TECHNISCHE UNIVERSITÄT WIEN

D I P L O M A R B E I T

Blockchain Technology and some Statistical Properties of Crypto-Currency Prices

ausgeführt am

Institut für Finanz und Versicherungsmathematik der Technischen Universität Wien

unter der Anleitung von

Univ.Prof. Dipl.-Math. Dr.rer.nat. Thorsten Rheinländer

und

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Friedrich Hubalek

durch

Dipl.-Ing. Lukas Fertl

Matrikelnummer: 01151896

Wien, am 20.Oktober.2019

(Betreuer 1)	(Betreuer 2)	(Verfasser)

Kurzfassung

Im Jahr 2009 stellte Satoshi Nakamoto die Blockchain Technology und Bitcoin der Öffentlichkeit vor. Seitdem ist das Interesse an Krypto-Währungen, ins besonders Bitcoin, und der innovative Blockchain Technology stark gewachsen. Der Preis von Bitcoin stieg spektakulär und erreichte an seinem Höchststand fasst 20 000 USD. In dieser Arbeit wurde die Funktionsweise der Blockchain, welche ein öffentliches Transaktionsregister ist, untersucht und erklärt. Heutzutage bedarf es für eine herkömmliche Transaktion einen Mittelsmann, meist eine Bank, der sicherstellt das die Überweisung ordnungsgemäß abgewickelt wird. Die Blockchain Technology eröffnet hingegen die Möglichkeit Transaktionen sicher ohne einen Mittelsmann abzuwickeln, und das Vertrauen in den Mittelsmann wird durch Kryptographie ersetzt.

Weiters wurde in dieser Arbeit die Frage untersucht ob sich Preise von Krypto-Währungen aus einer statistischen Sichtweise von Aktienpreisen unterscheiden. Es wurde schon lang beobachtet, dass Preise von Aktien sich an sogenannte stilisierte Fakten halten. Das heißt, dass gewisse statistische Eigenschaften universell bei fast allen Aktienpreisen auftreten, unabhängig davon welche Aktie oder Markt konkret betrachtet wird. Werden zum Beispiel die Returns (logarithmische Differenzen der Aktienpreise) betrachtet zeigt sich meist, dass diese nicht normalverteilt sind und die Verteilung schwerer Tails besitzt als die Normalverteilung. Weiters wird ein Volatilitäts-Clustering-Effekt beobachtet, der besagt, dass es nach Tagen mit großen Kursschwankungen wahrscheinlicher ist, das am nächsten Tag wieder eine große Schwankungsbreite vorherrscht.

In dieser Arbeit wurden daher die statistischen Werkzeuge zum untersuchen dieser Frage vorgestellt und auf die Frage angewendet ob bei Preisen von Krypto-Währungen ähnliche Effekte auftreten. Diese Fragestellung konnte positive beantwortet werden und Krypto-Währungspreise verhalten sich aus statistischer Sicht nicht anders als Aktienpreise. Weiters wurde festgestellt, dass Krypto-Währungen viel riskanter (schwankungsanfälliger) als Aktien sind aber, dass es trotzdem vorteilhaft sein kann in Sie zu investieren, da Sie kaum mit herkömmlichen Märkten korreliert sind und daher zur Diversifizierung benutzt werden können.

Abstract

The blockchain technology and bitcoin were introduced in 2009 by Satoshi Nakamoto. Since then there has been a great interest in the innovative technology and the price of bitcoin rose spectacular and hit nearly 20 000 USD. In this thesis, the blockchain technology is described in detail. For conventional transactions, a trusted third party (i.e. a bank) is required to secure the transaction. But the blockchain technology offers a new way to process transactions where the trusted third party is not required anymore and is replaced by cryptography.

Furthermore, the question if prices of crypto-currencies behave differently than stock prices from a statistical point of view is investigated. It is common knowledge that prices of stocks possess so called stylized facts and therefore the question arises if crypto-currencies conform to them too. The returns (logarithmic price differences) of stocks are not normally distributed and possess heavier tails. Furthermore, there is the so called volatility clustering effect, i.e. after days with high volatility it is much more likely that the next day is volatile. Therefore the statistical tools to analyze these questions were explained and then used to answer the question if crypto-currency prices display the same stylized facts as stock prices. The answer is positive, i.e. crypto-currency prices display the same statistical properties as stocks. Nevertheless, they are much more volatile than stocks but can offer benefits, from an investor's point of view, since they are uncorrelated to stock markets and therefore can be used for diversification.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei meiner Familie, insbesonders meiner Mutter Dr. Elisabeth Fertl, bedanken, da Sie mich immer ermutigt und unterstützt haben. Weiters bedanke ich mich bei meinen zwei Betreuen Thorsten Rheinländer und Friedrich Hubalek für die Unterstützung. Außerdem gilt mein Dank auch der Technischen Universität Wien.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 20.Oktober.2019

Dipl.-Ing. Lukas Fertl

Contents

1 Introduction

In the last years the blockchain technology and bitcoin, a peer to peer crypto-currency and electronic payment system, became quite famous. Its price against the USD, commonly referred to as bitcoin price, rose spectacular, although this rise was accompanied by extreme volatility and on 2017-03-02 it surpassed the price of gold (against the USD) and on 2017-12-16 reached a price of 19356 USD. It was invented and introduced in March of 2009 through a white paper [48] of Satoshi Nakamoto, whose identity is still unknown up to now. In a normal and modern currency regime (USD, EUR, ...) a central authority (i.e. central bank) controls the money supply and transactions are conducted through trusted third parties (banks). In contrast, bitcoin is a decentralized and open-source payment system which uses the blockchain technology and cryptography to secure transactions and control the money supply. The total supply of bitcoin is capped at 21 million and new bitcoins are produced (mined) and distributed according to a predefined schedule. This is one of the advantages proponents often cite, because the risk of diluting the monetary base by a central bank, which could happen via political pressure, is contained. Inflation on existing banknotes is therefore impossible. Thus, bitcoin is often compared or mentioned together with gold as a hedge against inflation.

Figure 1: historic development of bitcoin price versus Gold price on log scale.

In this master thesis, the basic structure and functionality of bitcoin and the blockchain technology are presented in section 2. In finance it is well known that stock market price data possess certain statistical properties, called stylized facts (see section 3, [39], [45], and [47] for more details). In section 3 the statistical tools to analyze and test for the stylized facts are presented in detail. Finally, in section 4 these tools are applied to analyze if crypto-currencies behave similar to stock markets, from a statistical point of view. The question is if crypto-assets conform to the same stylized facts as stocks. In section 5 a summary of the thesis is provided.

2 How does bitcoin work

Bitcoin is a peer to peer open-source payment system and one unit is called bitcoin, which is basically a private key that can be used to proof the ownership of the output of a previous transaction that is saved in the distributed ledger (i.e. the transaction history). It relies on the blockchain technology, a publicly available database of all transactions so far, consisting of blocks that contain the transactions and a reference to the previous block (i.e. the hash value of the previous block). This chain is openly available and redundantly saved on many computers (i.e nodes) participating in the peer to peer network which is governed by the bitcoin protocol. So the chain grows when a new block is published, which also creates new bitcoins. This is an incentive for participating and creating new blocks which require a proof-of-work (i.e. inverting a cryptographic hash function, for bitcoin case the *SHA256* hash is used). On average all 10 minutes, a new block is created and when a transaction is published and included in a block it is accepted by the network. The process of creating a new block is stochastic, since many PCs compete against each other to find a solution to a cryptographic problem, basically guessing the right input for a given output of the hash function, and two different miners may guess a solution at the same time creating a splitting in the chain. The network assumes the longest chain in terms of CPU power (total difficulty see mining) to be the true one and after a split, the chain with less CPU power behind it dies off. This makes cheating really hard or practically impossible, since to change a transaction after it was accepted by the network you would have to recreate the whole chain that came after the transactions and even surpass the growing honest one. Hence, the probability that a transaction can be changed drops of with the time that passes, Satoshi Nakamoto showed in [48] that it decays exponentially. For transactions, a Bitcoin-wallet or Bitcoin-Client (i.e an app or software on your PC) is needed for transactions and managing your account [2].

To understand the technology better we take a closer look at the structure of the blockchain, transaction, and mining. First, I will briefly summarize hash functions and Merkel trees.

2.1 Hash functions and Merkle Tree root

2.1.1 Hash function

A function $H(x)$ that has an input x of arbitrary length and output with fixed length is called Hash function. The output is called the hash value (hash). A cryptographic hash function is a hash function that is hard to invert but easy to evaluate (given input x it is computationally easy to calculate $H(x)$, but given an output $y = H(x)$ it is very hard to find input x) and where the output changes completely if the input is changed just a little bit.

