.

Thorough examination of the Tezos coherence scalograms revealed a time interval within which all media factors showed statistically significant positive correlation with the price, falling on November 2019. Searching through Omenics[14] news portal for news concerning the cryptocurrency, the could possible source for that occurrence, discovered, that on 6th November 2019 Coinbase[15], one of the largest crypto exchanges based in the United States, enabled XTZ holders to stake their Tezos coins straight from Coinbase wallets and earn some profit out of it. It also launched the so-called Coinbase Earn Tezos initiative, that allowed users from around the Coinbase community to complete simple tasks, that rewarded up to $6 in Tezos tokens every month. The promise on which Coinbase built up the campaign was about simplifying the entire staking for the users, empowered with a guarantee of extra security. Within next 48 hours Tezos price blew off and uncanny growth of 117%, which makes a viable explanation for the detected correlation.

As for Bitcoin, significant correlations on short-term period have not been found. The unstable and tempered short-term relationships may witness, that media factor might not be a price movement driver of strong impact, except for moments of extreme, extraordinary newsbreaks (like one described above), that drive a rash in the market evoking heated discussion. Meanwhile, traders on cryptocurrency exchanges are often following established daily trading strategies based on technical analysis instead of news background.

[14] https://omenics.com/

[15] https://www.coinbase.com/en/

Medium-term relationships appear less chaotic compared to those of short-term perspective. Considering the scalograms together allows to detect particular periods of strong correlation, separated by notable spaces of no-correlation. Mostly, the detected correlations are positive, with the price driving the on-line activity. All pairs of variables, except for Tezos price / Wikipedia views and Tezos price / Reddit sentiment, show positive mid-length correlation in April and May 2019. Again, scanning the news records found out, that on 2nd April, Bitcoin price for the first time in that year hit $4,800, leading to a growth in price for altcoins, with crypto markets noting a double-digit growth. Thus, after a long stagnation, that lasted since 2018, the price bounce preceded a new wave of social media activity and fueled up the interest about the crypto world.

Worthwhile to mention, that in medium-term perspective most of significant correlations for Tezos arise in the second half of the examined time period (after August 2019), and for Bitcoin, in contrary, in the first half of the period.

Extended relationships appear more consecutive and persistent in time, and do not seem to be directly affected by particular news. Almost all long-term relationships, once occur, come out positive, which implies long-ranged correlation between the price and on-line activity.

For Tezos, strong long-term correlation was detected with Google Trends and Wikipedia views across the whole explore period, and with Tweets and news volume on the period after August 2019. Correlation with other factors appeared just local and fugitive. As for Bitcoin, long-term correlations appear locally and, in most cases, are limited by 32-64 days periods (with only exception for Tweets). That is, the results suggest, that mass and social media drivers are insufficient as external factors for Bitcoin in the long run.

Looking at the whole picture, aside from examining separate terms, reveals, that for Tezos the correlation between price and media factors is notably more significant and persistent, especially when it comes to the long-term relationship. In Bitcoins early years (until the infamous events at the beginning of 2018), when every media report regarding the new cryptocurrency was a newsbreak spreading at the speed of light, a single article might trigger price spikes and drops. However, as the once new cryptocurrency matured, media coverage spread, and other drivers gained on weight, it got too hard to derive a particular factor, like media alone, carrying out critical effect on the price formation. As by Chainalysis team, the growth of Bitcoin rates in 2020 is something completely different than what was experienced in 2017[16]. Back on that period major demand was generated by individual investors of various experience and knowledge on the cryptocurrency, mainly from Asia, investing their own funds. By 2020 Bitcoin became objective for corporate, institutional investors mostly based in Europe and the US seeking for new opportunities for strategic investments rather than speculative profits. Another crucial Bitcoin price growth factor is a significant increase of demand on top of comparatively low number of tokes available for purchasing - of 14,8 million Bitcoins mined, now only 3,4 million remain in free circulation. So, the good old rule of supply and demand comes in. As already mentioned, Tezos is a relatively new project in the crypto world, so has by now managed to gather around a much smaller community and market cap than Bitcoin, and thus is a way easier pot to stir, so media factors may carry out faster and more tangible effect on it, especially in short-term perspective. The source of the long-term positive coherence revealed between the price and media metrics, may lie in the phase of development of the currency. As the project evolves, gains wider implementation, and attracts more deep pockets, its community is growing too, increasing on-line activity, pumping up the demand and, naturally, the cryptocurrency's price. Apart from this, new players – crypto start-ups, dev teams – benefit a lot from interacting with the media as much as possible, arousing the clamor and fueling up the discussion on social media to resound, earn people's credit for their cryptocurrency and therefore generate the demand.

For both explored cryptocurrencies it is, in most cases, the price, that precedes the media factors. That is, price changes turn to be the supreme prior to hype, even though it may push the price even higher. Changes in price, especially rapid positive changes, instantly appear on media frontpages, reactivating users' engagement and motivating new users to dive for information in Google and Wikipedia, and flood the Twitter news feeds. On top of that, crypto markets are full of amateur players, who tend to fall into lemming instincts instead of employing consistent analysis and making balanced decisions – as a result rushing to buy tokens whenever the price starts moving up, and thus pumping it even further. Additionally, social discussion is a continuous process, so particular news or events may stay under ventilation in social media for weeks.

