Universiteit Leiden

ICT in Business and the Public Sector

Cryptocurrencies:
Mutual relationships and event reactions

Name: Ivan Osenov
Student ID: 1933981

Date: 8th Nov 2018

First supervisor: Dr. Michael Emmerich
Second supervisor: Dr. Xishu Li

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. Michael Emmerich and Dr. Xishu Li, for assisting me with their expertise and immensely insightful feedback during the iterative process of writing. Their valuable contribution and cooperative attitude made the argumentation in this paper even more rigorous and explicit.

I would also thank Mrs. Esme Kaubo and Dr. Arno Knobbe for their administrative help and support for all organizational activities.

Finally, I am grateful to my family and friends for their constant encouragement and help regarding brainstorming ideas on solving challenges along the way of writing this paper.

Abstract

After the cryptocurrency total market capitalization peaked at ~$800 billion during the 2017's rally, it attracted even more the attention of governments, businesses and investors. In effect, numerous hearings, conferences, workshops and other social gatherings started to be organised frequently around the topic of proper regulation, new business models, and other social and profit opportunities in the cryptocurrency domain around the world. Bitcoin, starting in 2009, has proved to be a potential candidate of an innovative decentralized payment system that could not only serve the current financial needs but also solve the social problem of trust in multi-party interactions. It has set the stage for experimentation and further development of the field by providing an example of how a decentralized, secure and relatively scalable payment network could function. Many technological entrepreneurs deviated by creating their own cryptocurrency projects, focusing on different value propositions such as smart contract platforms, privacy and performance, thus expanding the market. In parallel, various industries initiated their exploration of new products and services based on distributed ledger technology, i.e. blockchain. Taking into account the nature of startups and their capital requirements to ensure short-term sustainability, many retail and institutional investors bought into various established projects and initial coin offerings with the assumption that they will be able to experience higher returns on their investment than in the existing markets.

This study focuses on understanding the relationships between cryptocurrency projects in terms of finding common patterns in their price movement over time. The followed methodology employs an industry recognised standard for approaching data analysis tasks, namely CRISP-DM. As part of it, correlation analysis has been performed on the top 150 projects by market capitalization, along with all preliminary data preparation issues that were encountered, i.e. stationarization of time series. This part of the thesis results in a summary of quantified mutual connections that could be further used in the process of constructing an efficient portfolio of assets, as described in Modern Portfolio Theory.

Furthermore, the impact of 15 events of diverse nature has been analysed by means of an event study methodology in order to find to what degree cryptocurrencies are susceptible to change their direction in regards to expected returns. Being the current market leaders, Bitcoin and Ethereum were chosen to represent the domain for this analysis. The produced outcome consists of events that were to found to have caused significant changes.

Keywords: Blockchain, Bitcoin, Ethereum, Cryptocurrency, Correlation Analysis, Event Study

Table of contents

Acknowledgements **2**

Abstract **3**

Table of contents **4**

List of figures **5**

List of tables **6**

List of abbreviations **6**

List of appendices **7**

Chapter 1: Introduction **8**
- 1.1 Research questions 9
- 1.2 Structure of the paper 10

Chapter 2: Literature review **10**
- 2.1 Blockchain 10
  - 2.1.1 Definition and internal workings 10
  - 2.1.2 Categorisation of Blockchain architectures 11
  - 2.1.3 Smart contracts 14
- 2.2 Bitcoin 14
- 2.3 Cryptocurrencies 17

Chapter 3: Correlation analysis **18**
- 3.1 Methodology 18
- 3.1 Business understanding 20
- 3.2 Data understanding 21
  - 3.2.1 Data collection 21
  - 3.2.2 Data format 22
- 3.3 Data preparation 23
  - 3.3.1 Establishment of one-to-one relationship 23
  - 3.3.2 Removal of cryptocurrency projects lacking enough data 26
- 3.4 Modeling 26
  - 3.4.1. Stationary and non-stationary time series 26
  - 3.4.2 Methodology 28
- 3.5 Results 29
  - 3.5.1 Relationships between individual cryptocurrencies and the cryptocurrency market 29

3.5.2 Relationships between individual cryptocurrencies and Bitcoin 31
3.5.3 Relationships between the top 20 cryptocurrencies 32

Chapter 4: Event study **36**
4.1 Methodology 36
4.1 Results 39
4.1.1 Bitfinex attack 39
4.1.2 Consensus summit 40
4.1.3 BitConnect shutdown 40
4.1.4 Brexit referendum results 41
4.1.5 Trump election victory 42

Chapter 6: Discussion **43**
6.1 Relationships in the cryptocurrency market 43
6.2 Influence of events in the cryptocurrency market 44

Bibliography **45**

Appendices **47**
Appendix A: Cryptocurrency ticker symbols to names mappings 47

List of abbreviations

ADF	Augmented Dickey-Fuller
API	Application Programming Interface
AR	Abnormal Return
BIP	Bitcoin Improvement Proposal
CAR	Cumulative Abnormal Return
CMC	CoinMarketCap
CMRM	Constant Mean Return Model
CRISP-DM	Cross-Industry Process for Data Mining
DLT	Distributed Ledger Technology
EMH	Efficient Market Hypothesis
EU	European Union
IPFS	InterPlanetary File System
MC	Market Capitalization
MPT	Modern Portfolio Theory
OLS	Ordinary Least Squares
P2P	Peer-to-Peer
PoS	Proof-of-Stake
PoW	Proof-of-Work
SEMMA	Sample, Explore, Modify, Model, Assess
TMC	Total Market Capitalization
TS	Time Series
UK	United Kingdom
USD	United States Dollar

List of figures

Figure 2.2.1 Bitcoin transaction model
Figure 2.2.2 Bitcoin block model
Figure 3.1 Phases of the CRISP-DM reference model
Figure 3.5.1.1 Density plot of correlations between Total Market Cap and cryptocurrencies
Figure 3.5.1.2 Distribution of correlations between Total Market Cap and cryptocurrencies
Figure 3.5.2.1 Density plot of correlations between Bitcoin and cryptocurrencies
Figure 3.5.2.2 Distribution of correlations between Bitcoin and cryptocurrencies
Figure 3.5.3.1 Heatmap of the correlations between the top 20 cryptocurrencies
Figure 3.5.3.2 Graph of the correlations between the top 20 cryptocurrencies

List of tables

Table 2.1.2.1 Blockchain categorization matrix
Table 2.2.1 Bitcoin ecosystem design
Table 3.2.1.1 Filtering criteria
Table 3.2.2.1 Record format of cryptocurrency dataset
Table 3.2.2.2 Record format of TMC dataset
Table 3.3.1 Cryptocurrencies that violate the one-to-one relationship
Table 3.3.2 Evaluation of cryptocurrency projects - remaining and discarded
Table 3.3.3 Evaluation of cryptocurrency projects - merged
Table 3.3.2.1 Evaluation of cryptocurrency projects - lack of enough records
Table 3.4.2.1 Executed methodology actions on cryptocurrencies dataset
Table 3.4.2.2 Executed methodology actions on TMC dataset
Table 3.5.1.1 Descriptive statistics of correlations between TMC and cryptocurrencies
Table 3.5.1.2 Top 8 positively and negatively correlated cryptocurrencies to the overall market
Table 3.5.1.3 Top 8 uncorrelated cryptocurrencies to overall market
Table 3.5.2.1 Descriptive statistics of correlations between Bitcoin and cryptocurrencies
Table 3.5.2.2 Top 8 strongly correlated coins to Bitcoin
Table 3.5.2.3 Top 8 weakly correlated coins to Bitcoin
Table 3.5.3.1 Top 20 coins by market capitalization
Table 4.1 Classification of events
Table 4.1.1.1 Analysis of Bitfinex attack
Table 4.1.2.1 Analysis of Consensus summit
Table 4.1.3.1 Analysis of BitConnect shutdown
Table 4.1.4.1 Analysis of Brexit referendum results
Table 4.1.5.1 Analysis of Trump election victory

List of appendices

Appendix A: Cryptocurrency ticker symbols to names mappings
Appendix B: Working files

Chapter 1: Introduction

Since the last large-scale financial crisis and the introduction of Bitcoin to the world in 2008, there has been observed a tremendous growth of interest in alternative decentralized monetary systems by the public - individuals, businesses, institutions, and governments. By solving the what is known to be the double-spending[1] problem, Bitcoin aims to provide a new disintermediated digital infrastructure for money-related activities such as secure storage and instantaneous transactions. Moreover, the digital currency is considered to be privacy-friendly as it is pseudonymous in design, meaning neither personal details are stored on the network, nor they are needed for its proper functioning. Not being linked to a central authority contributes to the fact that international payments are executed in a faster and cheaper way than current bank remittance operations. Bitcoin is also the first decentralized cryptocurrency network to gain massive user base traction that amounted to between 2.9 to 5.8 million unique users in 2017 (Hileman & Rauchs, 2017). Decentralized by design, the first cryptocurrency has set itself apart from the traditional government-backed fiat currencies and their corresponding issuance authorities, namely the central banks, by allowing individuals with standard computing devices to participate in a publicly accessible virtual network and validate financial transactions. This new form of a payment network backbone has the significant potential of rendering the need of a central entity obsolete. Bitcoin has set the idea and technology fundamentals for further innovation in the area of payment systems as it has proven itself to be sustainable during the last decade, albeit not flawless. Its mission to become money that is as widely adopted and used as traditional fiat currencies is challenged by many drawbacks, some of which are the high price volatility, security concerns and other barriers such as non-trivial interaction processes with software applications and regulatory uncertainty. Due to the immaturity of the Bitcoin's markets regarding liquidity and usage, it requires relatively small events such as manipulative business transactions or spread of fake news to cause an immense impact on the exchange rate to other currencies. Furthermore, the accumulation of participants investing and developing new products and services[2] around the Bitcoin platform has attracted many bad actors employing fraudulent methods with the purpose of scamming and eventually stealing funds from investors. Also, the virtual security of online marketplaces, where participants get together to trade different cryptocurrencies, has been breached numerous times leading to major losses of funds and trust undermining in the ecosystem.

The early days of a developing industry is a highly appealing opportunity to entrepreneurs who are risk tolerant and willing to speculate on the future value, in this case, of the cryptocurrency market. The magnitudes of potential profits are multiple times higher than other contemporary traditional financial assets. For example XRP's[3] appreciation in 2017 alone was 35,627%[4]. To date, a massive cryptocurrency ecosystem has been formed with over 2000 different projects and a total market capitalization of approximately $190 billion. However, its most recent peak, and all-time high, was in January 2018 at

[1] Double-spending is the process of spending a digital monetary unit more than once and is considered as a design flaw in digital cash systems.
[2] https://en.wikipedia.org/wiki/List_of_bitcoin_companies
[3] https://ripple.com/xrp/
[4] https://thebestbinaryoptionsbrokers.net/the-most-profitable-top-cryptocurrency-in-2017-is-ripple-with-35627

around $800 billion. The diverse range of projects with different business models and ecosystems presents an opportunity to any interested party to support and profit from the price fluctuations by building a portfolio of investments. The latter is needed as the size of the financial risk is as high as the potential monetary returns, following the risk-return spectrum[5] tradeoff. In order to mitigate the inherent risk of investing in any type of asset, a practice that has proven its effectiveness over time is to maintain a diversified portfolio, i.e. a group of uncorrelated assets (Markowitz, 1952). By identifying different cryptocurrency projects with little to no correlation, investors are able, for instance, to hedge the risk of all assets depreciate at the same time. Similarly, those with high correlation could be kept tracked and should a price movement occur in one of them, it might be a profitable decision to open market positions for the rest. A rigorous methodology of finding the interrelationships in a dynamic market with high volatility, as the cryptocurrency one, would improve investors' decision-making activities concerning which combinations of cryptocurrencies to include in their portfolios so that they can achieve an optimal exposure to that particular market.

1.1 Research questions

This study aims to provide interested parties in the cryptocurrency ecosystem with information about the current state of the market with respect to the strength of interrelationships between projects, as well as the magnitude of price action impact caused by various types of events. The results would contribute to improved investor decision-making processes when it comes to constructing efficient portfolios and risk expectations about future events. The problem of interrelationships is to be tackled by employing a correlation analysis onto a cryptocurrency subset of interest, whereas the measurement of events' price influence is to be examined by applying an event study methodology. In total, the cryptocurrency projects that are to be investigated mutually own around 91% market share in terms of capitalization as of June 2018.

In order to achieve the above objectives, this study will provide answers to the research questions stated below.

Main question: What are the relationships and event impact in regards to price level movement in the cryptocurrency market?

Sub-questions:

1. What is the relationship between individual cryptocurrencies and the cryptocurrency market?
2. What is the relationship between individual cryptocurrencies and Bitcoin?
3. What is the relationship between the top 20 cryptocurrencies by market capitalization?
4. What are some events that have had a significant impact on Bitcoin and Ethereum price fluctuations?

[5] https://en.wikipedia.org/wiki/Risk%E2%80%93return_spectrum

1.2 Structure of the paper

The report is divided into a total of six chapters. The first one gives an overview of the industry in question, its participants and related problems that, specifically, investors are currently facing. Its main purpose is to establish the context and scope within which the research will be conducted. The following chapter consists of more detailed background information about the cryptocurrency ecosystem along with a literature review on relevant topics which are considered of high importance to familiarizing the reader before executing the actual analysis. The latter also includes descriptions of concepts such as blockchain, Bitcoin and cryptocurrencies. In the next chapter, a full market analysis is conducted on specific projects which meet a set of predefined criteria by utilizing an industry-accepted framework as guidelines in the process. The results are illustrated with various visualisation techniques, following a systematic description of the processing steps, grouped into different categories such as business understanding, data understanding, data preparation, modelling, and evaluation. In another chapter, the impact of events on the cryptomarkets is measured. Lastly, a discussion of all findings is presented including limitations, future work ideas, challenges and potential ways of solving them, as well as multiple perspectives to be taken into account.