Bitcoin uses the Secure Hash Algorithm 256 (i.e. *SHA256*), that means the output has a fixed length of 256 bit, for the proof-of-work in the creation of new blocks and in the creation of bitcoin addresses (similar to the account number or IBAN, but in bitcoin you create a new one for each transaction). To increase the security of the hash function bitcoin uses $SHA256^2$, which means:

$$H(x) = SHA256(SHA256(x))$$

Where $SHA256(\cdot)$ is the standard *SHA256* hash algorithm, this algorithm partitions the input into 512 bit blocks (respectively 16 blocks of 32 bits each)and iteratively calculates them together with 4 logic operations and 64 constants. The constants are derived from the square roots of the first 64 prime numbers [13] [12]

A full inversion (i.e. given an output $y = H(x)$ find the shortest possible input x with output y) through a brute force attack has a known running time of $O(2^{256})$ which makes it computationally infeasible [3]. For the output the hexadecimal system is used meaning that the output of the hash function is a number with 256 bit displayed in hexadecimal. Lets see a small example [13]:

$$SHA256(\text{"This is the master thesis in finance of Lukas"}) =$$
$$6b3652a60b126db4e7dccfd6e99286056c540b1caabd567fa9d4d511df46a248$$

where the output is a 256 bit number in hexadecimal. The hexadecimal system needs 16 symbols with values from $\{0, 1, ..., 15\} = \{0, 1, ..., 8, 9, a, b, c, d, e, f\}$ so for example a 9 bit number:

$$271_{10} = 2 \cdot 10^2 + 7 \cdot 10^1 + 1 \cdot 10^0 = 1 \cdot 16^2 + 0 \cdot 16^1 + 15 \cdot 16^0 = 10f_{hex} = \\ = 1 \cdot 2^8 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2 + 1 \cdot 2^0 = 100001111_2$$

Remark. *For bitcoin a partial inversion of SHA256*^{2} *is required since the miner has to find a* 32 *bit nonce such that the output is below a certain threshold, see section 2.4 for more details. A brute force partial inversion by trying out all possible* 32 *bit nonce has run time of* $O(2^{32})$*.*

2.1.2 Merkle Tree

A Merkle Tree is a binary tree of hashes. The aim is to bundle the hash values of multiple transactions into one hash (i.e. Merkle root), for bitcoin all transactions in a given time-span are bundled in one block. The concept is pretty straight forward and best demonstrated by an example: Suppose we have three transactions a, b and c to bundle [4]:

$$d_1 = H(a) \qquad d_2 = H(b) \qquad d_3 = H(c) \qquad d_4 = H(c)$$
$$d_5 = H(d_1 \oplus d_2) \qquad d_6 = H(d_3 \oplus d_4)$$
$$\text{root} = H(d_5 \oplus d_6)$$

where $\oplus$ stands for the concatenation operation. In the first line, the transaction c is used twice so that the starting row has an even number.

2.2 Blockchain

The blockchain is a publicly available database of all recorded transactions saved on all nodes participating in the network. When a new block is created (see section 2.4), it is broadcasted to the network (i.e. it is sent simultaneously to all nodes participating in the network. A node is a server of a miner where the blockchain is saved). The database is a sequence of blocks.

Definition 1. ***Block***
A block consists of the header and the body. The body is essentially just a list of all transactions (see section 2.3) that are included in the block. Originally the body was restricted to 1Mb which limited the number of transactions that could be included in one block. The header connects the previous to the current block and consists of the hash of the previous block header, the hash of the transactions included in the body, a time stamp, and the nonce (see section 2.4 and table 1 for details).

Definition 2. ***Blockchain***
The blockchain is simply a distributed ledger where the blocks are stored in chronological order since each block contains the hash of the previous block. Therefore, the chain contains the transactions (see section 2.3 for details) and grows over time. This guarantees a chronological order and makes changing transactions, once included into blocks, hard, since all the blocks that came after the block you want to manipulate would have to be recreated.

Blocks are created through the mining process which requires a proof-of-work, so reproducing a chain is harder the longer it is. The length of a chain is measured in total difficulty (see section 2.4 for details), where difficulty is a measure for how hard it is to solve the cryptographic problem that created the block. This measure changes over time since for a fixed difficulty [see also mining][5] the average time it takes to produce a block with this difficulty level depends on how much CPU power is trying to solve the problem (i.e. how many miners are out there). The bitcoin protocol aims to create a new block every 10 minutes and adjusts the difficulty according to a predefined formula every two weeks (e.g. making it harder when the blocks are created too quickly or vice versa) to meet this goal on average.

It is also possible for temporary splits or forks to occur, for example if two miners come up with the solution of a block at the same time and publish them to different nodes. Whichever of the two blocks gets integrated into the next block becomes part of the main chain and survives. The shorter chain is used for nothing since all valid transactions in it get integrated into the longer chain, as new transactions that need to be processed, when it dies off, but the rewards (genesis transaction, see section 2.4 for details) for generating the blocks of the shorter chain do not exist in the longer chain [6].

Since the whole transaction data is public it is possible to know the whole history of bitcoin transactions and in fact, for conducting a transaction you just reference to a previous transaction where you were a receiver and sign it with your private key. So the crucial point for privacy is the link of the bitcoin address (i.e. a character string that is used for

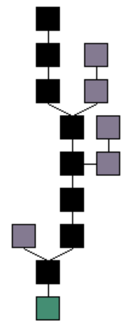

Figure 2: green: genesis block; black: longest chain; purple: dead blocks

identifying the sender and receiver of transactions, see section 2.3 for more details) and bitcoin client (e.g. wallet), if one is used, with the true identity of the bitcoin user. So if you download software (e.g. wallet), that helps you to keep track of your keys and addresses, with your true name and identity there exists the possibility that with this information your whole transactions could be traced down to you. Nowadays there are hundreds of different wallets to choose from. There are blockchain browser which displays the transaction history of the chain in human-friendly ways with hex editors [7] [8]

2.3 Transactions

A transaction is a transfer of bitcoin value where the input is a reference to a previous transaction output [8]. Transactions are published to the network and then incorporated into the next block. So it takes some time until a transaction is cleared, since it takes a random amount of time to create a block (in fact it is Poisson distributed [9][mining section]). This is one of the disadvantages of bitcoin which must be solved to become a feasible payment system for everyday life. There is the possibility to include transaction fees to give a transaction a higher priority so that it will get integrated into the next block that is created if there is an overload of pending transactions that have to be integrated into blocks.
Every time a new block is solved (i.e. the nonce, that solves the cryptographic problem, is found. See section 2.4 for details) also a genesis transaction occurs which means that bitcoins are created out of nowhere and sent to the address of the miner. This also makes

the problem to solve uniquely for each miner since each miner includes a different generation transaction, namely his address, as the first transaction into the transaction list of the block and the Merkle root of the transactions which is included in the problem is unique for every miner [see mining section][10]. All bitcoins in existence were created in this way. The smallest increment of one Bitcoin is one Satoshi in reference to the creator of bitcoin. 1 bitcoin = 100 000 000 Satoshi.

Here a little example of how a transaction looks like:

```
Input:
Previous tx: f5d8ee39a430901c91a5917b9f2dc19d6d1a0e9cea205b009ca73dd04470b9a6
Index: 0
scriptSig: 304502206e21798a42fae0e854281abd38bacd1aeed3ee3738d9e1446618c4571d10
90db022100e2ac980643b0b82c0e88ffdfec6b64e3e6ba35e7ba5fdd7d5d6cc8d25c6b241501

Output:
Value: 5000000000
scriptPubKey: OP_DUP OP_HASH160 404371705fa9bd789a2fcd52d2c580b65d35549d
OP_EQUALVERIFY OP_CHECKSIG
```

Figure 3: Data of a transaction [8]

The input takes 50 bitcoin (=5000000000 Satoshi) from output 0 (one transaction can send bitcoin from multiple addresses to multiple receiver addresses, therefore, they are indexed) of transaction f5d8ee..., signs them with the private key and sends them to the bitcoin address 40437... (all values except the index and value of the transaction are in the hexadecimal system) [8]. Since one transaction can source funds from multiple previous ones and send it to multiple addresses output 0 means that the first receiver address of transaction f58ee... sends the 50 bitcoin.

2.4 Mining

The process of solving a new block is called mining because brand new bitcoins are created when a block is finished and transferred to the miner. To do this the miners have to solve a difficult mathematical task (i.e. inverting the hash function). People who are part of the bitcoin network and provide computer resource to solving blocks are called miners. First, we will take a look at the structure of the block header which contains 6 fields

version:	block version number
hashPrevBlock:	256-bit hash of the previous block header
hashMerkleRoot:	256-bit Merkle root hash of the transactions in the block
Time:	Current timestamp as seconds since 1970-01-01T00:00 UTC
Bits:	Current target in compact format
Nonce:	32-bit number

Table 1: block header

The Bits and Nonce entries of the block header are related to solving the block. The nonce is the solution to the mathematical problem and the difficulty of the problem is controlled through the target in the Bits field. The version field is of minor importance and says which version is used.

The body of the block consists then of a list of all transactions (see section 2.3 and figure 3) that are compiled into the block by the miner.
Out of the header, the block hash is generated, namely concatenating all six fields and hashing them. The problem for solving the block is to find the nonce so that the hash of the block contains a certain number of leading zeros. Difficulty determines the number of leading zeros required to solve the block and is adjusted all 2016 blocks [5] (on average two weeks) so that on average all 10 minutes a block is created. The problem is unique to each miner, since the generation transaction that is different for each miner enters the header through the Merkle root of transactions.
So to use mathematical notation, given the current target T_t (i.e. a number set by the bitcoin protocol) the task is to find a nonce n given the information x of the first 5 rows such that:

$$H(x \oplus n) \leq T_t$$

which means a certain number of leading zeroes determined by the target (a 256-bit number in the hexadecimal system) [11] ($\oplus$ is again the concatenation operator). So the higher the target the lower the difficulty and vice versa. The $\leq$ instead of hitting the target exactly is to control difficulty and therefore the time it takes on average to create a block. The difficulty will depend on the hash rate of all miners combined (i.e. how often can they evaluate the hash function in a given time-span). For example, if all miners together can perform 10^6 evaluations in one second, they have a greater chance of finding the right nonce than if they have a hash rate of 10^5.