[16] Conforming to information from the company's blog https://blog.chainalysis.com/reports/bitcoin-price-surge-explained-2020/

4.2 Future research and limitations

Cryptocurrencies and their prices belong to research domain, that is till emerging. An original restraint for such studies lies in obscure and intangible nature of most information related to cryptocurrencies. Besides, media in general is a huge space, that includes hundreds of communication channels, and since covering them all at once is impossible, each researcher picks an array to one's own discretion. Therefore, the analysis of correlations between mass and social media, and cryptocurrencies, may provide overestimated or underestimated outcomes.

The period of exploration is also an important feature to choose for research matters. A lot of previous studies tended to scope time periods before 2018, as did, for example Rognone et al. (2020) and Neves (2020), who explored the fledging period of the cryptocurrency market between 2012 and 2018. This study deliberately covers a more recent period between August 2018 and October 2020 – the one after the early 2018 crypto market crash. The crash is known for causing not only a massive collapse of cryptocurrencies' market capitalization, but as well a notable churn of members of the ecosystem who had lost their trust in the concept, which inherently decreased the on-line activity. The following stagnation affected the entire year 2018 and early 2019, too. The year 2020 neither could be assumed a "normal" one in both the crypto world, and the real world due the circumstances carried out by COVID pandemic. However, on top of the global crisis and total uncertainty, Bitcoin has been showing sustainable price growth since May 2020 and by December 2020 has reached $23.000.

This study only employs the data in English language. This circumstance may potentially result in distorted measures, since there's a bulk of news and discussion (which potentially may even exceed the English one) ran across the world in other languages. After all, one should always keep in mind the large number of Asia-based cryptocurrency traders and holders. As follows from „The 2020 geography of cryptocurrency report" by Chainalysis, China is a sole keeper of 65% of Bitcoin's total hashrate, while from mid-2019 to mid-2020, 31% of ally cryptocurrency transactions occurred in East Asia.

Telegram provides its users with the advantage of safe and convenient closed groups and chats, which earned it huge popularity among the crypto community. The same feature, however, raises just as huge obstacles for researchers in terms of data collection, as the data can only be accessed through an administrator's pass permit, so the researcher is limited by publicly available data, which may, again compromise the analysis.

Tenuous correlation between Reddit sentiment and cryptocurrencies' price movements can be explained by several reasons. First, the imperfection of sentiment analysis tools. Posts in social media are often written informally, interspersed with scene specific slang, misspellings, allegoricality or cynicism. At the current stage of sentiment analysis development, processing such overly complicated data precisely is too difficult. Salač (2019) confirms VADER having trouble with evaluating negations in phrases. These results in way too many posts rated by the software as neutral, blurring the real sentiment picture.

Second, the analysis only reached for main subreddits (r/tezos, r/bitcoin), while the discussion each cryptocurrency may at the same time have a number of smaller sub-communities, like r/tezostrader or r/btc, or reversely be part larger, general ones like r/cryptocurrency, that consolidates over a million members. Thus, another massive layer of data might be omitted.

Third, a lot of submissions carry not only textual, but as well graphical or video content. Moreover, some post heading might be missing out on proper key words or semantic charge, so the real sentiment of the post is only encased in image or video. Apparently, text analysis ignores those types of content as well.

To avoid or potentially fix the described limitations, it might be useful for further studies to try and modify the sentiment analyzing software by adapting it to specific crypto semantics, employ machine learning advancement in terms of recognizing the sentiment carried by images or videos, and extend the scope of explored subreddits.

Generally, the research could potentially be elaborated and further advanced by altering the frequency of data, fracturing daily means into hourly or minutely, since prices may often be sore to certain events, but the leap may last for very short time and not get noted under day-scale examination. The number of explored cryptocurrencies can be extended as well to check whether the revealed trends are valid for the whole crypto market, or its only the case of particular currencies. Most viable approach would be to compare new entry cryptocurrencies and the old residents to see if there are substantive differences in the ways newcomers and well-established, reputable, and highly valued in the market projects react to factors of the same nature.

The quality of data collected from social media could also use some improvements – such as distinguishing not only the volume and sentiment of the content, but also taking into account the background and the impact of particular influencers or opinion shapers. Online crypto communities are highly differentiated and involve both amateurs and billion-dollar worth project creators. As a rule, proficient users are able to generate more insightful, far-sighted content and discuss more advanced future projects compared to novice authors. Separating the content by source may provide more informative data and assure more profound analysis. However, there still remains a risk of digesting content created by malicious users looking to fake their real expertise in pursuit for market manipulations. From the standpoint of econometrics, providing more comprehensive analysis could use simultaneous application of the generalized autoregressive conditional heteroskedasticity model and wavelet coherence analysis, as made by Kang et al. (2019) while exploring the relationship between Bitcoin and gold, or by Kumar and Anandarao (2019) who tried to unveil the source of volatility spillover in crypto markets.