Chapter 2: Literature review

The literature review is conducted by thorough extraction of existing scientific work, acquiring relevant literature from Google Scholar, in particular, Elsevier and Springer databases. The used keywords for querying the databases are: blockchain, ecosystem, consensus, public versus private blockchains, smart contracts, Bitcoin, cryptocurrency, correlation, event study, stationarity, markets.

Also, the main concepts and models are presented to the reader in order to support the relevance and need for research in the domain in general. Following, analyses from similar work are presented, critically discussed and used for the formation of hypotheses.

2.1 Blockchain

2.1.1 Definition and internal workings

Blockchain is commonly referred to as a peer-to-peer[6] (P2P) decentralised and distributed database, or Distributed Ledger Technology (DLT), which serves as a data store for keeping past transactions in an immutable manner. By eliminating the need for a central server in its operations, the blockchain is run by a network of usually geographically-dispersed software clients, called mining or full nodes, or validators, which host and maintain a full copy of the database locally and follow a predefined algorithmic protocol

[6] Peer-to-peer network is one in which two or more devices are connected, and are able to communicate and share resources without a central server device to mediate the connection.

to validate incoming user transactions. The latter are made persistent by grouping them in blocks, linked to each other in a chronological sequence. In order to ensure full database state consistency throughout the network, some standard rules are encoded into the protocol, and a consensus about the current state of the blockchain is achieved between validators near real-time. In addition, the technology makes use of cryptographic techniques that allow for access right to be established, as well as end-users to prove ownership of digital assets and transact them with each other.

The underlying data structure used in DLT, similar to a linked list[7], comprises a chain of chronologically linked blocks. Each block contains a header and a set of transactions. Usually, the header consists of block metadata, a reference to the previous block's header, and a fingerprint, which is a calculated hash value of all the data allocated in this concrete block. The ordering mechanism of the data structure is implemented using sequential referencing from each following block to its previous block's fingerprint. Instead of using a numerical sequence or a timestamp, the blocks' fingerprint values serve the purpose of achieving data validity as it becomes trivial to detect any inconsistencies. For example, a simple tampering check can be done by comparing a block's fingerprint with a secondary calculation of its contents' hash value; if the latter do not match, then the data in this block is invalid. Therefore, if an adversary tries to replace the data of an already existing block, they would have to regenerate all the following blocks' fingerprints, which in turn would lead to dramatic changes in the blockchain contents.

As part of the DLT goals to be a decentralised and distributed technology, a peer-to-peer network architecture is incorporated widely in its design. The file-sharing platform BitTorrent[8] is a time-tested example of the degree of availability, which P2P protocols can provide as a communication backbone. In such a network, the communication protocol in effect causes full or near-full data replicability to each node which is a less efficient process compared to the more traditional client-server way of computing. However, the independence of the nodes, participating in a P2P network, provides a robust operating environment by nearly removing any impact from separate nodes joining, quitting or losing connection with the network. Because of the lack of a central server, a P2P network requires more effort than other network topologies to be shut down. However, there is an inherent challenge to be dealt with when it comes to the synchronization activities undertaken by the participants taking part in P2P models. In the context of DLT, the problem of updating the data store at different speeds must be solved by introducing a mechanism, that as an output, constitutes an overall consensus about what is the single "real" data state at any given point in time. Moreover, this mechanism relies upon a pre-encoded set of rules, which when executed by every participant, as part of the protocol, resolves situations such as publishing multiple, potentially different in contents, updates from several nodes to the network.

2.1.2 Categorisation of Blockchain architectures

According to (Ølnes et al., 2017), one way of categorizing different types of blockchain architectures is in regards to two criteria (see Table 2.1.2.1): the requirement of a permission for a node to participate in the blockchain network as a validator and who is provided access to read and write data to it. Based on the

[7] Linked list is a linear data structure which in its basic form consists of separate elements referencing each other in a sequential fashion.
[8] BitTorrent - https://en.wikipedia.org/wiki/BitTorrent

first criterion, there can be differentiated two types of blockchains - a permissionless one, in which every participant that cares to be a validator can become one on a plug-and-play basis without asking for an explicit permission, and a permissioned one, in which there is an appointed central authority or a consortium of organizations that must provide authorization beforehand to each validator candidate. On the other hand, the second criterion describes a public blockchain to be one that has no restrictions on who is able to access the database for operations such as reading and writing. In contrast, the private version involves the formation of various groups, each with a specific set of permissions in regards to what activities are allowed for their participants in the network. In practice, the permissionless design features public no-restricted data access, as opposed to the permissioned variant wherein the access rights are distributed among a preselected group of nodes only. Due to the lack of need of authorisation, each participant that volunteers to become a validator node is able to provide their computational resources and start operating in the network by broadcasting valid transactions and verifying blocks, published by other validator nodes. In order to incentivise such behaviour among network participants, there must be a mechanism that rewards the validators' contribution, for instance by the issuance of new cryptographic scarce resources. Moreover, the nonexistence of a central party to keep track of users' asset balances calls for a distributed consensus algorithm that counteracts any adversarial attempt to tamper with ledger history, i.e. secure end users against double-spending attacks.

	Permissioned	**Permissionless**
Public	No restricted data access or transactions. Only a restricted set of nodes can participate in the consensus mechanism.	No restriction on access, transaction (data writing) or validation.
Private	Restricted access, data writing and validation. Only the owner determines who can participate.	Restrictions on access and who can transact. No restriction on participation in the consensus mechanism.

Table 2.1.2.1 Blockchain Categorization Matrix (Ølnes et al., 2017)

There are two families of techniques that are widely used in practice as distributed consensus algorithms, namely Proof-of-Work (PoW) and Proof-of-Stake (PoS).

PoW is leveraging computationally hard or memory hard mathematical problems that must be solved by a process called mining[9]. This technique relies heavily upon the computational power amount, measured in hashes per second, that is available to the network. As a side effect of the computational intensity involved, much electrical energy is consumed in the process. In case of a malicious attempt to changing the database history records, the adversary will need to deploy the same amount or more computational power in the network, so that they increase their chances to produce transaction blocks faster than the legitimate set of nodes. Eventually, this would lead to the production of a longer blockchain in terms of length and becoming the new "real" state of the database, proving the attack successful. On the other

[9] Mining is the process of solving computationally difficult problems in order to gain the privilege of validating transactions and creating new blocks. The problems are required to have the property of being difficult to solve, but trivial to be verified.

hand, Peercoin[10], described by (King & Nadal, 2012), is the first cryptocurrency that incorporated an alternative distributed consensus protocol called Proof-of-Stake[11]. It represents a security model whereby the creator of the next block is chosen based on the combination of randomization, wealth and coin age (1), i.e. the stake. The term 'stake' is used partly in a similar way as ownership. For example, in order to contribute and benefit from a firm's business operations, a stakeholder would need to obtain shares of it beforehand. Equally, the only way to participating in the future development of and profiting from a PoS-based network is to hoard the local to the network cryptocurrency.

$$Coin\ age = Number\ of\ coins\ \times\ Number\ of\ days\ coins\ being\ held \quad (1)$$

By utilizing a PoS-based blockchain design, the need for validators purchasing expensive and energy-intensive equipment is fully eliminated, thus proving the mechanism to be more energy efficient and environment-friendly compared to PoW. The rationale behind this mechanism is that instead of deploying scarce physical resources, validators who hold a stake in the network are inclined to behave correctly by preserving its security as, otherwise, their virtual assets will erode in value. However, (Bentov et al., 2016) discuss some inherent hurdles in pure PoS-based systems such as a fair initial distribution of the coin supply to all stakeholders, and network stability in case the majority of the nodes are behaving in a rational way rather than in an altruistic one.

Another possible variant of a blockchain is to be permissioned by design. Generally, this configuration is preferred over the permissionless one within individual companies and consortiums, wherein the group of validator nodes are preselected, and the addition of new ones must be agreed upon from a designated authority, which was given the power to do so. Due to the 'trusted' environment in which the nodes are operating in, they must be named and legally responsible for their actions. Usually, the data that is transacted on the network takes the form of digital representation of off chain assets such as securities, ownership titles, currency, and others. This type of blockchains provide the primary benefit of enabling business applications where the participants must be known and identifiable while not necessarily fully trusting each other. Due to its closed nature, permissioned networks are able to scale their computational capabilities in a convenient and structured manner by relying on an appointed authority to perform and ensure that all project requirements are fulfilled. More specifically, the controlling entities have the ability to plan and deploy necessary changes to the software and hardware network resources. There are further challenges in designing permissioned blockchains that have been addressed by (Vukolić, 2017). The author claims that incorporating mechanisms from permissionless into permissioned networks leads to operating limitations and are obsolete in a controlled context. However, decreasing the technology elements which provide the core features, i.e. lack of need for trusted intermediaries, would remove step-by-step the decentralised nature of blockchain, thus eliminating the properties of immutability and censorship resistance of data, as the control and power over the network is distributed among a small group of privately-owned full nodes. Another implication is the lack of need for economic incentives for validators as blockchain governing processes, in the case of consortiums, are performed by following a

[10] Peercoin - https://peercoin.net/
[11] Proof-of-Stake - https://en.wikipedia.org/wiki/Proof-of-stake

consortium consensus model (Valkenburgh, 2016). Some examples of permissioned blockchain networks include Hyperledger[12], Kadena[13], and Chain[14].

2.1.3 Smart contracts

(Szabo, 1994) introduces the concept of smart contracts that represents a computerized transaction protocol capable of executing encoded actions in computer code with regards to specific conditions, called contract terms. Furthermore, the paper contains highlights about several applications of the technology such as digital cash protocols, synthetic assets, and smart property, with the potential to minimize or completely remove the need of trusted intermediaries, as well as transaction costs.

The idea itself is finally materialized with the advent of blockchain development, whose combination is referred to as Blockchain 2.0. The self-enforcing property of smart contracts makes them an extremely attractive use case for automating transactions between engaging parties that must only occur if certain conditions are met.

When implemented on top of blockchain technology, the smart contracts are replicated and executed by each full node. Such contracts produce deterministic outputs as the network of validators come to a consensus on equal computed results, as opposed to data streams only in Bitcoin (Kosba et al., 2015). Two famous real-world examples of 'smart contract'-enabled blockchain platforms with Turing-complete programming languages, though still in experimental phase as of today, are Ethereum (Buterin, 2014) and EOS (Github, 2018).

Smart contracts can feature full-fledged computing capabilities such as loops, internal state, as well as being permanently stored on a network unless configured to self-destroy under specific circumstances. They are used as building blocks for implementing more complex services such as decentrantralized applications. The idea behind is to switch from the traditional trusted, centralised, cloud-based server architecture paradigm to a trustless, decentralized, blockchain-based one. The latter can leverage additional technologies such as IPFS[15] in order to store any voluminous files in a distributed manner.

2.2 Bitcoin

The concept of Blockchain was first introduced in the paper *Bitcoin: A Peer-to-Peer Electronic Cash System* by an anonymous person or group of people, which named themselves (Satoshi Nakamoto, 2009). The product of the paper is a digital payments ecosystem called Bitcoin[16], which is a peer-to-peer network version of electronic cash that allows participants to transact with each other without a central trusted authority. Instead of using a central party that mediates and keeps track of all transactions, Bitcoin incorporates Blockchain to secure payments against double-spending attacks. The system leverages a combination of PoW-based security model to prevent malicious behaviour together with the longest chain

[12] Hyperledger - https://www.hyperledger.org/
[13] Kadena - http://kadena.io/
[14] Chain - https://chain.com/
[15] IPFS - https://ipfs.io/
[16] Bitcoin - https://bitcoin.org/en/

rule[17] applied by mining nodes in order to reach an agreement on a specific version of transaction history that has been witnessed in the network (see Figure 2.2.1). One resulting security implication is that the system relies on the assumption of honest nodes controlling more computational power than any group of cooperating attacking nodes at any given point in the lifecycle of the system; a failure to do so would potentially lead to double-spending at the attackers' will.

A basic transaction model of transferal of coins between two participants is performed by combining the following virtual components:

- Receiver's public key. It represents the destination cryptographic address to which the ownership of the sent coins will be transferred, i.e. the new owner.
- Digitally signed transaction hash. It consists of a cryptographic hash of the sender's previous transaction, wherein the coins' ownership was transferred to themselves from another source, and the sender's private key.