To demonstrate this we will look at a small example [10]:

$$H(\text{"Hello, world!0"}) =$$
$$1312af178c253f84028d480a6adc1e25e81caa44c749ec81976192e2ec934c64$$

where $Hello, world!$ would correspond to x and 0 to the nonce n in the above structure. The work is to try out different values for n such that leading zeros are produced in the hash value (the higher the hash rate the faster the problem is solved).

$$H(\text{"Hello, world!1"}) =$$
$$e9afc424b79e4f6ab42d99c81156d3a17228d6e1eef4139be78e948a9332a7d8$$

and after some trial and error a solution $n = 4250$ is found for a given hypothetical target of for example $000F000...$

$$H(\text{"Hello, world!4250"}) =$$
$$0000c3af42fc31103f1fdc0151fa747ff87349a4714df7cc52ea464e12dcd4e9$$

We will look at the concept of difficulty in more detail:

Difficulty:

To calculate it you just need the highest possible target $T_{\max}$ (lowest difficulty) and the current target T_t [5].

$$\text{difficulty} = \frac{T_{\max}}{T_t}$$

If the $T_t = T_{\max}$ then the difficulty is one. You can also think about it in a probabilistic way, since for a given target $T_t \in [0, T_{\max}] =: I$ the probability when drawing a number randomly out of I to be lower than T_t is $\frac{T_t}{T_{\max}}$ [compare uniform distribution] and the difficulty is the inverse of this probability, the lower the probability of being under the target the higher the difficulty.

The highest possible target used in bitcoin is:

$$T_{\max} = 00000000fff0000000000000000000000000000000000000000000000000000$$

Lets look at an example where we calculate the difficulty for a target given by:

$$T_t = 00000000000404cb000000000000000000000000000000000000000000000000$$
$$\text{difficulty} = \frac{T_{\text{max}}}{T_t} = 16307.42$$

The difficulty and therefore the target is adjusted all 2016 blocks in response to the observed hash rate (how many hash values can be computed or tried out per second) of the bitcoin network (all miners taken together). The aim is to create a block all 10 minutes and after 2016 blocks the network compares the actual time it took and adjusts the current target

$$T_t^{\text{new}} = T_t^{\text{old}} \cdot \frac{t_{\text{obs}}}{t_2}$$

with the percentage deviation of time where t_2 is the time-span of two weeks and t_{o} is the observed time it took to create 2016 blocks. For example, if the used frequency is minutes, $t_2 = 60 \cdot 24 \cdot 14 = 20160$ and lets suppose it took the miners $t_{\text{obs}} = 30000$ minutes to create 2016 blocks, then the target is adjusted by multiplying with $30000/2016 = 1.49$ and difficulty difficulty $= \frac{T_{\text{max}}}{T_t^{\text{new}}}$ will decrease by the factor $1/1.49 = 0.67$ (i.e. the difficulty is lowered approx 33%).

A block is solved when a miner discovers a nonce such that the hash of the block is below the target, then he publishes to the network and each node receiving the block checks if the proof-of-work was done (i.e. hashing the block header and looking if the block hash is below target). Then the new block and with it the transactions are integrated into the public available chain and the outputs of the transaction can be referenced as inputs for new ones.
Any change in the block especially in the transactions will change the hash of the block completely and to make a valid block out of it again requires finding a new nonce such that the block hash is below target. And, since the blocks form a chain by referencing each block to its previous, a potential criminal would have to recreate all blocks that came after the changed block and even surpass the growing honest chain which prevents double-spending and gives the bitcoin network its security. An attacker would need the ability to create blocks faster than the whole bitcoin network combined which means he must have a higher hash rate than the network. So the more time passes after a transaction is accepted and build into a block the more secure the transaction is. A transaction is said

to be sufficiently secure if there are more than 6 new blocks after the block where it is in [9].

Remark. *On 2017-07-01 the blockchain was split by a hard fork that was not revised later. This split was by purpose and created two crypto-currencies (bitcoin and bitcoin cash) out of the original bitcoin. We won't dive into too many technical details here, but a short summary will be presented:*

Due to the limited block size (i.e. 1 Mb which limited the capacity of bitcoin to approximately 7 transactions per second), there was a growing need to enhance the block size. Since there is no central authority that could have changed the source code, all miners (i.e. CPU power devoted to the system which produce the blocks and process the transactions) have to agree to use a different source code and still accept the history of transactions (i.e. chain) up until now for this change to happen. Not all miners agreed to the change and a fork occurred on 2017-07-01 since only 97% agreed to the change to increase the block size limit to 8Mb. The remaining miners did not change the source code and created bitcoin cash. Basically, each user doubled the amount of currency he/she had since with their key they had access to the same amount of bitcoin cash as they had bitcoin before without using access to their bitcoins. The transaction history of both currencies is the same until 2017-07-01. Think of creating a new currency and every holder gets one unit of new currency for one unit of old currency he/she had. Nowadays bitcoin is, in terms of market capitalization, dominant compared to bitcoin cash (see figure 4). For more details see [1].

2.5 Summary

The bitcoin and blockchain technology depend upon a publicly available database of all transactions which is saved redundantly in a peer to peer network. The chain or database consists of blocks that are ordered chronological and each block contains transactions, a reference to the previous one and the solution to a difficult mathematical puzzle (the nonce that solves the block as a proof of work). Since all transactions are publicly available, all one needs to transfer bitcoins is the private key to sign and claim ownership of the output of a previous transaction and once the transaction is broadcasted to the network and included into a block of the chain the transaction is accepted and can be referenced to for new transfers. The problem of double-spending or cheating is solved by the proof-of-work that is required for creating a block since changing a block and especially the transactions in it is hard. An attacker would have to recreate the whole chain that came after the block the transaction is in and especially surpass the valid chain since the network considers the longest chain in terms of total difficulty to be the honest one. This gets computational harder the longer the chain is, since creating each block requires a proof-of-work and the chain grows all the time, because on average all 10 minutes a block is added (so the attacker would need a higher hash rate than the whole network). The incentive for denoting PC resources (mining) to block creation is that brand new bitcoins (see generation transaction) are created when a block is solved and transferred to the miner. All bitcoins are created through block creation and the total amount of bitcoin is capped at 21 million.

Advantages and Disadvantages of bitcoin payment system:

The capped money supply is one of the advantages since dilution of the money base due to political decisions is impossible. Another one is that no central authority and trusted parties are required so the whole payment system doesn't rely on trust but instead on a cryptographic proof of work and the idea that no attacker could amass more CPU power than the combined bitcoin network to corrupt the public transaction database (i.e. blockchain). This also means that no one can change a transaction after it is included in the chain which could be seen as advantage or disadvantage. Another point is the anonymity if one properly hides his identity when creating the wallet that manages and stores the private keys which I guess is especially loved by criminals. One clear drawback of the technology is that it can take a considerable time to clear a transaction (create a block) since time to creation is stochastic and on average only all 10 minutes a new one is created which means that on average you would have to wait so long until the transaction is accepted. This makes bitcoin infeasible for a payment system for everyday life.
Another point to mention here is that the price of bitcoin against major currencies (e.g. USD) is purely determined by supply and demand. Therefore the price is highly volatile and the value against currencies is only underpinned by the trust of the users that bitcoin will be used in the future and there is demand for bitcoins.

summary of bitcoin features:

- capped bitcoin supply, so no dilution due to political considerations
- new supply distributed into the system through predefined schedule
- no trusted third parties (i.e. banks) needed
- no national organization behind it

- integrity of the system is guaranteed by cryptography
- no changes to accepted transactions possible (for example no way of correcting fat finger errors)
- anonymity of transactions when identity is hidden properly
- takes a random amount of time to clear transactions (can range from minutes to hours)
- due to the increasing popularity of bitcoin a spectacular rise in the price from 0 to nearly 20000 USD
- extreme volatile price
- less regulation and security than for normal currencies
- open-source bitcoin protocol
- if private key lost no possibility of accessing ones bitcoins

3 Stylized facts of financial time series

In this section, we compare the statistical properties of the return times series of different financial markets. It is well known that financial return series possess specific statistical characteristics, see [47], [39], and [45], like

- log prices are integrated of order one (see 8 below), i.e. first differences are stationary
- returns are stationary, see definition 4
- returns possess little to no linear serial correlation
- squared or absolute returns have a high and slowly decaying serial correlation
- volatility clustering
- returns are not normally distributed and typically possess heavier tails than the normal distribution
- returns are asymmetrically distributed, i.e. the skewness is non zero

In this master-thesis, I analyze whether crypto-currencies (e.g. bitcoin, ether, ...) possess the same statistical characteristics like prices of well established major stock indices (e.g. S&P 500, Dow Jones, DAX, Nikkei 225, ...). So the question is if new crypto-assets behave differently than traditional financial assets from a statistical viewpoint.
First I will give a short review of the definitions and theory in section 3.1 used to analyze the aforementioned statistical properties.

3.1 Some facts of time series

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathbf{X} : \Omega \times I \to \mathbb{R}$ be a stochastic process where I denotes a index set that will represent the time domain, here it will be $I = \mathbb{N}$ or $I = \{1, ..., T\}$ due to the nature of financial data.