5 Conclusion

This paper investigates whether media hype actually exerts that much of influence upon cryptocurrencies price movements and whether it can be used as the basis for future movements prediction. For this purpose, two cryptocurrencies, Bitcoin and Tezos, and 7 media factors (Google trends, Wikipedia views, number of tweets, news volume, Telegram mentions, Reddit sentiment, Reddit total number of posts and comments) for each of them have been considered on daily basis from 08-01-2018 to 10-31-2020. Pearson correlation coefficient is used to estimate the liaison between explored variables. In order to go beyond correlations and develop a better understanding with regard to the causalities, the study utilizes Granger causality analysis. Finally, for the purpose of exploring the interdependence between media drivers and prices of the cryptocurrencies in short and long timespan this study deploys wavelet coherence approach.

The first hypothesis, starting the study, is: There is an undeniable correlation between mass and social media and cryptocurrencies rates fluctuations.

In general, through the whole explored period the correlations are majorly weak, except for Google Trends and Wikipedia views for Tezos, which can be, at a pinch, said to perform moderate positive correlation. Moreover, all the indicators for Bitcoin but Google trends, tweets, and Reddit total posts plus comments turned out to be not only close to zero, but also statistically insignificant. The weak correlation might be explained by several reasons. First, if the direction of the correlation within the period is lacking on consistency, the overall correlation is more likely to shrink across the whole timespan. Second, the correlation might be instable through the time, being strong at some partial periods and weak at other periods, which is also reflecting in the weak overall correlation. Another reason may consist in unstable users' activity across the web. Generally, for Tezos, the correlation between price and media factors is notably more significant and persistent, especially when it comes to the long-term relationship. In Bitcoins early years (until the infamous events at the beginning of 2018), when every media report regarding the new cryptocurrency was a newsbreak spreading at the speed of light, a single article might trigger price spikes and drops. However, as the once new cryptocurrency matured, media coverage spread, and other drivers gained on weight, it got too hard to derive a particular factor, like media alone, carrying out critical effect on the price formation. Thus, the correlation does actually exist, but is way too unstable in terms of time and direction.

Hypothesis two: monitoring of media hype around cryptocurrencies can be used as the tool for cryptocurrency rates prediction.

Examining the lead-lag relationship between the variables through wavelet coherence revealed, that for both explored cryptocurrencies it is, in most cases, the price, that precedes the media factors. That is, price changes turn to be the supreme prior to hype, even though it may push the price even higher. Changes in price, especially rapid positive changes, instantly appear on media frontpages, reactivating users' engagement and motivating new users to dive for information in Google and Wikipedia, and flood the Twitter news feeds. On top of that, crypto markets are full of amateur players, who tend to fall into lemming instincts instead of employing consistent analysis and making balanced decisions – as a result rushing to buy tokens whenever the price starts moving up, and thus pumping it even further. Additionally, social discussion is a continuous process, so particular news or events may stay under ventilation in social media for weeks. Thus, in most cases the hype in media cannot act as a reliable basis for price movement predictions. These implications coincide with the opinion of Vitalik Buterin, co-founder of the Ethereum project, tweeted in February 2020 - "Your daily reminder that 95%+ of articles of the form "event X will make crypto go (up | down)" are post-hoc rationalized bullshit". The tweet had and attached photo of two newspaper articles about the effect of coronavirus on BTC rates, with each of the articles offering completely opposite conclusions than the other one.

Hypothesis three: The influence of mass and social media is much stronger in the short run.

For both cryptocurrencies, the short-term relationship looks unstable and chaotic. The correlation appears and disappears, in one moment media factors precede the price, in another they are already falling back. The unstable and tempered short-term relationships may witness, that media factor might not be a price movement driver of strong impact, except for moments of extreme, extraordinary newsbreaks, that drive a rash in the market evoking heated discussion. Meanwhile, daily trends in cryptocurrency exchanges patterns stemming from technical analysis. Extended relationships appear more consecutive and persistent in time, and do not seem to be directly affected by particular news. Almost all long-term relationships, once occur, come out positive, which implies long-ranged correlation between the price and on-line activity. The source of the long-term positive coherence revealed between the price and media metrics, may lie in the phase of development of the currency. As the project evolves, gains wider implementation, and attracts more deep pockets, its community is growing too, increasing on-line activity, pumping up the demand and, naturally, the cryptocurrency's price. Apart from this, new players – crypto start-ups, dev teams – benefit a lot from interacting with the media as much as possible, arousing the clamour and fuelling up the discussion on social media to resound, earn people's credit for their cryptocurrency and therefore generate the demand. Thus, the hypothesis has failed to prove itself right.

Returning to the basic question of the study, the answer is, that price changes turn to be the supreme prior to hype, even though the growing ado may push the prices even higher. Thus, hype is failing to prove itself as a reliable cryptocurrency price predictor. Crypto investors, though, should anyways take the news background into account while building trading strategies, especially for new projects in the market.

This research contributes to the growing literature on cryptocurrency and investor activity around it. The research may also be useful for investors in a better understanding the connection between mass and social media and cryptocurrency prices. Especially, understanding which cryptocurrency is more affected and when by media platforms.

Bibliography

Abraham, J., Higdon, D., Nelson, J. & J. Ibarra (2018): "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis." SMU Data Science ices

Alahmari, S. (2019): "Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies." The ISC International Journal of Information Security, 11(3), pp. 139-144.

Auer, R. & S. Claessens (2018): "Regulating cryptocurrencies: Assessing market reactions." BIS Quarterly Review, September: pp. 51-65.