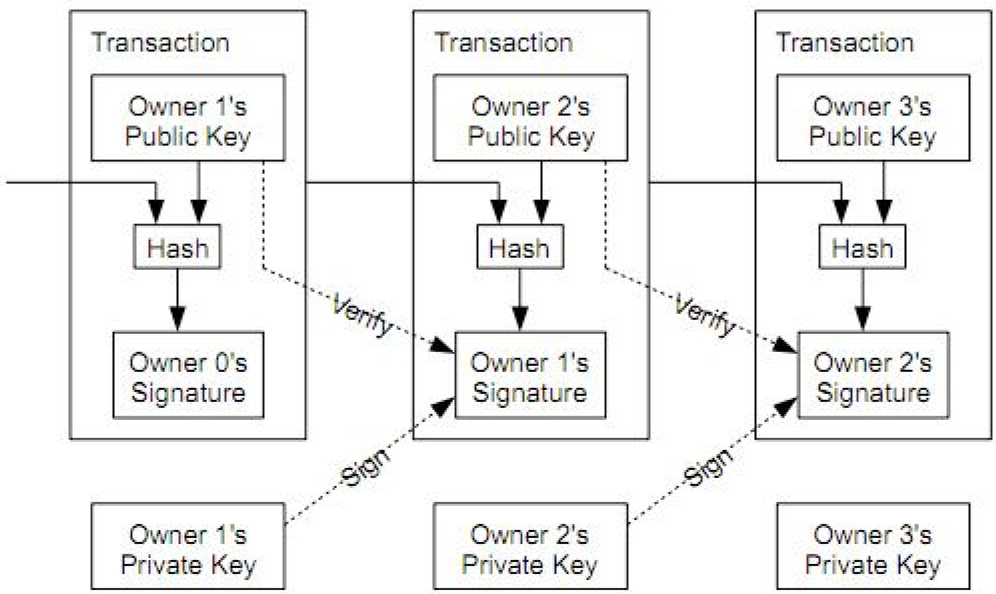

Figure 2.2.1 Bitcoin Transaction Model (Nakamoto, 2008)

Each mining node has at their disposition a copy of the ledger and validates each broadcast transaction that they are aware of so that only unspent coins can be transferred. At a regular time interval of approximately 10 minutes, the miners compete for the privilege of publishing a block of transactions, and get compensated in a certain amount of coins and transaction fees. Every new block appended to the blockchain contains the hash value of the previous block's contents (see Figure 2.2.2), thus reinforcing the chronological order of transactional data.

[17] Longest chain rule is part of the Bitcoin's consensus mechanism that in the event of multiple chain forks, the chain with the most blocks is considered to be the single source of truth, i.e. the only 'real' chain.

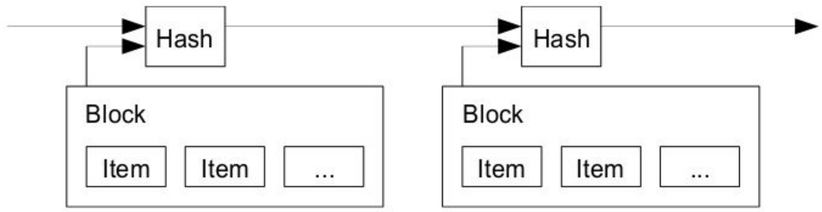

Figure 2.2.2 Bitcoin Block Model (Nakamoto, 2008)

In the case of multiple miners broadcasting their block to the rest of the network simultaneously, every other node receives them in some order and continue to work on the first one they were aware of, thus forming chain forks[18]. This temporary state inconsistency is trivially solved by the *longest chain rule* as all nodes are incentivized to choose the longest chain to produce new blocks on, otherwise, they would not be rewarded by the Bitcoin protocol.

Due to the permissionless nature of Bitcoin, a standardized change management process is established in the ecosystem that defines and structures the improvement cycle of the technology. There are two types of protocol rules that evolve - technical and business. The former provide software specification about data formats, necessary field inputs, and others, while the latter describe the economic constraints needed for the network to function as intended such as double-spending and to spend only coins that the user proves ownership of. The first step of the process is creating and submitting a Bitcoin Improvement Proposal (BIP) document, which describes an idea about a new feature or improvement change, for a discussion within the community. If it is approved, an integration of the corresponding code changes are published in the Bitcoin codebase, and a new version of the software is released for the miners to download and run.

Design aspect	Activity	Bitcoin
Data storage	Persisting data in the network	Blockchain
Data distribution	Synchronizing data between participants	Peer-to-peer
Consensus mechanism	Resolving conflicts/Dealing with chain forks	Longest chain rule
Upgrade mechanism	Changing protocol rules	BIPs and voting
Defence mechanism	Protecting the network against malicious behaviour	Proof-of-Work
Participation	Reading data	Everyone
Participation	Writing data/Submitting a transaction	Everyone
Participation	Validating data (transactions, blocks)	Everyone

[18] A fork represents a single version of the blockchain ledger. Occasionally, there are multiple versions of the ledger in existence due to mining nodes broadcasting new blocks to the network simultaneously.

Incentivisation	Incentivize block making	Block rewards and transaction fees

Table 2.2.1 Bitcoin ecosystem design. Table design and data readjusted from source: https://bitsonblocks.net/2015/09/09/a-gentle-introduction-to-blockchain-technology/

2.3 Cryptocurrencies

At the time of this writing, there are over 2000 different cryptocurrency projects, listed on CoinMarketCap. All of them, except Bitcoin, are referred to as alternative cryptocurrencies, or altcoins. Although Bitcoin has proven its capabilities to be a sustainable cryptocurrency throughout the years since its inception in 2009, there are inherent trade-offs within its design concering security, scalability, and degree of decentralization. A specific combination of adjustments to the latter three properties serves as a pillar around which the software design of each altcoin project has been chosen. In addition, the possibility of adding other unique features provides the opportunity to other-than-Bitcoin projects to experiment further and search for a more optimal solution, e.g. Litecoin and Zcash, or specialise in a specific use case such as creating a decentralized application platform - Ethereum and EOS. Furthermore, another possible classification of cryptocurrencies is by the purpose of the digital currency. The first type, and the largest one, are called utility tokens, which consumers pay to a payee in order to receive a specific good or service on the blockchain. Following, security tokens are used to represent ownership of physical or digital assets, including a special subtype of equity tokens that could increase the market liquidity of private company shares at the time of their issuance, compared to contemporary practices where waiting for an initial public offering is considered the way to go in order for early investors to materialize profits.

A significant reason for the exponential expansion of the overall market in the recent years is in part due to the development of various custodial exchanges, which fulfill the function of a gateway between two economic systems - traditional and cryptocurrency. The significant amount of 99% of the total trading volume is estimated to be processed by centralized exchanges - both custodial and non-custodial. This fact has in effect turned the exchanges into the backbone of the cryptocurrency market. This type of digital exchanges is referred to as ‘centralized’ as there is an actual registered company which supports the business operations and is responsible for acquiring all necessary licences, as well as comply with regulations. However, there have been quite many occasions of hacked exchanges, the most notable ones being the Mt. Gox (Pollock, 2018) and Coincheck (Bloomberg, 2018), and scams that resulted in a loss of large amounts of funds and trust in the ecosystem. (Vasek and Moore, 2015) conduct thorough research of various types of scams in the cryptocurrency space totalling in more than USD 10 million loss, from which the exchange-related account for some USD 500,000. Nevertheless, since the research had been done in 2015 when the magnitude of the market capitalization was two times smaller, the revenues of the malicious actors behind the scams who had been holding the stolen funds for the next few years would be a lot higher. In order to address similar concerns in the future, with the advances in the development of cryptocurrency platforms, the implementation of decentralized open-source exchanges has become possible. The latter represent autonomously functioning applications on a blockchain, which consist of a set of publicly accessible smart contracts that perform the matching of buy and sell orders. Additionally, the technology removes the need of a central entity to maintain and be responsible for the security of

users' assets. Some examples of operating popular decentralized exchanges are IDEX[19] and Waves DEX[20]
.

In the work of (Osterrieder et al., 2016), by studying the volatility of a few of the biggest cryptocurrencies an extreme value analysis has shown that the risks for investors are higher than any other traditional market such as fiat currencies and commodities. However, as part of the research methodology, correlation matrices have been computed on top of projects' returns without placing any information about their stationarity status. By skipping a step whose purpose is to reduce the risk of finding spurious correlations (Yule, 1926), could introduce noise in the final results which are based on the actual correlation coefficients. This same problem persists in a study by (Gandal & Halaburda, 2016). The correlations between crypto assets have been explored, but there is missing evidence about the stationary status of the underlying time series data. Also, the time period that has been explored in it is May 2013 until July 2014, which could render their conclusions obsolete today due to the immaturity of the market in the past and its dynamics over time.

Another study (Garcia et al., 2014) examines the association between social activity and price of Bitcoin, being the market leading cryptocurrency. The results consist of two positive feedback loops - social and user adoption - by conducting an analysis of data from social media platforms, Google trends engine, Bitcoin exchange platforms and user base. Essentially, the causal relationship starts with an increase in the number of users, which is positively correlated to Bitcoin popularity, in turn, increasing the number of searches for Bitcoin. Regardless, the occurring negative price changes are not explained by this research. Additionally, (Kristoufek, 2013) identifies a bidirectional relationship between price level and Internet search queries, specifically in Google Trends and Wikipedia.

As a result of the literature analysis above, one can conclude that there is a research gap of finding significant relationships, which are rigorously tested in the context of statistical properties. To the best of the author's knowledge, there was no event study explicitly conducted on Bitcoin and Ethereum combined. Thus, the following hypotheses are formed in relation to the research questions:

- Hypothesis 1
  There is no association between individual cryptocurrencies and the overall market.

- Hypothesis 2
  There is no association between individual cryptocurrencies and Bitcoin.

- Hypothesis 3
  There is no association between the top 20 cryptocurrencies by market capitalization.

- Hypothesis 4
  Events do not have an impact on the price levels of Bitcoin and Ethereum.

[19] IDEX - https://idex.market/
[20] Waves DEX - https://wavesplatform.com/product/dex

Chapter 3: Correlation analysis

3.1 Methodology

As an industry expert, (Rollins, 2015) outlines key reasons for the need of an appropriate strategy in executing data analytics tasks. The lack of such is a major cause for failure to arrive at solutions that address the right problem, as well as obtain a full understanding of its context. Following a proper methodology is a necessity in terms of producing an output which effectively solves the problem at hand in a timely manner.

Two process frameworks for data analysis, compared by (Azevedo et al., 2008), are considered during the initial stage of this paper with the final aim being to provide the researcher with guidelines and hints on ways to produce a systematic and reproducible report.

The first option to consider was SEMMA[21], developed by SAS Institute[22], which represents a sequential list of steps to guide the implementation of data mining applications. While it is known to be a general analysis methodology by the community, SAS refers to it as a "logical organization of the functional toolset of" one of their products SAS Enterprise Miner[23], which could potentially lead to ambiguous situations if applied in a different business context (Rohanizadeh et al., 2009). Furthermore, it is criticised for only focusing on the technical aspects and phases of a project life-cycle, and not covering the business ones (Azevedo et al., 2008).

Alternatively, CRISP-DM[24], developed under the ESPRIT funding programme in 1997, is an open standard leading methodology among industry data miners (Kdnuggets.com, 2014). It describes common activities that are performed to solve business problems, requiring data analytical procedures. One major advantage of the framework is that it allows for iterative processing, thus no strict order of tasks must be followed. Being an industry-agnostic methodology enables its application into various environments, which could benefit from the structured nature of planning activities related to data analysis (Mariscal, 2010). The framework is divided into two implementations - reference and user guide. The former provides a general overview of what phases, tasks and outputs should be elaborated on during the analysis process and, in general, answers to the question 'What?', while the latter describes the ways on how the actual activities should be conducted. Both implementations comprise the following phases: Business understanding, Data understanding, Data preparation, Modeling, Evaluation, Deployment (CRISP-DM,

---

[21] SEMMA (Sample, Explore, Modify, Model, and Assess) - http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbja4n1xueveo2uoujy.htm

[22] SAS Institute - https://www.sas.com

[23] SAS Enterprise Miner - https://web.archive.org/web/20120308165638/http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html/

[24] CRISP-DM (Cross-Industry Process For Data Mining) - https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

2000). By means of arrows, Figure 3.1 exhibits the most frequently occurring dependencies among different phases, as well as the logical transitions between them.

For the purposes of this research, the CRISP-DM framework is chosen to be followed.

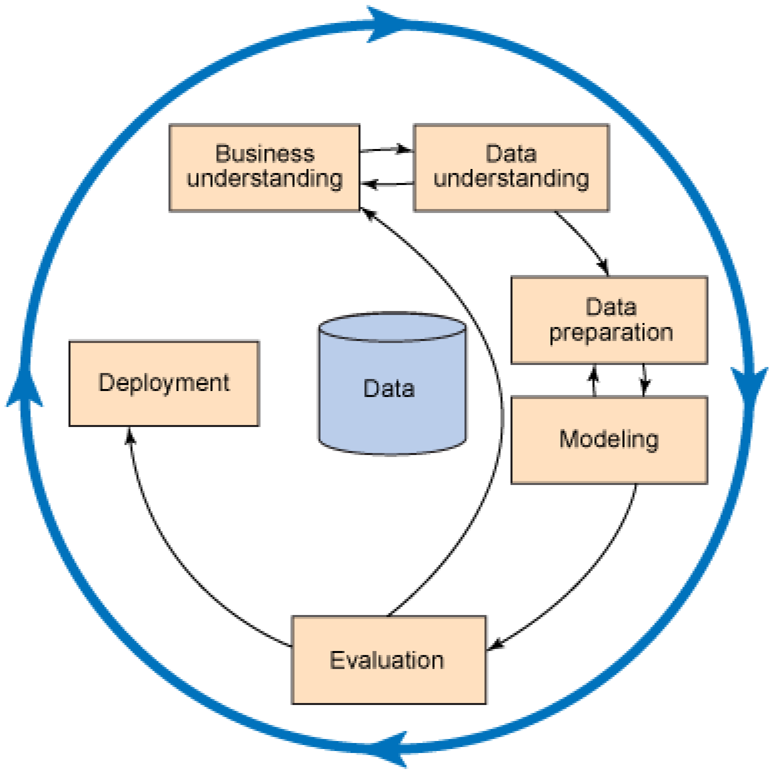

Figure 3.1 Phases of the CRISP-DM reference model (CRISP-DM, 2000)

3.1 Business understanding

Like any other financial market, the cryptocurrency one is no different in terms of investors searching for and applying custom trading strategies in order to profit from the price volatility of various assets. The combination of currently owned and managed by an investor various assets are contained in a set, called a portfolio. The issue of constructing and managing a portfolio with maximized reward and minimized risk at every point in time is firstly presented by (Markowitz, 1952). In his work, Markowitz discusses the general objective of each investor to maximize return for any level of risk, as well as to create a diversified portfolio of unrelated assets in order to reduce risk. The 'return' can be defined as the difference in the total price of the portfolio at two points in time plus any other financial and non-financial benefits generated from holding these assets, e.g. dividends. Following, each possible portfolio with potential reward, has some level of risk associated with it. 'Risk' is generally evaluated by calculating the standard deviation of a particular asset, thus giving an estimate of what the most likely price swings are. However, in the Markowitz' theory, it is stated that the risk of an asset should be measured not in isolation, but in combination with each other asset's contribution to the overall portfolio risk (Mangram, 2013). The term 'risk' refers to both systematic (common) and unsystematic (diversifiable) types of risk

("Lowering Portfolio Risk through Diversification", 2018). The former is said to have a significant influence over virtually all securities as it comprises macro-level factors such as inflation, unemployment rate, and interest rates (Jaffe et al., 2004). It can be dealt with by utilizing a hedging[25] strategy, which involves owning negatively correlated assets. On the contrary, the second type of risk is a micro-level risk that is tightly connected to a specific asset or group of assets. Although it is impossible to be fully eliminated, the influence of this threat could be substantially reduced by applying portfolio diversification, i.e. choosing different assets such that are uncorrelated to each other.