Definition 3. *A stochastic process $(\mathbf{X}_t)_{t \in I}$ is strict stationary if*

$$(\mathbf{X}_{t_1+k}, ..., \mathbf{X}_{t_n+k}) \stackrel{D}{=} (\mathbf{X}_{t_1}, ..., \mathbf{X}_{t_n})$$

for all $n \in \mathbb{N}, t_1, ..., t_n \in I$, $t_1 + k, ..., t_n + k \in I$, for all $k \in \mathbb{N}$

In this master thesis the concept of strict stationary is not used but instead the following

Definition 4. *A stochastic process $(\mathbf{X}_t)_{t \in I}$ is weakly stationary if*

(a) $\mathbb{E}(\mathbf{X}_t) = \mathbb{E}(\mathbf{X}_1)$ *for all* $t \in I$

(b) $\mathbb{E}(\mathbf{X}_t^2) < \infty$ *for all* $t \in I$

(c) $cov(\mathbf{X}_t, \mathbf{X}_{t+k}) = cov(\mathbf{X}_k, \mathbf{X}_1) =: \gamma(k)$ *for all* $t \in I$ *and* $k \in \mathbb{N}$

Remark. *$\gamma(k)$ is called the autocovariance function and $\rho(k) := \gamma(k)/\gamma(0)$ is called the autocorrelation function (acf).*

Definition 5. *A stochastic process $(\epsilon_t)_{t \in I}$ is white noise if*

(a) $\mathbb{E}(\epsilon_t) = 0$ *for all* $t \in I$

(b) $\mathbb{E}(\epsilon_t^2) = \sigma^2 < \infty$

(c) $cov(\epsilon_t, \epsilon_{t+k}) = 0$ *for all* $k \neq 0$

Remark. *Notation: $\epsilon_t \sim WN(\sigma^2)$*

Definition 6. *A stationary solution to the following equation (1) is called ARMA*(p, q) *process*

$$\mathbf{X}_t = a_1\mathbf{X}_t + ... + a_p\mathbf{X}_{t-p} + \epsilon_t + b_1\epsilon_{t-1} + ... + b_q\epsilon_{t-q} \tag{1}$$

where $\epsilon_t \sim WN(\sigma^2)$, $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}$ *and* $a_p \neq 0$ $b_q \neq 0$.

For further details about definition 6 see chapter 3 of [37].

Remark. *If* $q = 0$ *(i.e. all* $b_i = 0$*) in (1) we call* $\mathbf{X}_t$ *an AR*(p) *process.*

Definition 7. *A stochastic process* $(\mathbf{X}_t)_{t\in I}$ *is called integrated of order* l *if the* l*-th difference is stationary (where* $l \in \mathbb{N}$, *is the smallest* l *such that* $\Delta^l\mathbf{X}_t$ *is stationary with* $\Delta\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1}$ *and* $\Delta^l\mathbf{X}_t = \underbrace{\Delta\ldots\Delta}_{l-times}\mathbf{X}_t$*).*

Definition 8. *A stochastic process* $(\mathbf{X}_t)_{t\in I}$ *is called ARIMA*(l, p, q) *process (i.e. integrated of order* l*) if the* l*-th difference of* $\mathbf{X}_t$ *is an ARMA*(p, q) *process.*

Definition 9. *A process* $\mathbf{X}_t = x_0 + \sum_{j=1}^{t}\epsilon_j$ *for* $\epsilon_t \sim WN(\sigma^2)$ *iid and* $x_0 \in \mathbb{R}$ *(deterministic) is called random walk.*
$\tilde{\mathbf{X}}_t = t\theta + \mathbf{X}_t = x_0 + t\theta + \sum_{j=1}^{t}\epsilon_j$ *is called random walk with deterministic drift.*

For a random walk we can easily calculate:

- $\mathbb{E}(\mathbf{X}_t) = \sum_{j=1}^{t}\mathbb{E}(\epsilon_j) = 0$
- $\mathbb{V}\text{ar}(\mathbf{X}_t) = t\mathbb{V}\text{ar}(\epsilon_1) = t\sigma^2$
- $\text{cov}(\mathbf{X}_t, \mathbf{X}_s) = \sum_{j=1}^{\min(t,s)}\mathbb{V}\text{ar}(\epsilon_1) = \min(t, s)\mathbb{V}\text{ar}(\epsilon_1) = \min(t, s)\sigma^2$

Remark. *A random walk is integrated of order 1 and not stationary since the variance grows in* t*, but the first differences are iid white noise and therefore stationary. A random walk is a discrete-time analogon to a Brownian motion.*

Next I briefly introduce the estimation of the autocovariance and correlation function, since most of the test statistics used for analyzing the financial timeseries are based on these quantities.
The mean of a stationary process is estimated by the standard arithmetic mean:

$$\bar{\mathbf{X}}_T := \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t \tag{2}$$

and the autocovariancefunction can be estimated by

$$\hat{\gamma}(k) := \frac{1}{T}\sum_{t=1}^{T-k}(\mathbf{X}_{t+k} - \bar{\mathbf{X}}_T)(\mathbf{X}_t - \bar{\mathbf{X}}_T) \tag{3}$$

and the standard estimator for the autocorrelationfunction is

$$\hat{\rho}(k) := \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \tag{4}$$

for a detailed analysis under which conditions 2, 3 and 4 are consistent estimators for the population quantities and are asymptotically normal see [37].

The Box-Pierce and Ljung-Box tests are standard test for a white noise process. The H_0 hypothesis is that $\mathbf{X}_t$ is iid (therefore also $\mathbf{X}_t \sim \mathrm{WN}(\sigma^2)$). The test statistics are

- Box-Pierce: $Q(k) = T\sum_{i=1}^{k}\hat{\rho}(i)^2 \xrightarrow{D} \chi_k^2$ under H_0.
- Ljung-Box: $Q(k) = T(T+2)\sum_{i=1}^{k}\frac{\hat{\rho}(i)^2}{T-i} \xrightarrow{D} \chi_k^2$ under H_0. The Ljung-Box test has better performance for finite samples

Furthermore define the estimator for the $l \in \mathbb{N}$ centered moment $(\mathbb{E}\left(\mathbf{X}_t - \mathbb{E}(\mathbf{X}_t)\right)^l)$ as

$$m_l := \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t - \bar{\mathbf{X}}_T)^l \tag{5}$$

The skewness-estimate is

$$g_T := \frac{m_3}{m_2^{3/2}} \tag{6}$$

which is a measure of asymmetry of a distribution (i.e. $g_T < 0$ indicates that the distribution is skewed to the left and vice verse). An estimator of the kurtosis is

$$\kappa_T := \frac{m_4}{m_2^2} \tag{7}$$

which is a measure for heaviness of the tails. The standard normal distribution has kurtosis equal to 3, therefore a kurtosis higher than 3 indicates that the tails of the distribution are heavier than the normal density tails.

Under $\mathbf{X}_t$ iid standard normal it holds:

- $g_T\sqrt{\frac{T}{6}} \to N(0,1)$
- $(\kappa_T - 3)\sqrt{\frac{T}{24}} \to N(0,1)$

One of the most prominent tests for normality (H_0 states that $(\mathbf{X}_1, ..., \mathbf{X}_t)$ is iid normal distributed) is based on the skewness and kurtosis estimate and called Jarque Bera test, see [40]. The test statistic J is given by

$$J = \frac{T}{6}\left(g_T^2 + \frac{(\kappa_T - 3)^2}{4}\right) \xrightarrow{D} \chi_2^2 \tag{8}$$

3.2 Convergence of random walks

Definition 10. *A stochastic process $(W_t)_{t\in[0,1]}$ is called Brownian motion if*

- $W_0 = 0$ *a.s.*
- $t \to W_t$ *is continuous a.s.*
- *for all $n \in \mathbb{N}$ $t_0 < ... < t_n$ the vector with the increments $(W_{t_1} - W_{t_0}, ..., W_{t_n} - W_{t_{n-1}})$ has independent components with $W_{t_i} - W_{t_{i-1}} \sim N(0, t_i - t_{i-1})$*

Definition 11. *For $s \in [0,1]$ and $\epsilon \sim WN(\sigma^2)$ iid, define*

$$\mathbf{X}_T(s) := \frac{1}{\sqrt{\sigma^2 T}} \sum_{t=1}^{\lfloor sT \rfloor} \epsilon_t$$

Lemma 12. *$\mathbf{X}_T(s)$ converges for fixed but arbitrary $s \in [0,1]$ to $N(0,s)$ in distribution.*

Proof of Lemma 12: *From the CLT we have*

$$\mathbf{X}_T(s) = \underbrace{\frac{\sqrt{\lfloor sT \rfloor}}{\sqrt{T}}}_{\to\sqrt{s}} \underbrace{\frac{1}{\sqrt{\sigma^2 \lfloor sT \rfloor}} \sum_{t=1}^{\lfloor sT \rfloor} \epsilon_t}_{\to N(0,1)} \to N(0,s)$$

since $sT - 1 \leq \lfloor sT \rfloor \leq sT$ therfore $s - 1/T \leq \lfloor sT \rfloor / T \leq s$ □

Theorem 13. ***Donsker Theorem:***
The process $(\mathbf{X}_T(s))_{s\in[0,1]}$ converges to $(W_s)_{s\in[0,1]}$ in distribution

To be more prices it converges in the Skorokhod topology but this is beyond the scope of this thesis and for more details see chapters 2 and 3 of [18] or [32].
Furthermore there is a continuous mapping theorem for processes (i.e. for all continuous functionals $g(\cdot)$ we have $g((\mathbf{X}_T(s))_{s\in[0,1]}) \xrightarrow{D} g((W_s)_{s\in[0,1]})$, where continuity is with respect to the Skorokhod metric). All functionals that come up in this thesis are continuous, further details can be found in [52].