Betre, C. (2019): "Analysis of the relationship between Prices and Transactional Drivers of Bitcoin." Senior Independent Study Theses. Paper 8364.

Bianchi, D. (2017): "Cryptocurrencies as an Asset Class: An Empirical Assessment." Journal of Alternative Investments, forthcoming.

Bouoiyour, J. & R. Selmi (2015): "What Bitcoin Looks Like?" Annals of Economics and Finance, vol. 16, issue 2, pp. 449-492.

Bremmer, G. (2018): "Predicting tomorrow's cryptocurrency price using a LSTM model, historical prices and Reddit comments." Master thesis, Tilburg University, Tilburg, the Netherlands.

Bruns, S.B. & D.I. Stern (2019): "Lag length selection and p-hacking in Granger causality testing: prevalence and performance of meta-regression models." Empirical Economics, 56, pp. 797–830.

Buchholz, M., Delaney, J., Warren, J. & J. Parker (2012): "Bits and Bets Information, Price Volatility, and Demand for Bitcoin." Economics 312. Available at: http://www.bitcointrading.com/pdf/bitsandbets.pdf [Accessed: November 5, 2020].

Chainalysis (2020): "The 2020 Geography of Cryptocurrency Report." Available at: <https://go.chainalysis.com/2020-geography-of-crypto-report.html> [Accessed: December 26, 2020].

Chainalysis team (2020): "Why Bitcoin Is Surging And How This Rally Is Different From 2017 (Hint: It'S Who'S Buying)." Available at:

https://blog.chainalysis.com/reports/bitcoin-price-surge-explained-2020 [Accessed: December 21, 2020].

Chaum, D. (1982): “Blind Signatures for Untraceable Payments”. Advances in Cryptology Proceedings of Crypto 82, pp. 199-203.

Ciaian, P., Rajcaniova, M. & D. Kancs (2016): “The economics of BitCoin price formation.” Applied economics, 48(19), pp. 1799–1815.

Ciaian, P., Rajcaniova, M. & D. Kancs (2018): “Virtual relationships: Short- and long-run evidence from BitCoin and altcoin markets.” Journal of International Financial Markets, Institutions and Money, 52, pp. 173–195.

CipherTrace (2020): “Cryptocurrency Anti-Money Laundering and Crime Report, Spring 2020.” Available at: https://ciphertrace.com/cryptocurrency-anti-money-laundering-and-crime-report-spring-2020/ [Accessed 6 Nov. 2020].

Civitarese, J. & L. Menders (2018): “Bad News, Technical Development and Cryptocurrencies Stability.” Capital Markets: Market Efficiency eJournal.

Dai, W. (1998): “B-Money”. Available at: http://www.weidai.com/bmoney.txt [Accessed: November 3, 2020].

Dickerson, A. (2018): “Algorithmic Trading of Bitcoin Using Wikipedia and Google Search Volume.” Warwick Business School, Coventry, UK

ElBahrawy, A., Alessandretti, L. & A.Baronchelli (2019): “Wikipedia and Cryptocurrencies: Interplay Between Collective Attention and Market Performance.” Frontiers in Blockchain, 2.

Erzurumlu, Y.O., Oygur, T. & A. Kirik (2020): “One size does not fit all: external driver of the cryptocurrency world.” Studies in Economics and Finance, 37(3), pp. 545–560.

Foley, S., Karlsen, J.R. & T.J. Putniņš (2019): “Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?” The Review of Financial Studies, 32(5), pp.1798–1853.

Garcia, D., Tessone, C., Mavrodiev, P. & N. Perony (2014): “The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy.” Journal of The Royal Society Interface, 11(99), p.20140623.

Goodell, J.W. & S. Goutte (2020): “Co-Movement of COVID-19 and Bitcoin: Evidence from Wavelet Coherence Analysis.” SSRN Electronic Journal. Finance Research Letters, 101625.

Goodkind, A.L., Jones, B.A. & R.P. Berrens (2020): “Cryptodamages: Monetary value estimates of the air pollution and human health impacts of cryptocurrency mining.” Energy research & social science, 59(101281), p. 101281.

Granger, C.W.J. (1969): “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” Econometrica, 37, pp. 424-438.

Hamrick, J., Rouhi, F., Mukherjee, A., Feder, A., Gandal, N., Moore, T. & M. Vasek (2018): “The Economics of Cryptocurrency Pump and Dump Schemes.” The Foerder Institute for Economic Research and The Sackler Institute of Economic Studies, Working Paper No. 3-2019.

Hayes, A.S. (2017): “Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin.” Telematics and informatics, 34(7), pp. 1308–1321.

Hutto, C.J. & E. Gilbert (2014): “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” ICWSM, The AAAI Press.

Inan, S. (2018): “Are cryptocurrency price changes predictable.” Master thesis, Northeastern University Boston, Massachusetts, USA.

Kang, S.H., McIver, R.P. & J.A. Hernandez (2019): “Co-movements between Bitcoin and Gold: A wavelet coherence analysis.” Physica A: Statistical Mechanics and its Applications, 536, p.120888.

Kaya, Y. (2018): “Analysis of Cryptocurrency Market and Drivers of the Bitcoin Price Understanding the price drivers of Bitcoin under speculative environment.” Degree Project in Economics of Innovation and Growth, KTH Royal institute of technology, Stockholm, Sweden.