The scope of this research is limited to provide cryptocurrency relationship insights that tackle the unsystematic risk related to investing in the market. With the identification of the strength and direction of the correlation between any two cryptocurrency projects, a set of optimal portfolios could be crafted and eventually be filtered further according to an investor's subjective criteria.

3.2 Data understanding

The main purpose of this phase is to list and describe all data sources that have been used, as well as to illustrate activities which are part of the employed data collection process. Furthermore, a general perspective on the essence of data at hand is presented. Identification of quality issues and initial insights are also included.

3.2.1 Data collection

The Coinmarketcap (CMC) platform is used as a primary source of historical data. It represents a community-supported web service designed with the special purpose of tracking mainly financial information about cryptocurrency projects that cover low-barrier requirements such as being traded on at least one public exchange and have a non-zero trading volume. Due to price differences among various exchanges, CMC calculates each cryptocurrency's price by taking a volume-weighted average.

The actual data collection process is executed by means of utilizing third-party software libraries[26], whose objective is to extract raw data from CMC via API or scraping techniques. Consequently, some filtering criteria (see Table 3.2.1.1) are applied, so that the working set of data is aligned with the scope of this research paper.

#	Criterion
Comment	
1.	A dataset with the top 150 (one hundred and fifty) cryptocurrencies by market capitalization as of date 07.07.2018 are extracted.

[25] Hedging, in the financial services industry, is an investment strategy to reduce the overall risk of an investment or a portfolio.
[26] Coinmarketcappy (https://github.com/saporitigianni/coinmarketcappy) is an open source Python library, which implements an API client for Coinmarketcap's historical data endpoint.

The market capitalization serves as a popularity indicator. Naturally, cryptocurrencies with higher market capitalization imply higher trading volume and liquidity.	
2.	A dataset with the total cryptocurrency market capitalization is extracted.
It contains the historical records of the total value by summing up the market capitalizations of all separate projects, listed at that point in time on CMC.	
3.	The historical financial data is time-framed in the interval between 06.07.2015 and 07.07.2018 (dd/mm/yyyy).
The analysis time span is over the last 3 (three) years of market history only.	
4.	Each cryptocurrency must have at least 90 (ninety) days of historical records.
This criterion is set in order to tackle the possibility of a newly listed project on CMC to obtain a high market capitalization unnaturally due to manipulations or software bugs causing data inconsistencies.	
5.	The financial data for each dataset is collected on a weekly basis. For example, a project has 4 (four) or 5 (five) snapshots of data per month.
The choice of such time period aggregation leads to being able to process a vast amount of data for more extended periods in an efficient manner without losing representativeness of the sample.	

Table 3.2.1.1 Filtering Criteria

3.2.2 Data format

The data format in which the cryptocurrency information is structured is shown in Table 3.2.2.1, whereas Table 3.2.2.2 is related to the total market capitalization dataset.

Property	Description
Date	The point in time when the data snapshot was taken.
Name	The formal name of a cryptocurrency project.
Symbol	The ticker symbol uniquely identifies a cryptocurrency project.
Price	The current volume-weighted average market price.
Circulating Supply	The total amount of assets in circulation.
Market Cap	The market capitalization, calculated by ‘Price’ x ‘Circulating Supply’.

Table 3.2.2.1 Record format of cryptocurrency dataset

Property	Description
Date	The point in time when the data snapshot was taken.
Value	The total market capitalization.

Table 3.2.2.2 Record format of TMC dataset

3.3 Data preparation

This section discusses all preliminary manipulations of the raw dataset that aim for facilitating further analysis actions. As a guiding reference in the process, (Wickham, 2014)'s definition of tidy data is used, along with accompanying techniques to convert messy data into well-structured physical representation supporting the data semantics. The expected output is a completely prepared dataset to be fed into the Modeling phase.

3.3.1 Establishment of a one-to-one relationship

In order to provide meaningful results, a data consistency check was performed to ensure the existence of a one-to-one relationship between the *Symbol* and *Name* attributes throughout the entire dataset. In other words, a guarantee must be established that one and only one project can be uniquely identified by a specific *Symbol*. This operation mitigates the risk of an analysis to be thought to be conducted on a certain cryptocurrency project, whereas in reality that data comprise other cryptocurrency projects as well, leading to wrong conclusions. Initially, the problem at hand was thought to be trivially solvable by getting CMC data through filtering the results by both top 150 cryptocurrency symbols and names, however, it was noticed later on that some projects had been rebranded at some point in time, effectively changing their original name. Therefore, the only viable option for cleaning the inconsistent records was through manual analysis of all different groups of projects, sharing a common *Symbol*. Table 3.3.1 consists of cryptocurrency groupings which violate the one-to-one mapping and need to be handled further. It serves as a decision-making base on which projects should persist in the database and which ones should be discarded. As a solution, an approach that relies upon the knowledge base of three widely-used and time-proved different Internet services was undertaken: CMC, Coingecko[27] and Cryptoslate[28]. Table 3.3.2 illustrates the results from the analysis. Specifically, the column "To be preserved" contains the project that should remain, while the column "To be removed" designates the projects that should be filtered out. Additionally, Table 3.3.3 comprise projects that have undergone the process of rebranding. Thus, the only necessary operation to make the data consistent is renaming to their old or new name.

[27] Coingecko - https://www.coingecko.com/en
[28] Cryptoslate - https://cryptoslate.com/

Symbol	Name	Number of observations
bat	basic attention token	57
bat	batcoin	97
blz	blazecoin	82
blz	bluzelle	21
bnb	binance coin	49
bnb	bnb coin	6
bnt	bancor	54
bnt	bantam	50
btg	bitcoin gold	36
btg	bitgem	101
btm	bitmark	157
btm	bytom	47
cmt	comet	105
cmt	cybermiles	30
drop	dropil	15
drop	faucetcoin	29
etc	ethercoin	7
etc	ethereum classic	101
gtc	game.com	26
gtc	global tour coin	51
hot	holo	9

Symbol	Name	Number of observations
hot	hydro protocol	24
icn	icoin	64
icn	iconomi	92
icx	icon	36
icx	icon [futures]	2
knc	khancoin	62
knc	kingn coin	52
knc	kyber network	40
nas	nas	36
nas	nebulas	31
poly	polybit	6
poly	polymath	22
smart	smartbillions	11
smart	smartcash	50
sub	subcriptio	10
sub	substratum	40
xin	icoin	1
xin	infinity economics	39
xin	mixin	23
xrp	ripple	156
xrp	xrp	1

Table 3.3.1 Cryptocurrencies that violate the one-to-one relationship between Symbol and Name attributes

#	Symbol	To be preserved	To be removed	Cryptoslate	Coingecko	Coinmarketcap

1	**bat**	basic attention token	batcoin	Link	Link	Link
2	**blz**	bluzelle	blazecoin	Link	Link	Link
3	**bnt**	bancor	bantam	Link (no results)	Link (no results)	Link
4	**btg**	bitcoin gold	bitgem	Link	Link	Link
5	**btm**	bytom	bitmark	Link	Link	Link
6	**cmt**	cybermiles	comet	Link	Link	Link
7	**drop**	dropil	faucetcoin	Link (no results)	Link (no results)	Link
8	**etc**	ethereum classic	ethercoin	Link	Link	Link
9	**gtc**	game.com	global tour coin	Link	Link	Link
10	**hot**	holo	hydro protocol	Link	Link	Link
11	**icn**	iconomi	icoin	Link	Link (no results)	Link
12	**icx**	icon	icon [futures]	Link	Link	Link
13	**knc**	kyber network	khancoin kingn coin	Link (no results) Link	Link (no results) Link	Link Link
14	**nas**	nebulas	nas	Link (no results)	Link (no results)	Link
15	**poly**	polymath	polybit	Link (no results)	Link (no results)	Link
16	**smart**	smartcash	smartbillions	Link (no results)	Link	Link
17	**sub**	substratum	subcriptio	Link (no results)	Link (no results)	Link (no results)
18	**xin**	mixin	icoin infinity economics	Link Link	Link (no results) Link	Link Link

Table 3.3.2 Evaluation of cryptocurrency projects - remaining and discarded

19	**xrp**	xrp	ripple	Link Link	Link	Link
20	**bnb**	binance coin	bnb coin	Link	Link	Link

Table 3.3.3 Evaluation of cryptocurrency projects - merged

3.3.2 Removal of cryptocurrency projects lacking enough data

As listed in the criteria table (see Table 3.2.1.1), projects with less than 12 (twelve) data snapshots are to be dropped from the dataset. After applying the filter, the entities in Table 3.3.2.1 are removed.

#	Symbol	Name	Number of total observations
1	bft	bnktothefuture	9
2	ctxc	cortex	11
3	cvt	cybervein	11
4	hot	holo	9
5	ht	huobi token	10
6	moac	moac	5
7	mtc	docademic	8
8	nexo	nexo	8
9	ode	odem	10
10	tpay	tokenpay	11
11	wicc	waykichain	9
12	xin	mixin	10

Table 3.3.2.1 Evaluation of cryptocurrency projects - lack of enough records

3.4 Modeling

This phase concentrates on analyzing and solving data issues related to its nature and specific domain, with the end goal of providing a systematic approach to answer the research questions.

3.4.1. Stationary and non-stationary time series

Due to the time series[29] (TS) nature of the collected data, there are inherent issues to be addressed, so that the risk of finding spurious Pearson correlations (Yule, 1926) is eliminated. Firstly, the linear regression assumption of observations being independent is violated due to time-factor dependability, and secondly, the potential existence of trends and seasonality cycles in markets data imply varying statistical moments such as mean and variance. TS which incorporate the previously stated characteristics are called non-stationary (stochastic or random walk processes), and are known to have statistical moments,

[29] Time series is a sequence of data points, which are gathered at a regular time frame.

converging to integrals of Brownian motion, i.e. random variables. Since the Pearson correlation coefficient is calculated as a number, the need of working with stationary TS is of utmost importance - their sample moments probabilistically converge to constants. Therefore, each TS dataset must be checked for stationarity first in order to be able to find sensible relationships with other. In the event of finding a TS to be non-stationary, it must undergo a process of stationarization[30] before any further analysis.

A stationary time series is one that does not change its underlying distribution caused by a time shift. For the purposes of concept operationalization, a TS is considered to be stationary if it has constant mean and variance over time:

$$\mu(t) = \mu$$
$$\sigma^2(t) = \sigma^2$$

Given the number of datasets to be analysed, visual inferences about whether a TS is stationary or not by observing every dataset's graphical plot is not convenient and error-prone. Conversely, a combination of statistical tests will be used, following any mathematical transformations if found to be of necessity. Taking into account that there are generally two types of non-stationary time series - 'trend-stationary' and 'unit root-stationary' - some statistical tests will need to be used to determine the TS stationarity status.

The Augmented Dickey-Fuller (ADF) (Dickey and Fuller, 1981) test is a statistical test that is employed in order to check for the potential existence of unit root in univariate processes with autocorrelation. It is an improvement over the standard Dickey-Fuller test, invented by (Dickey & Fuller, 1979), as any serial correlation is handled internally, thus it applies to more complex models. The associated problem of choosing a 'right' maximum lag value is dealt with by using (Schwert, 2002)'s suggestion, where T is the number of observations:

$$\rho_{max} = 12(T/100)^{1/4}$$

As part of research for selection a unit root test on the basis of TS length, (Fedorová, 2016) argues that ADF is the most accurate one for short-length datasets. The test's null hypothesis is that the TS under observation has a unit root, i.e. is not stationary, and its results consist of a test statistic, which is to be compared with a preliminary chosen by the researcher critical value. The choice of it depends on the desired level of confidence. If the test statistic is less than the critical value, the null hypothesis is rejected, therefore proving the stationarity of the TS. Nonetheless, ADF's primary disadvantage is that it has moderately high Type I error rate.