Next a few convergence results for random walk processes are stated and proofed that follow from theorem 13. These results will be used to construct tests for stationarity, since this is one of the main properties of crypto-assets that we want to investigate. The importance of stationarity comes from the fact that this allows one to estimate expectations by averaging over the time domain, since otherwise we would have for just one observation for each t. Therefore, if we want to model something in a time-series context we have to have some sort of stationarity. The tests, applied in section 4, do indeed indicate a positve answere.

Observe that we have for $\mathbf{X}_t$ a random walk ($\mathbf{X}_t = \sum_{j=1}^{t} \epsilon_j$ for the rest of the section),

$$\frac{1}{\sqrt{T^3\sigma^2}}\sum_{t=1}^{T}\mathbf{X}_t = \frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{X}_t}{\sqrt{T\sigma^2}} = \int_0^1 \mathbf{X}_T(s)ds = g(\mathbf{X}_T(\cdot)) \xrightarrow{D} \int_0^1 W_s ds \tag{9}$$

where the last equality is due to $\mathbf{X}_T(\cdot)$ being a step function and the functional $g(h) = \int_0^1 h(s)ds$. The convergence follows from [32] and the continuous mapping theorem.

Remark. *For a stationary AR(p) process, $\mathbf{X}_t$, we would have (under some regularity conditions, see chapter 3 and 7 of [37])*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{X}_t - \mathbb{E}(\mathbf{X}_t)) \xrightarrow{D} N(0, \tilde{\sigma}^2)$$

where the long run variance is given by $\tilde{\sigma}^2 = \sum_{j=-\infty}^{\infty}\gamma(j)$ with $\gamma(j) = cov(\mathbf{X}_{t+j}, \mathbf{X}_t)$. Note the totally different scaling factor compared to the result for the white noise process.

Furthermore

$$\frac{1}{T^2\sigma^2}\sum_{t=1}^{T}\mathbf{X}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t}{\sqrt{T\sigma^2}}\right)^2 = \int_0^1 \mathbf{X}_T^2(s)ds = g(\mathbf{X}_T(\cdot)) \xrightarrow{D} \int_0^1 W_s^2 ds \tag{10}$$

for $g(h) = \int_0^1 h^2(s)ds$ and

$$\frac{1}{T}\sum_{t=1}^{T}(\Delta \mathbf{X}_t)^2 = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \to 1 \quad a.s. \tag{11}$$

due to the strong law of large numbers. Also

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T}\mathbf{X}_{t-1}\Delta\mathbf{X}_t = \frac{1}{2}\frac{1}{T\sigma^2}\sum_{t=1}(\mathbf{X}_t^2 - \mathbf{X}_{t-1}^2 - (\Delta\mathbf{X}_t)^2) = \frac{1}{2}(\underbrace{\frac{1}{T\sigma^2}\mathbf{X}_T^2}_{\mathbf{X}_T(1)} - \frac{1}{T}\sum_{t=1}^{T}(\Delta\mathbf{X}_t)^2) \tag{12}$$

$$\to \frac{1}{2}(W_1^2 - 1) = \int_0^1 W_s dW_s$$

where the last equality is due to the Ito formula for $f(x) = x^2$. Next

$$\frac{1}{\sqrt{T^3\sigma^2}}\sum_{t=1}^{T}t\Delta\mathbf{X}_t = \frac{\mathbf{X}_T}{\sqrt{T\sigma^2}} - \frac{1}{T}\sum_{t=1}^{T-1}\frac{\mathbf{X}_t}{\sqrt{T\sigma^2}} \to W_1 - \int_0^1 W_s ds = \int_0^1 s dW_s \tag{13}$$

where the last equality is due to the Ito formula for $f(x,t) = xt$.

$$\frac{1}{\sqrt{T^5\sigma^2}}\sum_{t=1}^{T}t\mathbf{X}_t = \frac{1}{T}\sum_{t=1}^{T}\frac{t}{T}\frac{\mathbf{X}_t}{\sqrt{T\sigma^2}} = \int_0^1 s\mathbf{X}_T(s)ds \to \int_0^1 sW_s ds \tag{14}$$

Next we will look at a random walk with deterministic drift, $\tilde{\mathbf{X}}_t = t\theta + \mathbf{X}_t$ and collect some results about convergence. First observe that

$$\frac{1}{T^{d+1}}\sum_{t=1}^{T}t^d = \sum_{t=1}^{T}\frac{1}{T}\left(\frac{t}{T}\right)^d \to \int_0^1 x^d dx = \frac{1}{d+1} \tag{15}$$

since the sum in (15) is the upper sum (e.g. compare with Riemann integral) of the integral on the right hand side. Then we have

$$\frac{1}{T^2}\sum_{t=1}^{T}\tilde{\mathbf{X}}_t = \underbrace{\frac{\theta}{T^2}\sum_{t=1}^{T}t}_{\to\theta/2} + \underbrace{\frac{1}{\sqrt{T}}}_{\to 0}\underbrace{\frac{1}{T^{3/2}}\sum_{t=1}^{T}\mathbf{X}_t}_{\to\int_0^1 W_s ds} \to \frac{\theta}{2} \tag{16}$$

where the first term converges due to (15) with $d = 1$ and the second term due to (9).

Remark. *Note that for a random walk with drift ($\tilde{\mathbf{X}}_t$) the deterministic drift dominates the random walk and different scaling factors (i.e. the scaling factor of $\mathbf{X}_t$ is multiplied by $1/T^{1/2}$ to get the scaling factor for $\tilde{\mathbf{X}}_t$) are needed to obtain convergence. Analogous to (16) we can handle all the sums we have analyzed for $\mathbf{X}_t$ also for $\tilde{\mathbf{X}}_t$. Further details can also be found in section 17.4 in [37].*

Collecting results for $\mathbf{X}_t$ and $\tilde{\mathbf{X}}_t$ we get:

- $\frac{1}{T^{3/2}}\sum_{t=1}^T \mathbf{X}_t \to \sigma\int_0^1 W_s ds$
- $\frac{1}{T^2}\sum_{t=1}^T \mathbf{X}_t^2 \to \sigma^2\int_0^1 W_s^2 ds$
- $\frac{1}{T}\sum_{t=1}^T (\Delta\mathbf{X}_t)^2 \to \sigma^2$
- $\frac{1}{T}\sum_{t=1}^T \mathbf{X}_{t-1}\Delta\mathbf{X}_t \to \sigma\int_0^1 W_s dW_s$
- $\frac{1}{T^{5/2}}\sum_{t=1}^T t\mathbf{X}_t \to \sigma\int_0^1 sW_s ds$
- $\frac{1}{T^{3/2}}\sum_{t=1}^T t\Delta\mathbf{X}_t \to \sigma\int_0^1 s dW_s$
- $\frac{1}{T^{1/2}}\sum_{t=1}^T \Delta\mathbf{X}_t \to \sigma^2 W_1$
- $\frac{1}{T^2}\sum_{t=1}^T \tilde{\mathbf{X}}_t \to \frac{\theta}{2}$
- $\frac{1}{T^3}\sum_{t=1}^T \tilde{\mathbf{X}}_t^2 \to \frac{\theta^2}{3}$
- $\frac{1}{T}\sum_{t=1}^T (\Delta\tilde{\mathbf{X}}_t)^2 \to \theta^2+\sigma^2$
- $\frac{1}{T^2}\sum_{t=1}^T \tilde{\mathbf{X}}_{t-1}\Delta\tilde{\mathbf{X}}_t \to \frac{\theta^2}{2}$
- $\frac{1}{T^3}\sum_{t=1}^T t\tilde{\mathbf{X}}_t \to \frac{\theta}{3}$
- $\frac{1}{T^2}\sum_{t=1}^T t\Delta\mathbf{X}_t \to \frac{\theta}{2}$
- $\frac{1}{T}\sum_{t=1}^T \Delta\tilde{\mathbf{X}}_t \to \theta$

From this we can deduce

$$\frac{1}{T}\hat{\gamma}(0) = \frac{1}{T^2}\sum_t(\mathbf{X}_t-\bar{\mathbf{X}}_T)^2 = \frac{1}{T^2}\sum_t \mathbf{X}_t^2 - \frac{1}{T}\bar{\mathbf{X}}_T^2 \to \sigma^2\left(\int_0^1 W_s^2 ds - (\int_0^1 W_s ds)^2\right) \quad (17)$$

$$\begin{aligned}\frac{1}{T}\hat{\gamma}(1) &= \frac{1}{T^2}\sum_t(\underbrace{\mathbf{X}_{t+1}}_{\mathbf{X}_t+\epsilon_{t+1}}-\bar{\mathbf{X}}_T)(\mathbf{X}_t-\bar{\mathbf{X}}_T) = \frac{1}{T^2}\sum_t \mathbf{X}_t^2 + \frac{1}{T}\underbrace{\frac{1}{T}\sum_t \mathbf{X}_t\underbrace{\epsilon_{t+1}}_{\Delta\mathbf{X}_{t+1}}}_{\to\sigma\int_0^1 W_s dW_s} - \frac{1}{T}\bar{\mathbf{X}}_T^2 \\ &\to \sigma^2\left(\int_0^1 W_s^2 ds - (\int_0^1 W_s ds)^2\right) \\ \hat{\rho}(1) &= \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \to 1 \quad \text{in probability} \\ \hat{\rho}(k) &= \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \to 1 \quad \text{in probability} \quad (18)\end{aligned}$$

where the statement (18) uses the fact that convergence in probability is equivalent to convergence in distribution if the limit is a constant. Further (18) can be proofed analogous to the previous statement.