Kraaijeveld, O. & J. De Smedt (2020): “The predictive power of public Twitter sentiment for forecasting cryptocurrency prices.” Journal of International Financial Markets, Institutions and Money, 65, p. 101188

Krause, M.J. & T. Tolaymat (2018): “Author Correction: Quantification of energy and carbon costs for mining cryptocurrencies.” Nature Sustainability, 1(12), pp.814–814.

Kristoufek, L. (2013): "BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era." Scientific reports, 3(1), p. 3415.

Kristoufek, L. (2015): "What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis." PloS one, 10(4), p. e0123923.

Kumar, A. S., & Anandarao, S. (2019). Volatility spillover in crypto-currency markets: Some evidences from GARCH and wavelet analysis. Physica A: Statistical Mechanics and its Applications, 524, pp. 448-458.

Lamon, C., Nielsen, E. & E. Redondo (2017): "Cryptocurrency Price Prediction Using News And Social Media Sentiment." SMU Data Science, 1(3), pp. 1-22.

Lansky, J. (2018): "Possible State Approaches to Cryptocurrencies." Journal of Systems Integration, 9(1), pp. 19–31.

Li, T., Chamrajnagar, A., Fong, X., Rizik, N. & F. Fu (2019): "Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model." Frontiers in Physics, 7.

Li, X. & C.A. Wang (2017): "The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin." Decision support systems, 95, pp. 49–60.

Liu, J. & A. Serletis (2019): "Volatility in the Cryptocurrency Market." Open Economies Review, 30(4), pp.779–811.

Liu, Y. & A. Tsyvinski (2018): "Risks and Returns of Cryptocurrency." NBER Working Paper, No. 24877, August.

Ma, Y. (2020): "NLP: How does NLTK.Vader Calculate Sentiment?" Medium. Available at: https://medium.com/ro-data-team-blog/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b [Accessed: December 11, 2020].

Mai, F., Bai, Q., Shan, Z., Wang, X. & R. Chiang (2015): "From Bitcoin to Big Coin: The Impacts of Social Media on Bitcoin Performance." Thirty Sixth International Conference on Information Systems, Fort Worth.

Mai, F., Shan Z., Bai Q., Wang X. & R. Chiang (2018): "How does social media impact bitcoin value? A test of the silent majority hypothesis.” Journal of management information systems : JMIS, 35(1), pp. 19–52.

Matta, M. & Lunesu M.I. & M. Marchesi (2015): “Bitcoin Spread Prediction Using Social And Web Search Media.” Conference paper, Università degli Studi di Cagliari, Cagliari, Italy.

Mirtaheri, M., Abu-El-Haija, S., Morstatter, F., Steeg, G. & A. Galstyan (2019): “Identifying and Analyzing Cryptocurrency Manipulations in Social Media.” arXiv preprint arXiv:1902.03110

Nakamoto, S. (2008): “Bitcoin: A Peer-to-Peer Electronic Cash System.” Available at: https://bitcoin.org/bitcoin.pdf [Accessed: November 2, 2020].

Neves, R.H. (2020). Bitcoin pricing: impact of attractiveness variables, Financial Innovation, 6(1), pp. 1–18.

Pearson, K. (1930): “The Life, Letters and Labors of Francis Galton.” Cambridge University Pres

Perry-Carrera, B. (2018): “Effect of Sentiment on Bitcoin Price Formation.” Duke Journal of Economics, 4/2018.

Phillips, R. & D. Gorse (2018): “Cryptocurrency price drivers: Wavelet coherence analysis revisited.” PLoS one, 13(4), p.e0195200.

Phillips, R. (2019): “The Predictive Power of Social Media within Cryptocurrency Markets.” UCL (University College London).

Pitta, J. (1999). “Requiem for a Bright Idea.” Forbes. Available at: https://www.forbes.com/forbes/1999/1101/6411390a.html?sh=6b1710f2715f [Accessed 3 Nov. 2020].

Polasik, M., Piotrowska, A., Wisniewski, T., Kotkowski, R. & G. Lightfoot (2015): “Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry.” International journal of electronic commerce, 20(1), pp. 9–49.

Popsters (2020): “*Social Networks Users Activity Report*.” Available at: https://popsters.com/app/docs/Popsters_research_2019_en.pdf/ [Accessed 15 December 2020].

Poyser, O. (2017): "Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series." arXiv:1706.01437 [q-fin].

Rognone, L., Hyde, S. & S. Zhang (2020): "News sentiment in the cryptocurrency market: An empirical comparison with Forex." International Review of Financial Analysis, 69, p.101462.

Salač, A. (2019): "Forecasting of the cryptocurrency market through social media sentiment analysis." Bachelor's essay, University of Twente, the Netherlands.

Salles, R., Belloze, K., Porto, F., Gonzalez, P.H. & E. Ogasawara (2019): "Nonstationary time series transformation methods: An experimental review." Knowledge-Based Systems, 164, pp.274–291.

Shanaev, S., Satish S., Binam G. & A. Shuraeva (2020): "Taming the blockchain beast? Regulatory implications for the cryptocurrency Market." Research in international business and finance, 51(101080), p. 101080.