[30] Stationarization is the process of converting non-stationary time series into stationary ones, for example, by applying a mathematical transformation.

3.4.2 Methodology

In order to ensure the stationarity characteristics of all datasets, a systematic approach, shown in Table 3.4.2.1 and Table 3.4.2.2, is crafted and employed into action. It comprises a sequential order of steps to be executed on each dataset prior to analyzing interrelations and making inferences.

With the objective of transforming a non-stationary TS into a stationary one in mind, a reduction of the influence of potential trends and seasonality components in the data is needed. Taking into account that the data describes market information of cryptocurrencies, the researcher makes the assumption of the underlying statistical distributions to be non-normal. Therefore, the natural logarithmic function will be applied on the TS with the goal of transforming the distributions closer to Gaussian ones, thus minimizing the effect of outliers, as well as reducing unequal variances.

A suggested technique by (Granger & Newbold, 1974) for eliminating autocorrelation caused by trends and seasonality is calculating the first differences between TS observations over a specified time range *t*:

$$X_{diff} = X_t - X_{t-1}$$

Step	Description	Results: stationary/non-stationary
1	Conduct ADF test on raw cryptocurrency projects' pricing data.	22 / 115
2	Apply logarithmic transformation on the non-stationary subset. Conduct ADF test.	13 / 102
3	Apply 'first differences' transformation on the non-stationary subset. Conduct ADF test.	75 / 27
4	Apply 'second differences' transformation on the non-stationary subset. Conduct ADF test.	10 / 17
5	Apply 'third differences' transformation on the non-stationary subset. Conduct ADF test.	2 / 15

Table 3.4.2.1 Executed methodology actions on cryptocurrencies dataset

Step	Description	Results
1	Conduct ADF test on raw total market capitalization data.	non-stationary
2	Apply logarithmic transformation. Conduct ADF test.	non-stationary
3	Apply 'first differences' transformation. Conduct ADF test.	stationary

Table 3.4.2.2 Executed methodology actions on TMC dataset

3.5 Results

At this stage of the exploration process, interrelationships between entities of interest are measured and presented. Pearson correlation coefficients (Rodgers & Nicewander, 1988) lower than *-0.20* or higher than *0.20* are considered to designate existing relationship, while the rest - lack of one. For the purposes of significance testing, the Student's two-sided t-test is used by leveraging a third party Python3 library ("statsmodels.stats.weightstats.ttest_ind").

3.5.1 Relationships between individual cryptocurrencies and the cryptocurrency market

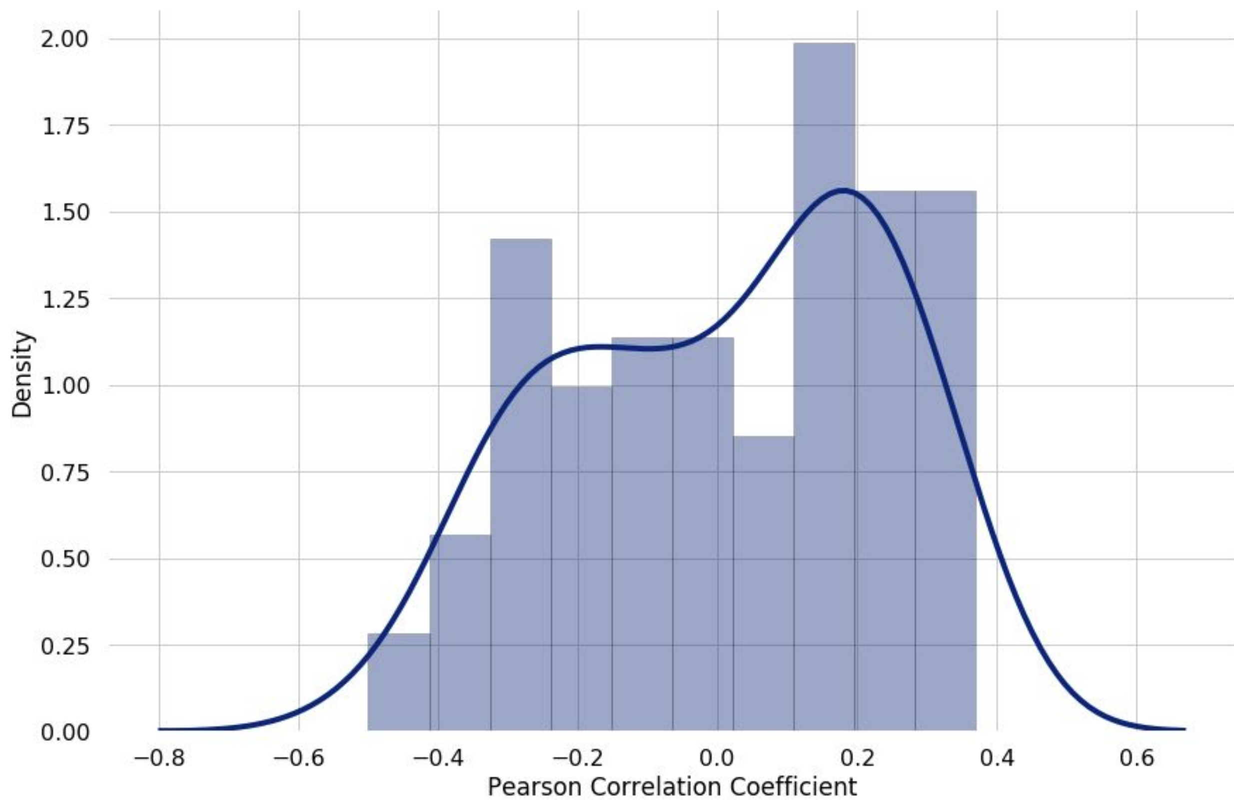

Figure 3.5.1.1 Density plot of correlations between Total Market Capitalization and cryptocurrencies

Total analysed	Total significant	Mean	Std	Min	Max
122	**81**	**0.01**	**0.23**	**-0.50**	**0.37**

Table 3.5.1.1 Descriptive statistics of correlations between Total Market Capitalization and cryptocurrencies

	Positively correlated				Negatively correlated			
Symbol	storm	bts	btc	mco	wax	bix	etn	mkr
Name	Storm	BitShares	Bitcoin	Monaco	WAX	Bibox Token	Electron-eum	Maker
Corr.	**0.37**	**0.36**	**0.35**	**0.32**	**-0.35**	**-0.39**	**-0.42**	**-0.50**

Table 3.5.1.2 Top 8 positively and negatively correlated coins to the overall market

Symbol	mds	payx	bch	rdd	grs	aion	btcp	iost
Name	MediShares	Paypex	Bitcoin Cash	ReddCoin	Groestlcoin	Aion	Bitcoin Private	IOST
Corr.	**0.03**	**0.01**	**0.01**	**0**	**-0.02**	**-0.02**	**-0.03**	**-0.04**

Table 3.5.1.3 Top 8 uncorrelated coins to the overall market

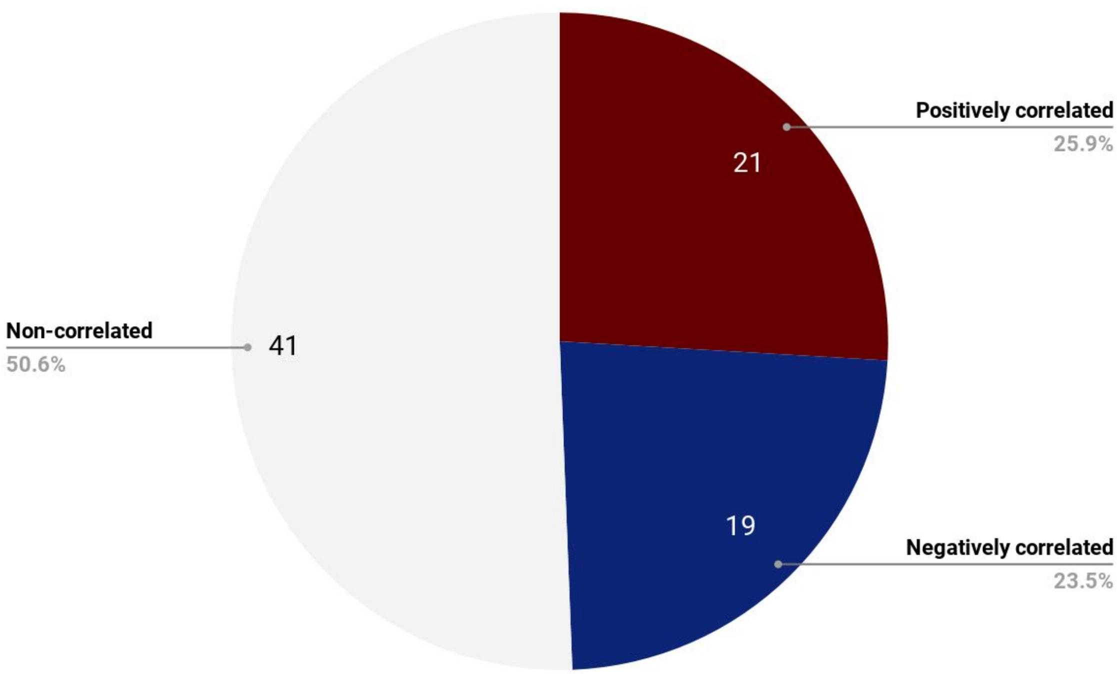

Figure 3.5.1.2 Distribution of correlations between Total Market Capitalization and cryptocurrencies

Figure 3.5.1.1 provides a general overview of the analysis as it presents the frequency distribution of the correlation coefficients. In order to get a more detailed picture, Table 3.5.1.1 provides descriptive statistics where the extrema and variation can be observed. Also, it is of importance to take notice at the first two columns: the first one represents the total quantity of cryptocurrencies, whereas the second one shows that amount of them which have been found to have a significant correlation with the overall market by conducting a standard two-tail t-test at the confidence interval of 5%. As a result, *Hypothesis 1* is rejected. Furthermore, Table 3.5.1.2 illustrates 8 projects with the most substantial positive and negative relationship to the market, compared to Table 3.5.1.3 where the coins with the lowest coefficients can be found. Figure 3.5.1.2 supports the view that in general there is 50% chance for a project to be uncorrelated to the market's movement, while there is 25% chance to be positively or negatively linked to it.

3.5.2 Relationships between individual cryptocurrencies and Bitcoin

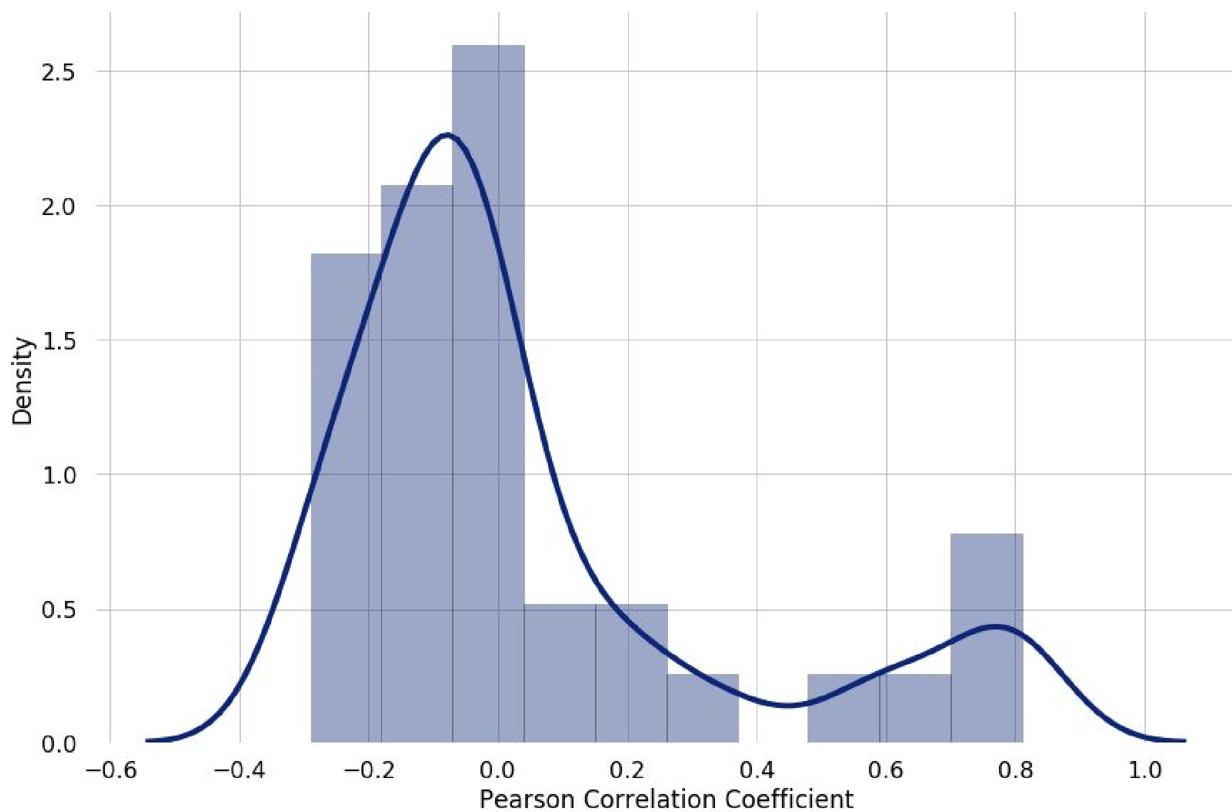

Figure 3.5.2.1 Density plot of correlations between Bitcoin and cryptocurrencies

Total analysed	Total significant	Mean	Std	Min	Max
122	**35**	**0.05**	**0.31**	**-0.29**	**0.81**