3.3 Unit Root Tests

Now we are ready to derive the Dickey Fuller test (DF test) for more details see [31] and [50].

DF Test Typ A:

The basic DF test (Type A) starts from a simple AR(1) model:

$$\mathbf{X}_t = a\mathbf{X}_{t-1} + \epsilon_t \tag{19}$$

where $\epsilon \sim \text{WN}(\sigma^2)$ is iid. It can be easily seen that if $a = 1$ and $\mathbf{X}_0 = 0$ the solution is a white noise process. We want to test if $\mathbf{X}_t$ is a white noise process (i.e. H_0: $a = 1$) versus the hypothesis that $\mathbf{X}_t$ is a stationary AR(1) process (i.e. H_1: $a < 1$). Testing for $a = 1$ is called unit root test.
Therefore we subtract $\mathbf{X}_{t-1}$ on both sides of (19) and obtain

$$\Delta\mathbf{X}_t = \underbrace{(a-1)}_{=:\delta}\mathbf{X}_{t-1} + \epsilon_t = \delta\mathbf{X}_{t-1} + \epsilon_t \tag{20}$$

The idea is simple that we estimate δ by ordinary least squares and test $H_0 : \delta = 0$ versus $H_1 : \delta < 1$. This is done by using a t-test-statistic but under H_0 it will not have a t-distribution.

To see this, define:

- $Y = (\Delta\mathbf{X}_T, ..., \Delta\mathbf{X}_2)' \in \mathbb{R}^{T-1}$
- $X = (\mathbf{X}_{T-1}, ..., \mathbf{X}_1)' \in \mathbb{R}^{T-1}$

Then the estimate $\hat{\delta}$ is given by

$$\hat{\delta} = (X'X)^{-1}X'Y = \frac{\sum_t \mathbf{X}_{t-1}\Delta\mathbf{X}_t}{\sum_t(\Delta\mathbf{X}_t)^2} = \frac{1}{T}\frac{\frac{1}{T\sigma^2}\sum_t \mathbf{X}_{t-1}\Delta\mathbf{X}_t}{\frac{1}{T^2\sigma^2}\sum_t(\Delta\mathbf{X}_t)^2}$$

therefore we have using (12) and (11) that

$$T\hat{\delta} = \frac{\frac{1}{T\sigma^2}\sum_t \mathbf{X}_{t-1}\Delta\mathbf{X}_t}{\frac{1}{T^2\sigma^2}\sum_t(\Delta\mathbf{X}_t)^2} \xrightarrow{D} \frac{\int_0^1 W_s dW_s}{\int_0^1 W_s^2 ds} \tag{21}$$

Remark. *Note that in (21) we again have a different scaling factor (i.e. T) compared to standard statistical theory where usually $\sqrt{T}$ occurs as the right scaling to get convergence to a random variable. This is called superconsistentency (i.e. $\hat{\delta} \to 0$ with rate $1/T$).*
For an $AR(1)$ process we would have $\sqrt{T}\hat{\delta} \to N(0, 1 - a^2)$. The distribution especially quantiles of $\int_0^1 W_s dW_s / \int_0^1 W_s^2 ds$ are computed using simulations. Furthermore note that $\{t \in [0,1] \times \omega \in \Omega : W_t(\omega) = 0\}$ is a $\lambda \times P$ (where λ is the Lebesgue measure) null set therefore we have no problem in the denominator.

The least squares estimator for σ^2 is given by (recall $\hat{Y} = \hat{\delta}X$ and $\hat{Y} \perp (Y - \hat{Y})$)

$$\hat{\sigma}^2 = \frac{1}{T}\|Y - \hat{Y}\|_2^2 = \frac{1}{T}\|Y\|_2^2 - \frac{1}{T}\underbrace{\|\hat{Y}\|_2^2}_{=\hat{\delta}^2\|X\|_2^2} = \underbrace{\frac{1}{T}\sum_t (\Delta\mathbf{X}_t)^2}_{\to\sigma^2} - \overbrace{\underbrace{\hat{\delta}^2}_{O_P(T^{-2})} \, T \underbrace{\frac{1}{T^2}\sum_t \mathbf{X}_{t-1}^2}_{\to\sigma^2\int_0^1 W_s^2 ds}}^{\to 0} \to \sigma^2 \tag{22}$$

by using (11), (10) and (21).

The test statistic τ of a t-test for $H_0 : \delta = 0$ is given by

$$\tau = \frac{\hat{\delta}}{\sqrt{\hat{\sigma}^2(X'X)^{-1}}} = \frac{\sum_t \mathbf{X}_{t-1}\Delta\mathbf{X}_t}{\hat{\sigma}(\sum_t \mathbf{X}_t^2)(\sum_t \mathbf{X}_t^2)^{-1/2}} = \frac{\frac{1}{T\sigma^2}\sum_t \mathbf{X}_{t-1}\Delta\mathbf{X}_t}{\frac{\hat{\sigma}}{\sigma}(\frac{1}{T^2\sigma^2}\sum_t \mathbf{X}_t^2)^{1/2}} \to \frac{\int_0^1 W_s dW_s}{\sqrt{\int_0^1 W_s^2 ds}} \tag{23}$$

The quantiles q_α of the distributions are again calculated by simulations and can be found in [37] the most important ones are in Table 2

Table 2: DF Test Type A cirtical values

α	10%	5%	2.5%	1%
q_α for $T\hat{\delta}$	-5.7	-8.1	-10.5	-13.8
q_α for τ	-1.62	-1.95	-2.23	-2.58

The test and critical values are implemented in R in the library tseries. For comparison in the classic statistical setting $\tau \sim t_T \to N(0,1)$ and the critical values for the standard normal are given by $-1.28, -1.65, -1.96, -2.33$.

DF Test Type B:

The type B DF test is a slight modification of type A. The model is given by:

$$\Delta \mathbf{X}_t = \theta + \delta \mathbf{X}_{t-1} + \epsilon_t \tag{24}$$

and the H_0 hypothesis is again that $\delta = 0$ and $\theta = 0$ (i.e. $\mathbf{X}_t = \mathbf{X}_0 + \sum_{j=1}^{t} \epsilon_j$ is a random walk plus a constant if $\mathbf{X}_0 \neq 0$). The H_1 is $\delta < 1$, i.e. $\mathbf{X}_t$ is an AR(1) process plus constant. We again estimate (θ, δ) by ordinary least squares. Define

- $Y = (\Delta \mathbf{X}_T, ..., \Delta \mathbf{X}_2)' \in \mathbb{R}^{T-1}$
- $X = (\mathbf{X}_{T-1}, ..., \mathbf{X}_1)' \in \mathbb{R}^{T-1}$

then through the estimate for δ and θ are given by

$$\begin{aligned}
\hat{\delta} &= \frac{\hat{\text{cov}}(Y, \mathbf{x})}{\hat{\mathbb{V}\text{ar}}(\mathbf{x})} = \frac{\frac{1}{T}\sum_t \Delta \mathbf{X}_t \mathbf{X}_{t-1} - (\frac{1}{T}\sum_t \Delta \mathbf{X}_t)(\frac{1}{T}\sum_t \mathbf{X}_{t-1})}{\frac{1}{T}\sum_t \mathbf{X}_{t-1}^2 - (\frac{1}{T}\sum_t \mathbf{X}_{t-1})^2} = \\
&= \frac{1}{T} \frac{\frac{1}{T}\sum_t \Delta \mathbf{X}_t \mathbf{X}_{t-1} - (\frac{1}{T}\sum_t \Delta \mathbf{X}_t)(\frac{1}{T}\sum_t \mathbf{X}_{t-1})}{\frac{1}{T^2}\sum_t \mathbf{X}_{t-1}^2 - (\frac{1}{T^{3/2}}\sum_t \mathbf{X}_{t-1})^2} \\
\hat{\theta} &= \bar{Y} - \hat{\delta}\bar{X}
\end{aligned}$$

where bar over a vector indicates the ermpirical mean. With the same techniques as for type A we get

$$T\hat{\delta} = \frac{\frac{1}{T}\sum_t \Delta \mathbf{X}_t \mathbf{X}_{t-1} - (\frac{1}{T}\sum_t \Delta \mathbf{X}_t)(\frac{1}{T}\sum_t \mathbf{X}_{t-1})}{\frac{1}{T^2}\sum_t \mathbf{X}_{t-1}^2 - \frac{1}{T^3}(\sum_t \mathbf{X}_{t-1})^2} \rightarrow \frac{\int_0^1 W_s dW_s - W_1 \int_0^1 W_s ds}{\int_0^1 W_s^2 ds - (\int_0^1 W_s ds)^2} \tag{25}$$

Note due to Jensen inequality and since $\{t \times \omega : W_t(\omega) = 0\}$ is a set with measure 0 the denominator is always strictly positive. The t-test statistic τ is given by

$$\tau = \frac{\frac{1}{T}\sum_t \Delta \mathbf{X}_t \mathbf{X}_{t-1} - (\frac{1}{T}\sum_t \Delta \mathbf{X}_t)(\frac{1}{T}\sum_t \mathbf{X}_{t-1})}{\hat{\sigma}\sqrt{\frac{1}{T^2}\sum_t \mathbf{X}_{t-1}^2 - \frac{1}{T^3}(\sum_t \mathbf{X}_{t-1})^2}} \rightarrow \frac{\int_0^1 W_s dW_s - W_1 \int_0^1 W_s ds}{\sqrt{\int_0^1 W_s^2 ds - (\int_0^1 W_s ds)^2}} \tag{26}$$

Furthermore, we can also test for $(\theta, \delta) = 0 \in \mathbb{R}^2$ by the Wald test statistics (i.e. comparing the unrestricted model M1 (with 2 parameters) with the restricted one M2 (i.e. with 0 parameter since $(\theta, \delta) = 0 \in \mathbb{R}^2$)) which is $F(2, T-2)$-distributed in classic regression setups but here again follows a non standard distribution.

$$F = \frac{\frac{1}{2-0}(\text{RSS}_1 - \text{RSS}_2)}{\frac{1}{T-2}\text{RSS}_2} = \frac{\frac{1}{2}(\sum_t (\Delta \mathbf{X}_t)^2 - \sum_t (\Delta \mathbf{X}_t - \hat{\theta} - \hat{\delta}\mathbf{X}_{t-1})^2)}{\frac{1}{T-2}\sum_t (\Delta \mathbf{X}_t - \hat{\theta} - \hat{\delta}\mathbf{X}_{t-1})^2}$$

where RSS_i is the residual sum of squares of model i. Since it does not add a lot value no more details about the wald statistc are presented.