Shumway, R. & D.Stoffer (2017): "Time Series Analysis And Its Applications." Cham: Springer International Publishing AG.

Smith, J. (2016): "An Analysis of Bitcoin Exchange Rates." SSRN Working Paper, No.2493797, May.

Smuts, N. (2019): "What Drives Cryptocurrency Prices?" ACM SIGMETRICS Performance Evaluation Review, 46(3), 131–134.

Sovbetov, Y. (2018): "Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero." Journal of Economics and Financial Analysis, 2(2), pp. 1-27.

Statistics Solutions (2019): "Pearson's Correlation Coefficient" Available at: https://www.statisticssolutions.com/pearsons-correlation-coefficient/.[Accessed 10 Dec. 2020].

Steinert, L. and Herff, C. (2018). Predicting altcoin returns using social media. PLOS ONE, 13(12), p.e0208119.

Stenqvist, E. & J. Lönnö (2017): "Predicting Bitcoin price fluctuation with Twitter sentiment analysis." Degree project in technology, KTH Royal institute of technology, Stockholm, Sweden.

Torrence, C. & G.P. Compo (1998): "A Practical Guide to Wavelet Analysis." Bulletin of the American Meteorological Society, 79(1), pp. 61-78.

Torrence, C. & P.J. Webster (1999): "Interdecadal Changes in the ENSO-Monsoon System." Journal of Climate, 12, pp. 2679-2690.

Valencia, F., Gómez-Espinosa, A. & B. Valdés-Aguirre (2019): "Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning." Entropy, 21(6), p.589.

Viglione, R. (2015). Does governance have a role in pricing? Cross-country evidence from bitcoin markets. SSRN Electronic Journal. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2666243 [Accessed: November 7, 2020].

Wang, J., Xue, Y. & M. Liu (2016): "An analysis of bitcoin price based on VEC model." In Proceedings of the 2016 International Conference on Economics and Management Innovations. Paris, France: Atlantis Press, pp. 180–186.

Wijk, D. (2013): "What can be expected from the Bitcoin." Working Paper No. 345986, Erasmus Rotterdam Universiteit.

Wooley, S., Edmonds, A., Bagavathi A. & S. Krishnan (2019): "Extracting Cryptocurrency Price Movements from the Reddit Network Sentiment." 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, pp. 500-505.

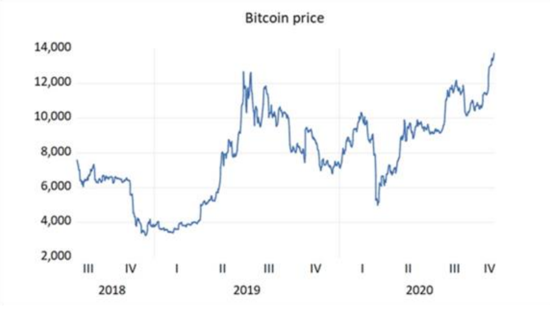
Yao, W., Xu, K. & Q. Li (2019): "Exploring the Influence of News Articles on Bitcoin Price with Machine Learning." 2019 IEEE Symposium on Computers and Communications (ISCC).

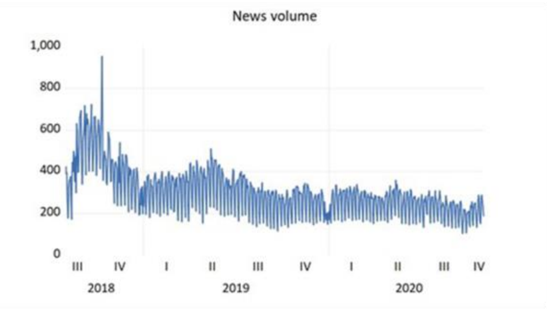
Yoshida, M., Arase, Y., Tsunoda, T. & M. Yamamoto (2015): "Wikipedia page view reflects web search trend." In Proceedings of the ACM Web Science Conference (Oxford, UK: ACM), 65.

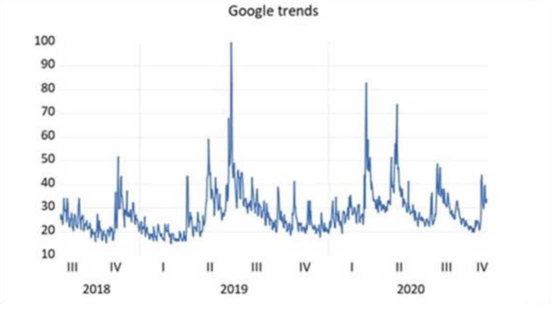
Zuo, X. (2019): "Several Important Unit Root Tests." 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP).