Table 3.5.2.1 Descriptive statistics of correlations between Bitcoin and cryptocurrencies

	Positively correlated				**Negatively correlated**			
Symbol	gto	dbc	icx	hpb	smart	nebl	eng	ethos

Name	Gifto	Deep-Brain Chain	ICON	High Performance Blockchain	SmartCash	Neblio	Enigma	Ethos
Corr.	**0.81**	**0.8**	**0.76**	**0.65**	**-0.25**	**-0.25**	**-0.28**	**-0.29**

Table 3.5.2.2 Top 8 positively and negatively correlated coins to Bitcoin

Symbol	mkr	zec	nas	gbyte	bix	wax	nxs	gas
Name	Maker	Zcash	Nebulas	Byteball Bytes	Bibox Token	WAX	Nexus	Gas
Corr.	**0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	**-0.03**	**-0.04**	**-0.06**

Table 3.5.2.3 Top 8 uncorrelated coins to Bitcoin

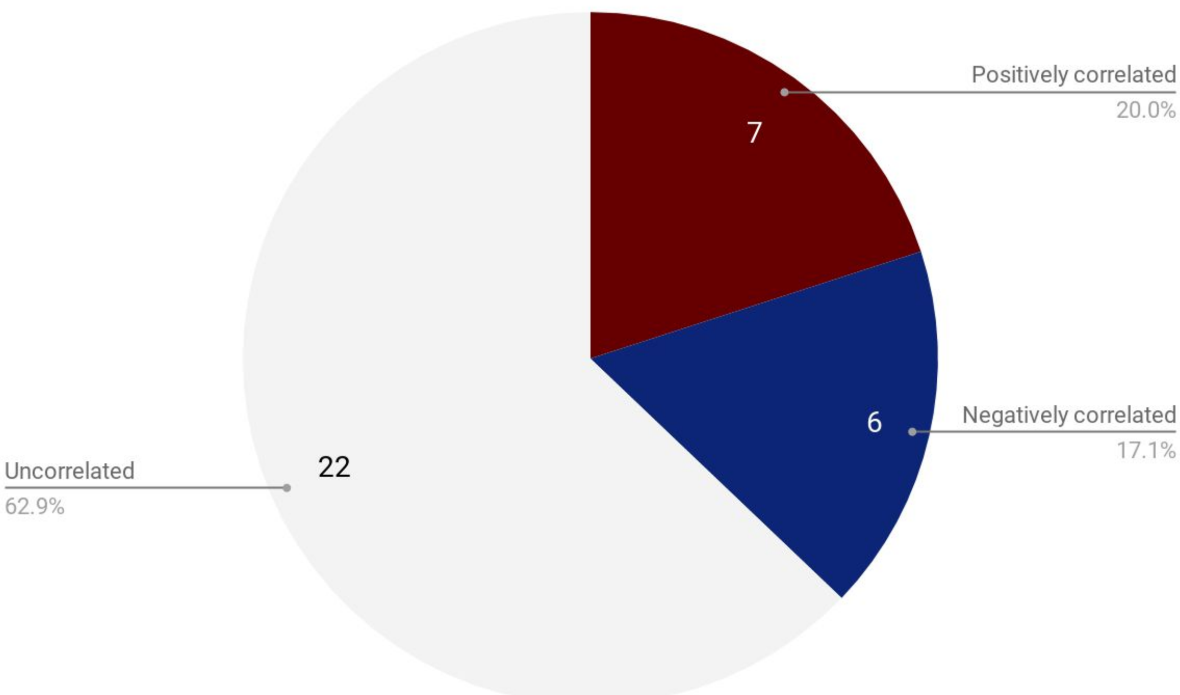

Figure 3.5.2.2 Distribution of correlations between cryptocurrencies and Bitcoin

The same analysis steps as in the previous section are followed, with Bitcoin being used as an anchor. As shown in Figures 3.5.2.1 and 3.5.2.2, the majority of the coins are found to be uncorrelated to Bitcoin. Details about which coins are found to have the weakest link are listed in Table 3.5.2.3. However, it is safe to reject *Hypothesis 2* as there is a subset that exhibits strong 'follower' behaviour with a maximum value of .81, a sample of which can be observed in the 'positively correlated' section in Table 3.5.2.2.

Furthermore, the negatively correlated subset, which is similar in size to the positively correlated one, exhibits a weak relationship with a minimum value of -.29.

3.5.3 Relationships between the top 20 cryptocurrencies

1	btc	Bitcoin	**11**	trx	TRON
2	eth	Ethereum	**12**	xmr	Monero
3	xrp	XRP	**13**	dash	Dash
4	bch	Bitcoin Cash	**14**	etc	Ethereum Classic
5	eos	EOS	**15**	xem	NEM
6	ltc	Litecoin	**16**	bnb	Binance Coin
7	xlm	Stellar Lumens	**17**	ven	VeChain
8	ada	Cardano	**18**	omg	OmiseGO
9	miota	IOTA	**19**	qtum	QTUM
10	neo	NEO	**20**	zec	Zcash

Table 3.5.3.1 Top 20 coins by market capitalization

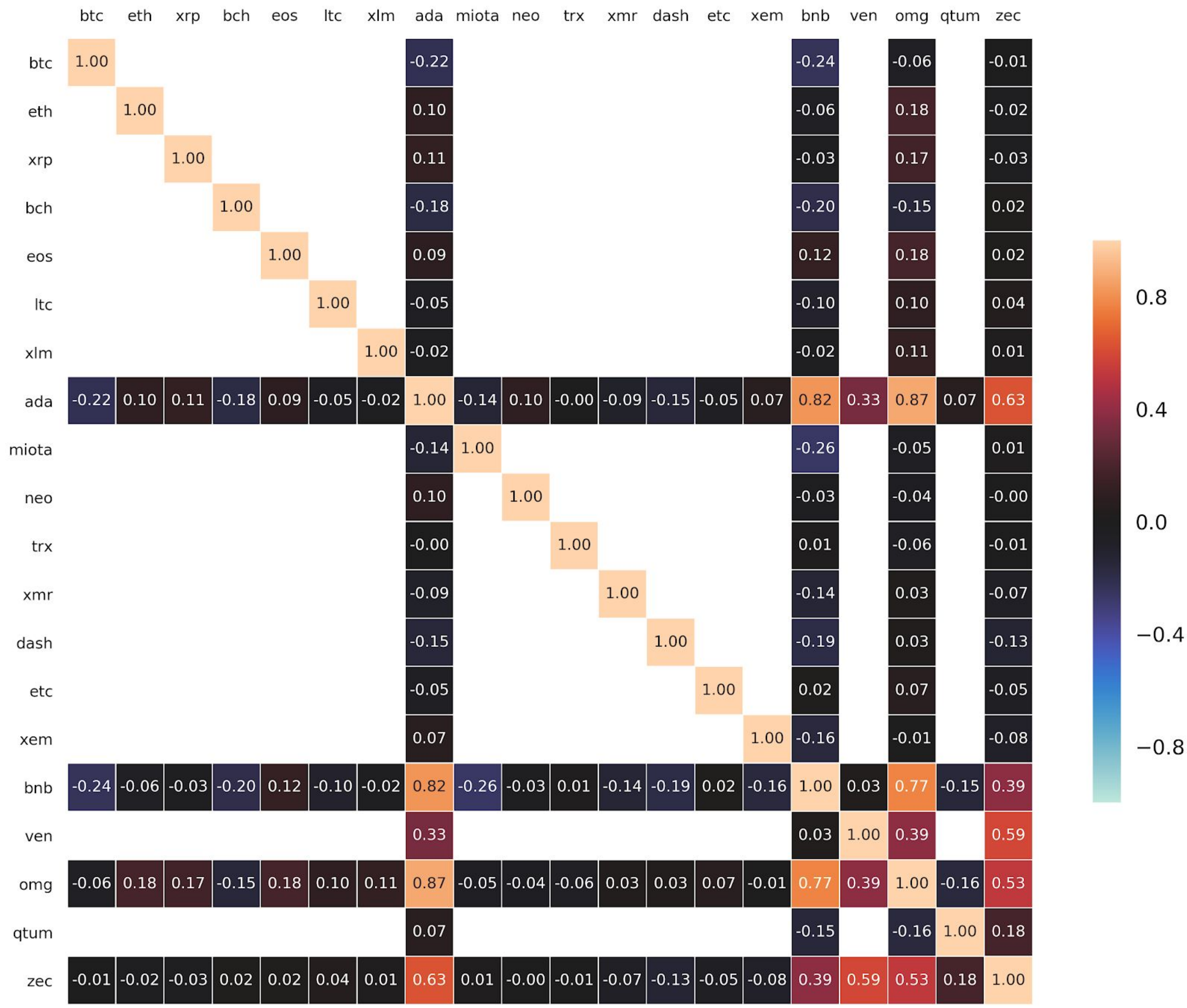

Figure 3.5.3.1 Heatmap of the correlations between the top 20 cryptocurrencies

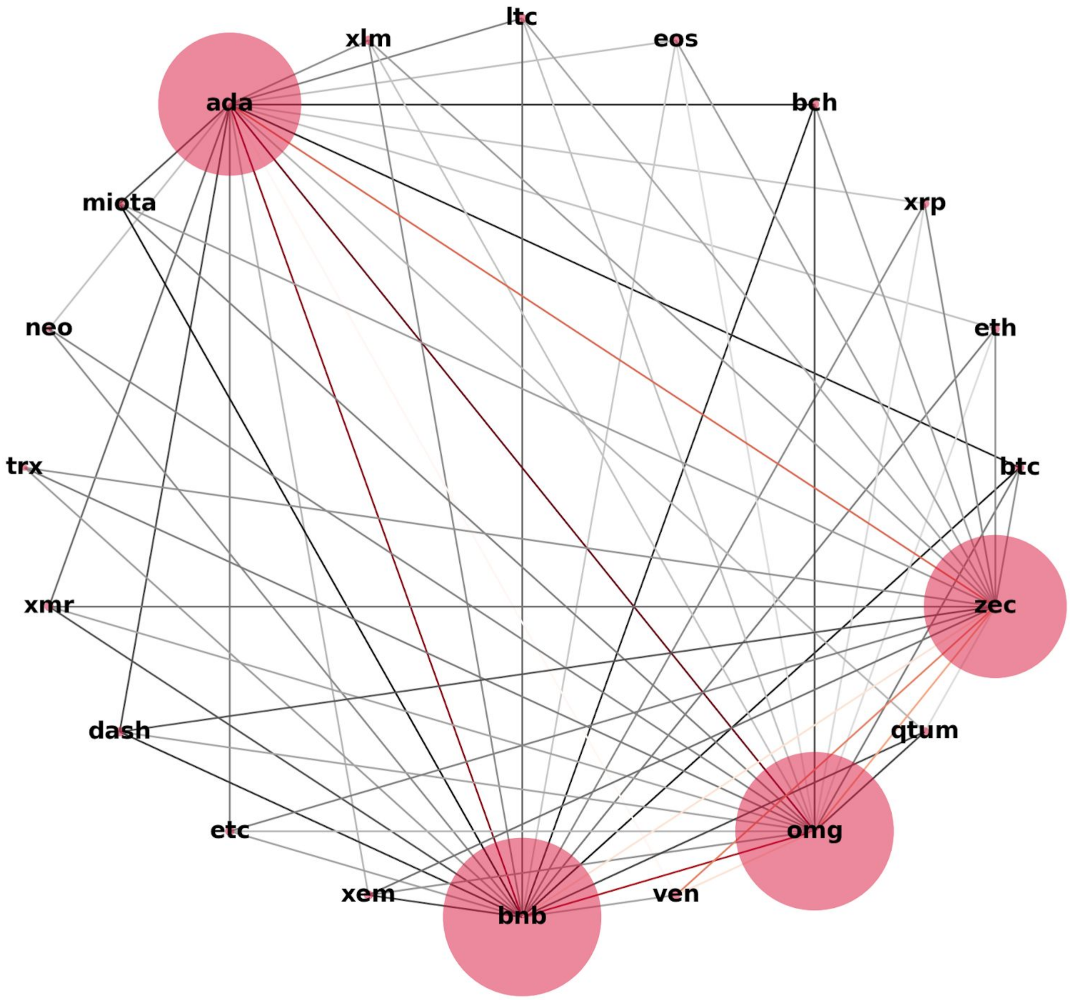

Figure 3.5.3.2 Graph of correlations between the top 20 cryptocurrencies

In this section, the scope of analysis is reduced to the top twenty coins which, at the time of this writing, are the most popular in regards to market capitalization (see Table 3.5.3.1). The latter are more likely to have development teams that have proven their engineering and innovation capabilities to the ecosystem, thus achieving higher price appreciation. Also, they have a higher potential to realize a product-market fit faster than their competitors due to higher availability of resources. A slight change is done in Table 3.5.3.1 by removing the 10th coin Tether, with a ticker symbol 'usdt', from the original ranking. The motive behind this is the fact that this cryptocurrency aims to serve as a one-to-one peg with the USD, therefore it is considered to be a 'stable' coin rather than an investment vehicle. Therefore, any potential existing correlations with this coin would be false as none should exist.

For practical understanding of the relationships, Figure 3.5.3.1 and Figure 3.5.3.2 provide a heatmap and a weighted graph, respectively. It can be observed that the majority of the calculated correlation coefficients are found to be not t-test significant, thus missing from the figures. However, a group consisting of the coins Cardano, Binance Coin, OmiseGO and Zcash. Those four differentiate themselves by having a strong positive relationship with each other while remaining uncorrelated to the rest. This information is of utmost importance in efficient portfolio generation. These results contribute to the rejection of *Hypothesis 3*.