DF Test Type C:

For the type C test the model is:

$$\Delta \mathbf{X}_t = \theta_0 + \theta_1 t + \delta \mathbf{X}_{t-1} + \epsilon_t \tag{27}$$

Here the H_0 is $(\theta_1, \delta) = 0 \in \mathbb{R}^2$, i.e. $\mathbf{X}_t = \mathbf{X}_0 + \theta_0 t + \sum_j^t \epsilon_j$ a random walk with linear deterministic trend plus constant (if $\mathbf{X}_0 \neq 0$) if $\theta_1 \neq 0$ a quadratic deterministic trend would be present. The $H_1 : \delta < 1$ is that $\mathbf{X}_t$ is an AR(1) process with linear deterministic trend and plus a constant. Then again $(\theta_0, \theta_1, \delta)$ is estimated by ordinary least squares, considere the same test statistics as in type B but the distribution under the H_0 is different again. I will not present more details here, since it is quite technical and does not offer additional insight.

Augmented DF Test (ADF):

For the normal DF test, the errors ϵ_t are assumed to be iid. This might be too restrictive in some cases since no correlation structure can be modeled in this way. There are a few tests (Phillips-Perron, ADF, KPSS test) that circumvent this problem. The most prominent one is the Augmented Dickey Fuller (ADF) test where the error terms in the AR equation for $\mathbf{X}_t$ are modeled as AR(p) stationary process. As for the DF test, there are type A,B, and C.

ADF Type A:

The type A model is given by:

$$\mathbf{X}_t = a\mathbf{X}_{t-1} + u_t \tag{28}$$

$$u_t = a_1 u_{t-1} + ... + a_p u_{t-p} + \epsilon_t \tag{29}$$

We want to test $H_0 : a = 1$ (i.e. $\mathbf{X}_t$ is an ARIMA$(p-1, 1, 0)$ process, see 8) versus $H_1 : a < 1$ (i.e. $\mathbf{X}_t$ is an AR(p) process, see equation (30)). Next the model is rewritten by inserting $u_t = \mathbf{X}_t - a\mathbf{X}_{t-1}$ into

$$\Delta \mathbf{X}_t = (a-1)\mathbf{X}_{t-1} + u_t = (a-1)\mathbf{X}_{t-1} + a_1 \underbrace{u_{t-1}}_{\mathbf{X}_{t-1} - a\mathbf{X}_{t-2}} + ... + a_p \underbrace{u_{t-p}}_{\mathbf{X}_{t-p} - a\mathbf{X}_{t-p-1}} + \epsilon_t \tag{30}$$

it can be shown that (30) can be reparametrized as

$$\Delta \mathbf{X}_t = \delta \mathbf{X}_{t-1} + \tilde{a}_1 \Delta \mathbf{X}_{t-1} + ... + \tilde{a}_p \Delta \mathbf{X}_{t-p} + \epsilon_t$$

where $\delta = (a-1)\underbrace{(1 - a_1 - ... - a_p)}_{>0}$ (since we assumed u_t to be stationary we have $a(z) = (1 - a_1 z - ... - a_p z^p) \neq 0$ for all $|z| \leq 1$ and since $a(0) = 1 > 0$ we also have $a(z) > 0$ for all $|z| \leq 1$). Therefore we test $H_0 : \delta = 0$ versus $H_1 : \delta < 0$ again. The parameters are estimated using least squares and follow a non standard distribution. The critical values are computed by simulations.

ADF Type B:

The type B model is given by:

$$\Delta \mathbf{X}_t = \theta + \delta \mathbf{X}_{t-1} + \tilde{a}_1 \Delta \mathbf{X}_{t-1} + ... + \tilde{a}_p \Delta \mathbf{X}_{t-p} + \epsilon_t$$

Analog to the DF type B the $H_0 : \delta < 0, \theta = 0$ hypothesis states that $\mathbf{X}_t$ is an ARIMA($p-1, 1, 0$) process plus a constant (i.e. $\mathbf{X}_0 \neq 0$) versus $H_1 : \mathbf{X}_t$ is an AR(p) process plus constant.

ADF Type C:

The type C model is given by:

$$\Delta \mathbf{X}_t = \theta_0 + \theta_1 t + \delta \mathbf{X}_{t-1} + \tilde{a}_1 \Delta \mathbf{X}_{t-1} + ... + \tilde{a}_p \Delta \mathbf{X}_{t-p} + \epsilon_t$$

Analog to the DF type C the $H_0 : \delta < 0, \theta = 0$ hypothesis states that $\mathbf{X}_t$ is an ARIMA($p-1, 1, 0$) process plus a linear trend (plus a constant if $\mathbf{X}_0 \neq 0$) versus $H_1 : \mathbf{X}_t$ is an AR(p) process with linear trend plus constant.

Remark. *For applying the ADF test one needs to choose the lag order p. It can be shown that the asymptotic theory described above holds as long as $p \leq O(T^{1/3})$ and in practice if no knowledge about how to choose the lag is available one sets $p_{max} = \lfloor (T-1)^{1/3} \rfloor$ or tests with different $p \in \{1, ..., p_{max}\}$ and one hopes to find a consistent pattern.*

Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Test:

The KPSS test is another test that can be used to distinguish between a stationary or integrated process. The main difference is that for the ADF tests the H_0 represents non-stationarity (i.e. ARIMA) whereas for the KPSS test the roles are switched around. The model is given by:

$$\Delta \mathbf{X}_t = u_t - b u_{t-1} \tag{31}$$

where $u_t = \sum_{j\geq 0} k_j \epsilon_{t-j}$ is a centered stationary process with certain technical assumptions (i.e. ϵ_j iid white noise with finite fourth moments and $\sum_{j\geq 0} j|k_j| < \infty$). For more details about the KPSS test see [43].
We test $H_0 : b = 1$ (i.e. $\mathbf{X}_t = \theta + u_t$ stationary[1], i.e. $\theta = \mathbf{X}_0 = \mathbb{E}(\mathbf{X}_t)$ if $\mathbf{X}_0$ is deterministic) versus $H_1 : b < 1$ (i.e. $\mathbf{X}_t = \theta + \sum_{j=1}^{t}(u_t - b u_{t-1})$ non stationary).

Next, for simplicity, I will briefly introduce the test statistic under the assumption u_t iid (i.e. $k_j = 0$ for $j > 0$). The KPSS test is related to the (unobserved) random walk $S_t = \sum_{j=1}^{t} u_j$. The estimates are:

$$\hat{\theta} = \bar{\mathbf{X}}_T \tag{32}$$

$$\hat{u}_t = \mathbf{X}_t - \hat{\theta} = \mathbf{X}_t - \bar{\mathbf{X}}_T \tag{33}$$

$$\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j = S_t - \frac{t}{T} S_T \tag{34}$$

Then we have with the same reasoning as in theorem 12 and 13 that:

$$\frac{1}{\sqrt{T}} \hat{S}_{\lfloor sT \rfloor} = \frac{1}{\sqrt{T}} S_{\lfloor sT \rfloor} - \frac{\lfloor sT \rfloor}{\sqrt{T^3}} S_T \to \sigma(W_s - sW_1) =: \sigma B_s \tag{35}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} \hat{S}_t^2 \to \sigma^2 \int_0^1 B_s^2 ds \tag{36}$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 \to \sigma^2 \tag{37}$$

[1] This is the solution to $\Delta \mathbf{X}_t = \Delta u_t$

Remark. *The process* $(B_s)_{s\in[0,1]}$ *is called a Brownian bridge*[2] *since it is tied down at* 0 *for* $s = 0, 1$ *(i.e.* $B_0 = B_1 = 0$.