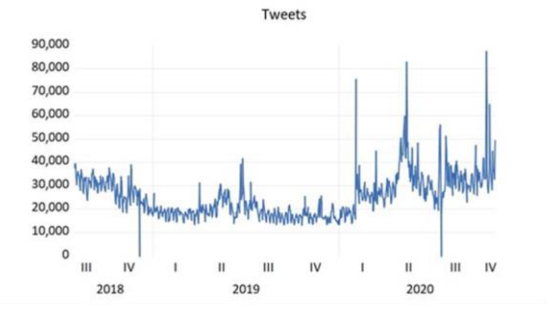
Appendix A: Plots of explored variables

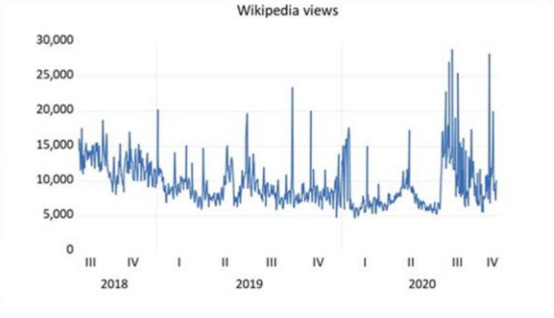
Bitcoin data set

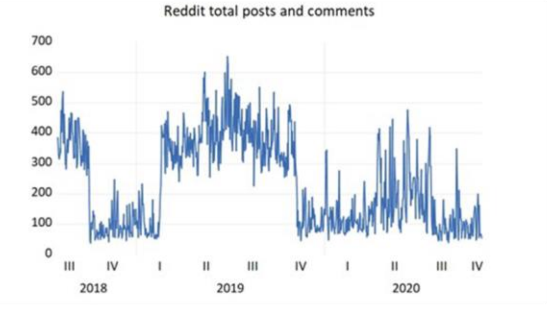


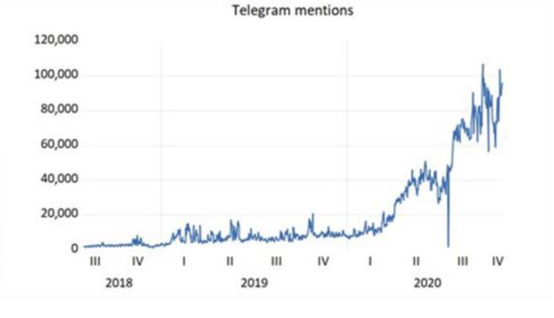


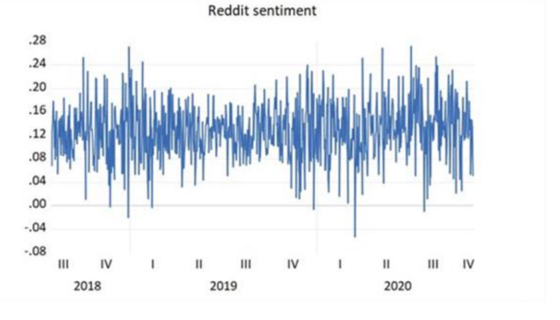




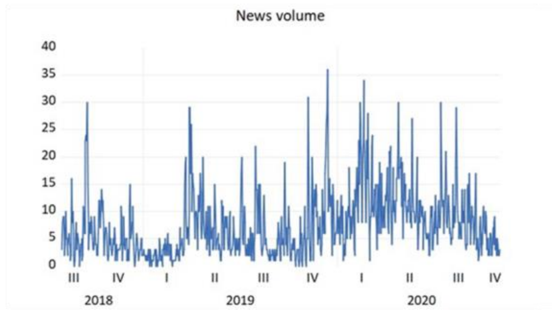


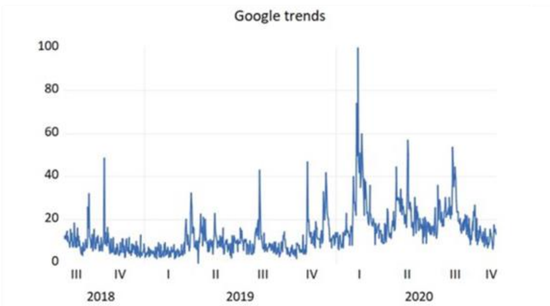


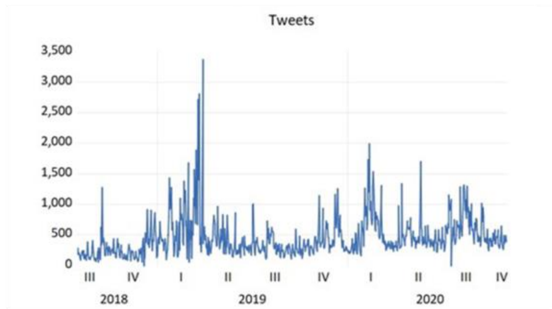


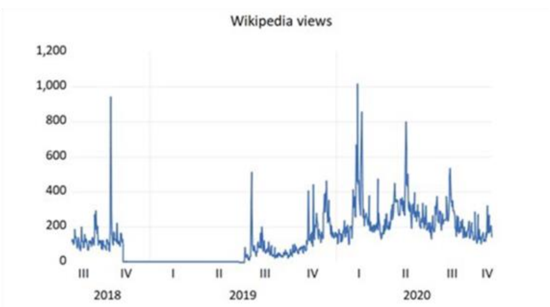


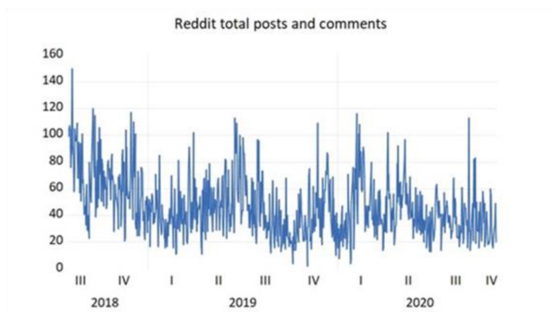