Chapter 4: Event study

4.1 Methodology

Following the correlation analysis conducted in Chapter 3, an event study of the degree of impact is of interest to this research in order to provide further insights about any mutual connections among cryptocurrency projects reacting to new information. The applied methodology is the same as the one that has been used to analyse stock price changes due to firm announcements. The concept of event study is firstly described in (Fama et al., 1969), while (Malkiel and Fama, 1970) define the Efficient Market Hypothesis (EMH) as one that a market is efficient if assets' prices reflect all information, available to investors. The hypothesis in the event study in (Fama et al., 1969) is tested by measuring the price effect on stocks, once there is an announcement about planned stock splits in the future. Due to the positive likelihood of increased dividends after a stock split, the market reevaluates the firm stock price positively. Concerning the power of event study methodologies, (Brown and Warner, 1980) and (Brown and Warner, 1985) examine the impact of using monthly versus daily stock returns, respectively. Simulations in both research papers show that the application of methods that leverage the Ordinary Least Squares (OLS) market model and parametric statistical tests is straightforward and leads to correct results.

Furthermore, (MacKinlay, 1997) provides a general description of a methodology that can be applied in event studies concerning financial securities. Therefore, due to the lack of regulatory clarity about cryptocurrencies' asset status, they are assumed to be securities by the researcher. (Dodd and Warner, 1983) employ similar methodology in their analysis of events where proxy contests for seats of firms' board of directors are organised. In the center of the analysis, there is the concept of abnormal return (1), whose purpose is to illustrate the impact of a specific event on a security price in a quantitative way. It is calculated by subtracting the normal return, which does not account for the event in question, from the actual return. It can be observed that the more the actual return is different than the normal return, the more significant the event impact is to the security.

$$AR_{i,t} = R_{i,t} - E(R_{i,t} \mid X_t) \quad (1)$$

Structurally, the $AR_{i,t}$ signifies the abnormal return for a specific security *i* and event date *t*, the $R_{i,t}$ represents the actual return, and finally the $E(R_{i,t} \mid X_t)$ describes the expected return which is calculated based on historical data under the conditional information that the event had not happened. Basically, in order to calculate the latter, there are two options to model it - using the constant mean return model (CMRM) (2) or market model (3). The first one relies on the assumption of security returns to always converge to their historical mean return over time plus some disturbance term, whereas the second one incorporates the return rate of the market as a whole and uses it as a predictor of a security's price. (MacKinlay, 1997) argues that there is a potential gain to be made by preferring the market model over

CMRM. In essence, the higher the R^2 coefficient of the market model regression, the more the effect of the variance, caused by the market, to the abnormal return is reduced. Therefore, a potential contribution to a more accurate event effect on the price levels could be achieved.

$$R_{i,t} = \mu_i + \epsilon_{i,t} \quad (2)$$
$$E[R_{i,t} \mid X_t] = \alpha_i + \beta_i R_{mt} \quad (3)$$

In the context of the cryptocurrency event study, the market model is used with estimating an asset price by calculating a linear regression over the whole market returns. In order to avoid double counting, the market capitalization (4) of each analyzed cryptocurrency is subtracted from the total market capitalization before calculating the logarithmic market returns and using them in the regression.

$$R_{m,i,t} = \ln(\mathrm{ret}(TMC_t - MC_{i,t}) + 1) \quad (4)$$

Furthermore, the cumulative abnormal return across the event window is of interest as it will show the total impact that an event caused on the final return of a specific asset *i* (5).

$$CAR_i = \sum_{t=t_1}^{t_n} AR_{i,t} \quad (5)$$

The measurement of the event impact is expressed by a t-value, which is associated with its abnormal return and calculated only in the event interval. If the absolute value of a t-value is greater than the critical value 1.96 (MacKinlay, 1997), then it is 95% confident that the corresponding event had a significant effect on the price.

Given the results in Chapter 3, and their corresponding interpretation of Bitcoin being the major cryptocurrency to which most of the others are highly correlated with, the projects of choice for this event study are Bitcoin and Ethereum. Being the first and second coins in terms of market capitalization as of 7 July 2018, respectively, the latter is a potential substitute and direct competitor for the market leader spot. The time frame for their estimation window is determined to be 120 days, while the event window spans over 2 days before and after the actual event date. However, the latter is expanded to 7 days around the event date for macroeconomic news as these are not directly related to the cryptocurrency market, therefore might have a rather sluggish effect (if any).

A set of events with their corresponding dates are identified with the help of the Google search engine. The choice of them is made in a time-framed subjective manner according to the researcher's previous experience of being involved in the domain. Table 4.1 designates a categorisation of the events for this study. Common criteria that the sample satisfies is occurrence between 1 Jan 2016 and 1 Sep 2018 and mass media attraction. Three types of events have been identified and analysed:

Virtual attacks

The cryptocurrency ecosystem is known to be one with frequently occurring events about breached virtual exchange or wallet security, phishing scams and other malicious activities eventually leading to major fund losses. (Magas, 2018) has summarized a list of the most notable hacks, along with their corresponding theft of funds.

Crypto ecosystem

This section comprises events of different nature, however, all of them are playing a vital role in providing information to interested individuals about the potential future of the industry in a direct or indirect way.

Macroeconomic

To further expand the scope of this study, a few events of politico-economic nature are included in the sample under research. It is interesting to examine their effect on cryptocurrencies as that could open new trends and trading opportunities.

Virtual attacks		Crypto ecosystem		Macroeconomic	
2016-06-17	DAO attack[31]	2017-05-(22-24)	Consensus summit[32]	2016-06-24	Brexit referendum results[33]
2016-08-02	Bitfinex attack[34]	2017-08-01	Bitcoin Cash hardfork[35]	2016-11-08	Trump election victory[36]
2017-11-07	Parity multi-sig wallet[37]	2017-10-31	CME Group BTC futures[38]	2018-03-01	Trump tariffs[39]
2017-12-06	NiceHash attack[40]	2017-12-01	CBOE BTC futures[41]		
2018-01-26	Coincheck attack[42]	2018-01-16	BitConnect shutdown[43]		
2018-02-09	BitGrail	2018-03-26	Twitter ban[45]		

31 https://blockgeeks.com/guides/cryptocurrency-hacks/
32 https://www.coindesk.com/events/consensus-2017/
33 https://en.wikipedia.org/wiki/Brexit
34 https://blockgeeks.com/guides/cryptocurrency-hacks/
35 https://en.wikipedia.org/wiki/Bitcoin_Cash
36 https://en.wikipedia.org/wiki/United_States_presidential_election,_2016
37 https://medium.com/cybermiles/i-accidentally-killed-it-and-evaporated-300-million-6b975dc1f76b
38 https://www.cmegroup.com/media-room/press-releases/2017/10/31/cme_group_announceslaunchofbitcoinfutures.html
39 https://en.wikipedia.org/wiki/Trump_tariffs
40 https://blockgeeks.com/guides/cryptocurrency-hacks/
41 http://www.cboe.com/blogs/options-hub/2017/12/01/cboe-bitcoin-futures-announcement
42 https://www.coindesk.com/cryptocurrency-exchange-coincheck-abruptly-suspends-withdrawals/
43 https://thenextweb.com/hardfork/2018/01/16/bitconnect-shut-down-closed/

	attack[44]				

Table 4.1 Classification of events

4.1 Results

This section illustrates the outcome of the aggregated event study analysis. In the following subsections, individual occasions that proved to have had a significant impact in terms of price deviations of Bitcoin or Ethereum are provided, therefore rejecting *Hypothesis 4*. The listing of events answers RQ4.

4.1.1 Bitfinex attack

Bitfinex is a cryptocurrency trading platform, which according to Coinmarketcap is third in market share by volume at the time of this writing. During 2016, shortly after the exchange partnered with a third-party firm BitGo and deployed a multi-signature 'hot' wallet security system, 120,000 BTC were stolen, worth $72 million back then. The official announcement was made on 2 August 2016 ("5 High Profile Cryptocurrency Hacks", 2018). On the day after the incident, the price of Bitcoin dropped by ~10%, but that trend did not persist.

Date	BTC return	BTC AR	BTC T-value	BTC CAR
2016-07-31	-0.03	-0.02	-1.29	-0.02
2016-08-01	-0.02	-0.03	-1.36	-0.05
2016-08-02	-0.03	-0.03	-1.72	-0.08
2016-08-03	-0.09	-0.11	-5.78	-0.2
2016-08-04	0.04	0.05	2.63	-0.14

Table 4.1.1.1 Analysis of Bitfinex attack

4.1.2 Consensus summit

'Consensus', organized by CoinDesk, is an annual technology conference, where various stakeholders are experimenting with the application of blockchain-related systems and contributing to the development of the surrounding ecosystem. Interested parties include enterprises of all sizes, investors, institutions, policy groups and others. It is an event, which every investor keeps track of as they could gather information about the current and potential future state of current projects in the field. The occurrence of the summit is

45
https://www.cnbc.com/2018/03/26/twitter-bans-cryptocurrency-advertising-joining-other-tech-giants-in-crackdown.html
[44] https://bitgrail.com/news#february

seen as ‘bullish’ news from an investor’s perspective and can be observed by the significant contributions to the expected returns of both Bitcoin and Ethereum with +20% and +50%, respectively.

Date	BTC return	BTC AR	BTC T-value	BTC CAR	ETH return	ETH AR	ETH T-value	ETH CAR
2017-05-20	0.02	0.03	1.26	0.03	0.13	0.12	0.76	0.12
2017-05-21	0.05	0.05	2.29	0.07	0.04	0.03	0.2	0.15
2017-05-22	0.07	0.07	3.52	0.14	0.37	0.38	2.33	0.53
2017-05-23	0.02	0.02	0.86	0.16	-0.1	-0.1	-0.64	0.43
2017-05-24	0.05	0.06	3	0.22	0.13	0.11	0.68	0.54
2017-05-25	0.13	0.13	6.4	0.35	0	-0.01	-0.05	0.53
2017-05-26	-0.15	-0.15	-7.32	0.2	-0.04	-0.03	-0.21	0.5

Table 4.1.2.1 Analysis of Consensus summit

4.1.3 BitConnect shutdown

BitConnect[46] (BCC) was a highly controversial cryptocurrency throughout the whole ecosystem. It was claimed to be a Ponzi scheme by separate individuals due to BCC’s high return on investment (1% daily interest rate) and its multi-level marketing structure. In December 2017, the coin hit an all-time high of $463, while after its collapse, it hit $5.92 on 30 January 2018. An official announcement that confirmed the scam theory was the BCC statement of shutting down after the US states Texas and North Carolina issued an order for ceasing operations ("BitConnect is shutting down its lending and exchange platform", 2018). In this event window, both Bitcoin and Ethereum suffered a loss of -30% of their expected return.

Date	BTC return	BTC AR	BTC T-value	BTC CAR	ETH return	ETH AR	ETH T-value	ETH CAR
2018-01-14	0	0.02	0.32	0.02	-0.05	-0.05	-0.43	-0.05
2018-01-15	-0.04	-0.05	-0.85	-0.03	-0.01	-0.01	-0.13	-0.06
2018-01-16	-0.04	-0.04	-0.61	-0.07	-0.24	-0.24	-2.22	-0.3
2018-01-17	-0.18	-0.26	-4.23	-0.33	-0.18	-0.2	-1.84	-0.5
2018-01-18	0.03	0.04	0.64	-0.3	0.2	0.2	1.86	-0.3

[46]BitConnect - https://bitconnect.co/

Table 4.1.3.1 Analysis of BitConnect shutdown

4.1.4 Brexit referendum results

Following a referendum about whether the United Kingdom (UK) should stay in the European Union (EU) or not, conducted on 23 June 2016, the outcome was the latter. The majority of the voters supported leaving the EU, which is due on 29 March 2019. The exact politico-economic implications are still unclear leading to high uncertainty. With interest towards whether some parties decided to hedge their financial risk against potential emerging crisis, the event study shows that there was a small spike in demand for Bitcoin on the day of the results announcement (24 June 2016), as well as on the following one.

Date	BTC return	BTC AR	BTC T-value	BTC CAR
2016-06-17	0.01	0.04	1.82	0.04
2016-06-18	0.02	-0.01	-0.4	0.03
2016-06-19	-0	-0.01	-0.26	0.03
2016-06-20	0.01	0.02	0.78	0.04
2016-06-21	-0.07	-0.06	-2.66	-0.02
2016-06-22	-0.05	-0.07	-3.32	-0.09
2016-06-23	-0.14	-0.12	-5.59	-0.21
2016-06-24	0.1	0.09	3.95	-0.12
2016-06-25	0.04	0.04	2.04	-0.08
2016-06-26	-0.03	-0.02	-0.84	-0.1
2016-06-27	-0.02	-0.03	-1.44	-0.13
2016-06-28	0.01	0.01	0.56	-0.12
2016-06-29	-0.02	-0.02	-0.71	-0.13
2016-06-30	0	0	0.02	-0.13
2016-07-01	0.05	0.05	2.34	-0.08

Table 4.1.4.1 Analysis of Brexit referendum results

4.1.5 Trump election victory

In 2016, the 58th American presidential elections took place in which the representative of the Republican party Donald Trump surprisingly won over the Democrat Hillary Clinton. The date of the results announcement is 8 November 2016. In terms of fiscal policy, the Republican party tends to favor less government intervention in everyday business activities, thus supports deregulation and free market economy. By promoting supply-side economics, tax cuts for both businesses and workers aim to boost overall investment levels ("Do Republican Economic Policies Work?", 2018). The impact of this news was minuscule. Scoring a positive abnormal return the day following the event day is the only direct impact of that event that the author can infer.

Date	BTC return	BTC AR	BTC T-value	BTC CAR
2016-11-01	0.03	0.04	2	0.04
2016-11-02	-0	-0	-0.23	0.03
2016-11-03	0.02	0.02	0.94	0.05
2016-11-04	-0.04	-0.04	-2.02	0.01
2016-11-05	-0	-0.01	-0.32	0.01
2016-11-06	0.01	0.01	0.49	0.02
2016-11-07	-0.01	-0.02	-0.8	0
2016-11-08	0	0	0.22	0.01
2016-11-09	0.04	0.04	2.19	0.05
2016-11-10	-0.02	-0.03	-1.35	0.02
2016-11-11	-0	0	0.03	0.02
2016-11-12	-0.01	-0.01	-0.7	0.01
2016-11-13	-0.02	-0.02	-1.03	-0.01
2016-11-14	0.01	0.01	0.59	0
2016-11-15	0.01	0.01	0.7	0.01

Table 4.1.5.1 Analysis of Trump election victory

Chapter 6: Discussion

6.1 Relationships in the cryptocurrency market

By means of the widely adopted data science framework across industries CRISP-DM, quantitative research was performed in order to explore current relationships in the cryptocurrency market. Its sequence of logically structured steps played a significant role in regarding task breakdown and business-IT aspects alignment. The data was gathered from Coinmarketcap, which is a popular web platform with more than 90 million visits only during September 2018 (Coinmarketcap August Overview, 2018), therefore considered to be a trusted source for this research. After that, a correlation analysis methodology based on theoretical foundations about stationary time series was derived and implemented on a subset of all cryptocurrency projects in existence at this point. The results show approximately half of the cryptocurrencies have a relationship with the overall market regarding price action with 25% of them moving in the same direction. Another finding includes that Bitcoin, the first made cryptocurrency and the one with the highest market capitalization ever since, has ceased to be the most prominent as around 63% of the other cryptocurrencies are uncorrelated to it. Using the newly found information about the strength and direction of relationships between cryptocurrencies, diversified portfolios can be generated and considered by investors for their current and future allocation of assets. In combination with data about expected returns and quantified risks, which is out of scope in this paper, the set of portfolios can be used as an input to MPT-based tools and further graded in terms of efficiency. The ones with the highest potential reward and lowest risk are to be added to the efficient frontier - a set of all 'good' portfolios.

On the other hand, the approach applied in this paper does not take into account the dynamic nature of correlations between assets across time. In order to build a model which is not to be rendered obsolete with time, a rolling correlation should be used as it is a more rigorous metric that keeps track of changes in the relationships. It is calculated by calculating the correlation between two different assets on a rolling window basis. Thus, periodic portfolio rebalancing strategies are enabled based on solely distinct relationships. For example, two assets having no correlation could become highly interrelated over the next few months, thus increasing the portfolio's exposure to the same risk factors. In this case, it should be rebalanced by factoring in the new correlation coefficients between assets and redistributing the total value accordingly. This way, an initially efficient portfolio could stay as such by re-adjusting itself to accommodate to changed return expectations and environmental risks.

6.2 Influence of events in the cryptocurrency market

Being a relatively new industry in existence, around 10 years old, it is of interest to measure the degree of impact that various events could have on projects' returns. The employment of an event study methodology provided this research with the statistical toolbox, needed to identify patterns of discrepancies in coins' expected returns in a quantitative manner. To discover a specific type of events that are consistently causing abnormalities, the dataset has been divided into three categories: virtual attacks, ecosystem and macroeconomic. Five out of fifteen events, approximately uniformly distributed by category, were identified to have had a significant contribution to the forecasted returns.

However, a limitation of this methodology is the absolute necessity to isolate the event window in such a way that there are no other accompanying events happening simultaneously with the event under investigation. This requirement could only be dealt with in a subjective manner due to the lack of a publicly available exhaustive list of historic cases that could potentially cause significant influence and distort the outcomes of the research. Thus, one should be cautious with absolute conclusions about the actual quantitative effects on price returns.

Bibliography

Hileman, G., & Rauchs, M. (2017). Global cryptocurrency benchmarking study. Cambridge Centre for Alternative Finance.
Gandal, N., & Halaburda, H. (2016). Can we predict the winner in a market with network effects? Competition in cryptocurrency market. Games, 7(3), 16.
Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing.
Van Valkenburgh, P. (2016). Open Matters—Why Permissionless Blockchains are Essential to the Future of the Internet. Coin Center, December, 14, 2016-12.
Nakamoto, S. (2008). A Peer-to-Peer Electronic Cash System. Bitcoin.org. Retrieved from https://bitcoin.org/bitcoin.pdf
King, S., & Nadal, S. (2012). Ppcoin: Peer-to-peer crypto-currency with proof-of-stake. self-published paper, August, 19.
Bentov, I., Gabizon, A., & Mizrahi, A. (2016, February). Cryptocurrencies without proof of work. In International Conference on Financial Cryptography and Data Security (pp. 142-157). Springer, Berlin, Heidelberg.
Vukolić, M. (2017, April). Rethinking permissioned blockchains. In Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts (pp. 3-7). ACM.
Szabo, N. (1994). Smart contracts. Unpublished manuscript.
Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. (2016, May). Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In 2016 IEEE symposium on security and privacy (SP) (pp. 839-858). IEEE.
Buterin, V. (2014). A next-generation smart contract and decentralized application platform. white paper.
GitHub. (2018). EOSIO/Documentation. [online] Available at: https://github.com/EOSIO/Documentation/blob/master/TechnicalWhitePaper.md [Accessed 12 Jul. 2018].
Osterrieder, J., Lorenz, J., & Strika, M. (2016). Bitcoin and cryptocurrencies-not for the faint-hearted.
Vasek, M., & Moore, T. (2015, January). There's no free lunch, even using Bitcoin: Tracking the popularity and profits of virtual currency scams. In International conference on financial cryptography and data security (pp. 44-61). Springer, Berlin, Heidelberg.
Bloomberg. (2018, January 31). How Hackers Stole $500 Million in Digital Currency. Retrieved September 11, 2018, from http://fortune.com/2018/01/31/coincheck-hack-how/
Pollock, D. (2018, March 09). The Mess That Was Mt. Gox: Four Years On. Retrieved September 11, 2018, from https://cointelegraph.com/news/the-mess-that-was-mt-gox-four-years-on
Garcia, D., Tessone, C. J., Mavrodiev, P., & Perony, N. (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. Journal of the Royal Society Interface, 11(99), 20140623.
Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Scientific reports, 3, 3415.

Rollins, J. (2015). Why we need a methodology for data science. Retrieved from http://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science
Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM.
Rohanizadeh, S. S., & BAMENI, M. M. (2009). A proposed data mining methodology and its application to industrial procedures.
Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137-166.
Kdnuggets.com. (2018). What main methodology are you using for your analytics, data mining, or data science projects? Poll. [online] Available at: https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html [Accessed 17 Jul. 2018].
CRISP-DM, S. P. S. S. (2000). Step-by-step Data Mining Guide.
Markowitz, H. (1952). Portfolio selection. The journal of finance, 7(1), 77-91.
Mangram, M. E. (2013). A simplified perspective of the Markowitz portfolio theory.
M. (n.d.). Lowering Portfolio Risk through Diversification. Retrieved August 28, 2018, from https://www.money-zine.com/investing/investing/lowering-portfolio-risk-through-diversification/
Jaffe, J., & Randolph Westerfield, R. (2004). Corporate finance. Tata McGraw-Hill Education.
Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series?--a study in sampling and the nature of time-series. Journal of the royal statistical society, 89(1), 1-63.
Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23.
Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica: Journal of the Econometric Society, 1057-1072.
Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.
Schwert, G. W. (2002). Tests for unit roots: A Monte Carlo investigation. Journal of Business & Economic Statistics, 20(1), 5-17.
Fedorová, D. (2016). Selection of unit root test on the basis of length of the time series and value of ar (1) parameter. STATISTIKA, 96(3), 3.
Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. Journal of econometrics, 2(2), 111-120.
Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1), 59-66.
Statsmodels.stats.weightstats.ttest_ind¶. (n.d.). Retrieved October 22, 2018, from https://www.statsmodels.org/dev/generated/statsmodels.stats.weightstats.ttest_ind.html
Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. International economic review, 10(1), 1-21.
Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), 383-417.
MacKinlay, A. C. (1997). Event studies in economics and finance. Journal of economic literature, 35(1), 13-39.
Dodd, P., & Warner, J. B. (1983). On corporate governance: A study of proxy contests. Journal of financial Economics, 11(1-4), 401-438.

Brown, S. J., & Warner, J. B. (1980). Measuring security price performance. Journal of financial economics, 8(3), 205-258.
Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. Journal of financial economics, 14(1), 3-31.
Magas, J. (2018, August 31). Crypto Exchange Hacks in Review: Proactive Steps and Expert Advice. Retrieved September 24, 2018, from https://cointelegraph.com/news/crypto-exchange-hacks-in-review-proactive-steps-and-expert-advice
B., @brenfife, B. F., & Dulger, I. (2018, April 01). 5 High Profile Cryptocurrency Hacks - (Updated). Retrieved September 26, 2018, from https://blockgeeks.com/guides/cryptocurrency-hacks/
M. (2018, January 16). BitConnect is shutting down its lending and exchange platform. Retrieved September 26, 2018, from https://thenextweb.com/hardfork/2018/01/16/bitconnect-shut-down-closed/
Amadeo, K. (2018, March 1). Do Republican Economic Policies Work? Retrieved September 26, 2018, from https://www.thebalance.com/do-republican-economic-policies-work-4129139
Coinmarketcap August 2018 Overview. (2018, October 1). Retrieved October 1, 2018, from https://www.similarweb.com/website/coinmarketcap.com

Appendices

Appendix A: Cryptocurrency ticker symbols to names mappings

Symbol	Name
btc	Bitcoin
eth	Ethereum
xrp	XRP
bch	Bitcoin Cash
eos	EOS
ltc	Litecoin
xlm	Stellar
ada	Cardano
miota	IOTA
usdt	Tether
neo	NEO
trx	TRON
xmr	Monero

Symbol	Name
rdd	ReddCoin
dcn	Dentacoin
knc	Kyber Network
cnx	Cryptonex
veri	Veritaseum
emc	Emercoin
cmt	CyberMiles
pivx	PIVX
ctxc	Cortex
ela	Elastos
poly	Polymath
drop	Dropil
eng	Enigma

dash	Dash
etc	Ethereum Classic
xem	NEM
bnb	Binance Coin
ven	VeChain
xtz	Tezos
omg	OmiseGO
qtum	Qtum
zec	Zcash
ont	Ontology
icx	ICON
zil	Zilliqa
lsk	Lisk
bcn	Bytecoin
zrx	0x
dcr	Decred
ae	Aeternity
btg	Bitcoin Gold
bts	BitShares
steem	Steem
rep	Augur
btm	Bytom
sc	Siacoin
xvg	Verge
nano	Nano
mkr	Maker
dgb	DigiByte
gnt	Golem
bcd	Bitcoin Diamond
doge	Dogecoin
ppt	Populous
waves	Waves
snt	Status
rhoc	RChain
strat	Stratis

mana	Decentraland
kin	Kin
moac	MOAC
powr	Power Ledger
theta	Theta Token
etn	Electroneum
qash	QASH
nxt	Nxt
bix	Bibox Token
dent	Dent
wicc	WaykiChain
wax	WAX
smart	SmartCash
sys	Syscoin
payx	Paypexca
gto	Gifto
storm	Storm
tusd	TrueUSD
sub	Substratum
fct	Factom
nuls	Nuls
cennz	Centrality
ddd	Scry.info
xzc	ZCoin
man	Matrix AI Network
gtc	Game.com
zen	ZenCash
nxs	Nexus
link	ChainLink
fsn	Fusion
mtc	Docademic
gbyte	Byteball Bytes
salt	SALT
block	Blocknet
storj	Storj

wan	Wanchain
hsr	Hshare
bat	Basic Attention Token
kcs	KuCoin Shares
wtc	Waltonchain
btcp	Bitcoin Private
xin	Mixin
iost	IOST
nas	Nebulas
dgd	DigixDAO
lrc	Loopring
ht	Huobi Token
aion	Aion
mith	Mithril
kmd	Komodo
gxs	GXChain
elf	aelf
ardr	Ardor
ark	Ark
bnt	Bancor
maid	MaidSafeCoin
loom	Loom Network
fun	FunFair
mona	MonaCoin
ethos	Ethos
mco	Monaco
gas	Gas

drgn	Dragonchain
cvc	Civic
icn	Iconomi
pay	TenX
tpay	TokenPay
agi	SingularityNET
btcd	BitcoinDark
dbc	DeepBrain Chain
act	Achain
blz	Bluzelle
nexo	Nexo
hpb	High Performance Blockchain
hot	Holo
mds	MediShares
rlc	iExec RLC
r	Revain
ode	ODEM
tnb	Time New Bank
bft	BnkToTheFuture
ncash	Nucleus Vision
sky	Skycoin
gno	Gnosis
ant	Aragon
cvt	CyberVein
grs	Groestlcoin
req	Request Network
nebl	Neblio

Appendix B: Working files

Public access to all thesis-related working files: https://github.com/iiosenov/cryptocurrency-thesis