Tezos data set

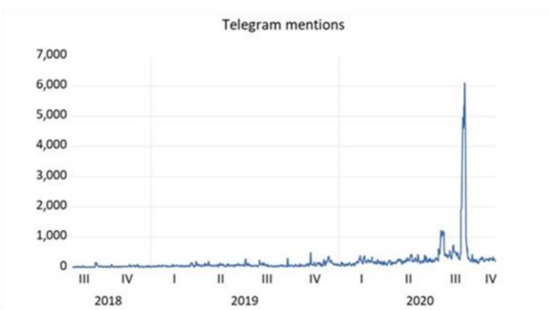


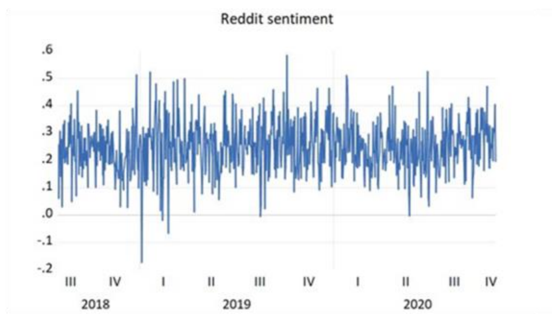




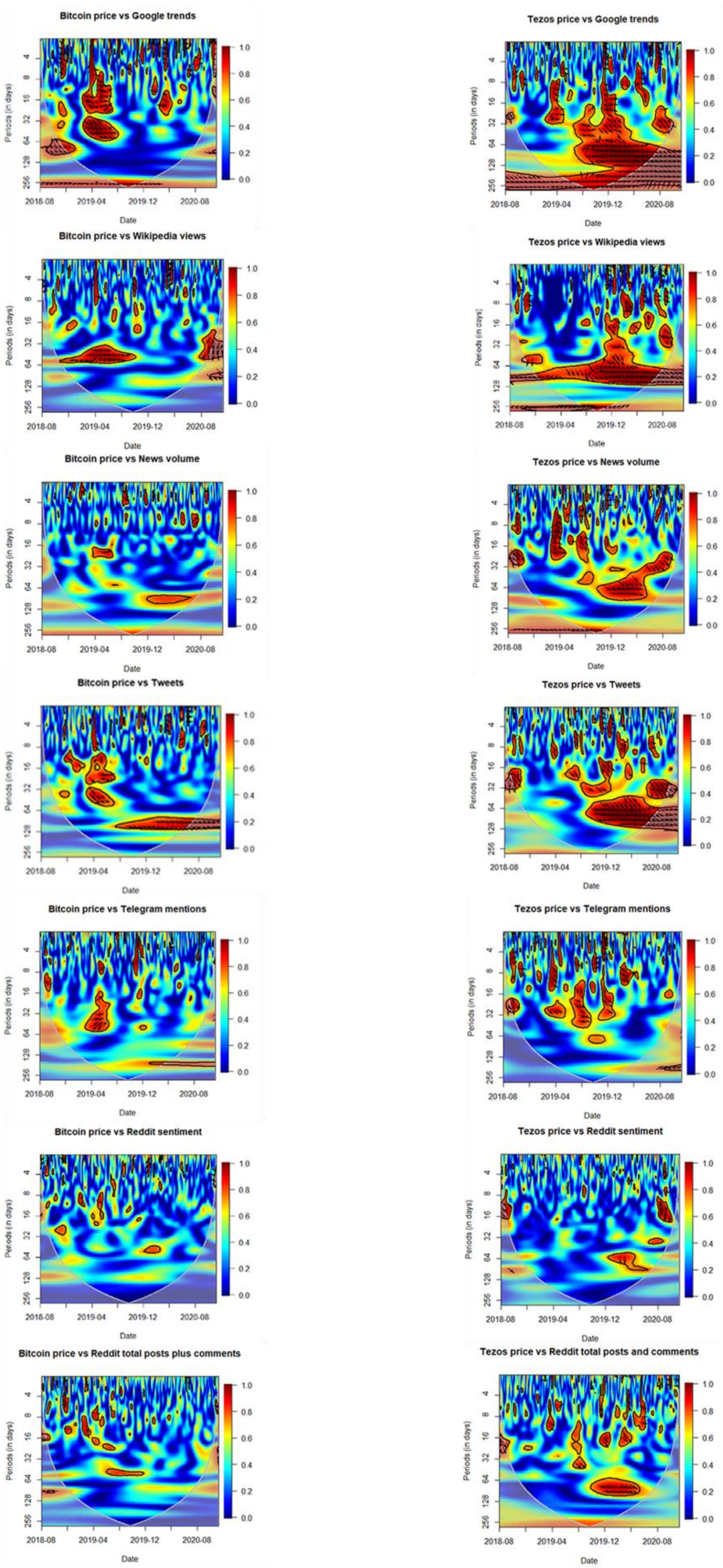








Appendix B: Wavelet coherence



Appendix C: Data availability

The study's text content, consolidated data set and sentiment extracted from Reddit are available under the following link: shorturl.at/eITW7. The access to RStudio and Python code can be obtained via direct request to the author.