The test statistic for the KPSS test is then given by:

$$\frac{1}{T^2}\sum_{t=1}^{T}\frac{\hat{S}_t^2}{\hat{\sigma}^2} \to \int_0^1 B_s^2 ds = \|B.\|^2_{L^2([0,1])} \tag{38}$$

Remark. *The critical values of the KPSS test (38) are normally obtained by simulation. But using the Karhunen–Loève-Theroem, see [44], one sees that the distribution is a weighted* χ^2 *distribution.*
The Karhunen–Loève-theroem states that the Brownian bridge can be represented as:

$$B_t = \sum_{j=1}^{\infty} \xi_j \lambda_j \sqrt{2}\sin(j\pi t)$$

where $\lambda_j = 1/(j\pi)$ *and* $e_j(\cdot) = \sqrt{2}\sin(j\pi\cdot)$ *are the eigenvalues and eigenfunctions of the covariance operator*[3] *of the Brownian bridge and* $\xi_j(\omega) = \int_0^1 B_s(\omega)e_j(s)ds$ *are iid standard normal. The Karhunen–Loève-theroem states that this Fourier expansion (for a fixed* $\omega \in \Omega$ *view* $B.(\omega)$ *as element of* $L^2([0,1])$*) has certain optimality properties and the eigenfunctions form an orthonormal base of* $L^2([0,1])$*. Therefore we have:*

$$\|B.\|^2_{L^2([0,1])} = \int_0^1 B_s^2 ds = \sum_{i,j}\xi_i\xi_j \underbrace{\int_0^1 e_i(s)e_j(s)ds}_{\delta_{i,j}} = \sum_{j=1}^{\infty}\lambda_j^2\xi_j^2$$

Therefore the distribution is a infinitely weighted χ^2 *distribution with weights* λ_i^2 *and can be approximated by cutting of the infinite sum. The Karhunen–Loève-theroem states also that the eigenbasis is optimal to approximate by cutting of the sum.*
The same trick can be used for the Brownian motion:

$$W_t = \sum_{j=1}^{\infty}\xi_j \frac{1}{(j-1/2)\pi}\sqrt{2}\sin((j-1/2)\pi t)$$

[2] $\mathrm{cov}(B_s, B_t) = \min(t,s) - st$

[3] The covariance operator is given by: $C : L^2([0,1]) \to L^2([0,1])$ by $C(f)(t) := \int_0^1 \mathrm{cov}(B_s, B_t)f(s)ds$ and $C(e_j)(\cdot) = \lambda_j e_j(\cdot)$

Next I briefly touch the general case where u_t is not iid. In this setup the procedure stays the same except that we have to correct (i.e. divide) for the long run variance $\tilde{\sigma}^2 = \sum_{j=-\infty}^{\infty} \gamma_u(j) = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{t+j}u_t)$ since it can be shown that for $u_t = \sum_{j\geq 0} k_j \epsilon_{t-j}$:

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor sT \rfloor} u_t \to \tilde{\sigma}^2 W_s$$

To correct for the long run variance $\tilde{\sigma}^2$ we have to estimate it and replace the division by $\hat{\sigma}^2$ in (38) by $\hat{\tilde{\sigma}}^2$- To estimate the long run variance the procedure of the KPSS test is slightly adapted:

$$\begin{aligned}
\hat{\theta} &= \bar{\mathbf{X}}_T \\
\hat{u}_t &= \mathbf{X}_t - \hat{\theta} = \mathbf{X}_t - \bar{\mathbf{X}}_T \\
\hat{S}_t &= \sum_{j=1}^{t} \hat{u}_j = S_t - \frac{t}{T} S_T \\
\hat{\gamma}_u(k) &= \frac{1}{T} \sum_{j=1}^{T-k} \hat{u}_{t+k}\hat{u}_t \\
\hat{\tilde{\sigma}}^2 &= \sum_{k=-T+1}^{T-1} K(\frac{k}{m_T})\hat{\gamma}_u(k)
\end{aligned}$$

where $K(\cdot)$ is a kernel function (i.e. symmetric, density with expectation 0 and support $x : |x| \leq 1$, for example the Epanechnikov kernel) and m_T is a bandwidth parameter that controls the bias variance trade-off with $m_T \to \infty, m_T/T \to 0$.
The general KPSS teststatistic is then:

$$\frac{1}{T^2} \sum_{t=1}^{T} \frac{\hat{S}_t^2}{\hat{\tilde{\sigma}}^2} \to \int_0^1 B_s^2 ds = \|B.\|^2_{L^2([0,1])}$$

Remark. *Also for the KPSS test there exists a type B test where the model is enhanced by a constant:*

$$\Delta \mathbf{X}_t = \theta_1 + u_t - bu_{t-1} \tag{39}$$

and under $H_0 : b = 1$ *the solution is given by* $\mathbf{X}_t = \theta + \theta_1 t + u_t$ *versus* $H_1 : b < 1$ $\mathbf{X}_t = \theta + \theta_1 t + \sum_j^t u_j - bu_{j-1}$. *The procedure stays the same,* (θ_0, θ_1) *is estimated by* *least squares and only the asymptotic distribution of the test statistic (3.3) is the* $L^2([0,1])$ *norm of a second level Brownian bridge* $B_s^{(2)} := W_s + (2s - 3s^2)W_1 + 6(s^2 - s)\int_0^1 W_s ds$ *(*$B_0^{(2)} = B_1^{(2)} = 0$ *and* $\int_0^1 B_s^{(2)} ds = 1$*).*

3.4 Tail index estimation

In this section, the tools for analyzing the heaviness of the tails of a distribution are presented. One measure is kurtosis (7) but here I will focus on the tail index of a distribution. Heuristically speaking, we assume that the survival function (for the right tail) of a random variable **X** decays with a power law and we want to estimate the power. To be more precise define $\bar{F}(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x)$ and assume that for x large we have ($\alpha > 0$):

$$\bar{F}(x) \asymp Cx^{-\alpha} \tag{40}$$

for some $C > 0$ where $\asymp$ means asymptotically equivalent. Therefore α is a measure for the speed of decay of the tail and also gives you information which moments exist.

If a distribution has Pareto-like tails:

$$\lim_{x \to \infty} \bar{F}(x)x^{\alpha} = \beta > 0 \tag{41}$$

then we can look at the one sided moments, define $X_+ := \max(0, X)$ and $X_- = \min(0, X)$ ($f(x)$ is the density of X). From measure theory we know that for a positive random variable Y the following formula holds:

$$\mathbb{E}(Y) = p \int_0^{\infty} y^{p-1} \bar{F}_Y(y) dx$$

Therefore we can conclude

$$\mathbb{E}(X_+^p) = p \int_0^{\infty} x^{p-1} \bar{F}_{X_+}(x) dx \tag{42}$$

and recall that for $M > 0$

$$\int_M^{\infty} x^{p-1-\alpha} = \begin{cases} < \infty & \text{if } p - 1 - \alpha < -1 \\ \infty & \text{if } p - 1 - \alpha \geq -1 \end{cases} \tag{43}$$

Thus from (42) and (43) we can conclude that if $p \geq \alpha$ then $\mathbb{E}(X_+^p) = \infty$. If $p < \alpha$ then $\mathbb{E}(X_+^p) < \infty$.
Analog we can analyze the left tail and $\mathbb{E}(|X_-|)$ or we combine the both tails by analyzing the random variable $|X|$.

Next two procedures to estimate the (left and right) tail index are presented:

Estimation of tail index method 1:

First we observe that if the distribution of X has pareto-like tails we have for x large:

$$\log(P(X > x)) = \log(\bar{F}(x)) \approx \log(\beta x^{-\alpha}) = \log(\beta) - \alpha \log(x) \tag{44}$$

and we can estimate $P(X > x)$ from the order statistics $\mathbf{X}_{(1)} \geq ... \geq \mathbf{X}_{(T)}$ by

$$P(X > \mathbf{X}_{(t)}) \approx \frac{t}{T}$$

Therefore if we plot the points $(\log(t/T), \log(\mathbf{X}_{(t)}))_{t=1,...,K}$ they should lay approximately on a line with slope $-\alpha$. Therefore we can estimate α by the regression $\log(t/T) = \tilde{\beta} + \alpha \log(\mathbf{X}_{(t)})$. One important aspect is that one should choose a reasonable cut-off point K since we assume the power law tail behavior only for large X. One possibility is to use only the order statistics that are above a certain threshold like the 80% empirical quantile. Another idea is based on the plot and to choose the K in such a way that the points are on a line as good as possible.

Estimation of tail index method 2, Hill estimator:

Another estimator for the tail index is the so called Hill-estimator. It is based one the same idea as presented above, i.e. $\log(t/T) \approx \tilde{\beta} + \alpha \log(\mathbf{X}_{(t)})$, therefore for $k \in \{2, ..., T\}$

$$\sum_{t=1}^{K-1} \log(\frac{\mathbf{X}_{(t)}}{\mathbf{X}_{(K)}}) = \sum_{t=1}^{K-1} \log(\mathbf{X}_{(t)}) - (K-1)\log(\mathbf{X}_{(K)}) \approx$$
$$\frac{1}{\alpha}\left(-\sum_{t=1}^{K-1} \log(t/T) + (K-1)\log(K/T) \right) = \frac{1}{\alpha}\left(-\sum_{t=1}^{K-1} \log(t) + (K-1)\log(K) \right) =$$
$$\frac{1}{\alpha}\left(- \underbrace{\log((K-1)!)}_{\approx (K-1)\log(K-1)-(K-1)} + (K-1)\log(K) \right) \approx \frac{1}{\alpha}(K-1)$$

where for last approximation the Stirling formula $\log(K)! \approx \log(K^K e^{-K})$ [4] and the Taylor approximation $\log(K-2+1) \approx K-3$ was used. Then rearranging for α gives the Hill estimator:

$$\hat{\alpha} = \frac{K-1}{\sum_{t=1}^{K-1} \log(\frac{\mathbf{X}_{(t)}}{\mathbf{X}_{(K)}})} \tag{45}$$

Furthermore it can be shown that if $\mathbf{X}_t$ iid with tail index α then the Hill estimator is consistent and asymptotic normality holds, for more details see [30]:

$$\sqrt{K}(\hat{\alpha} - \alpha) \to N(0, \alpha^2)$$

Here again, choosing the cut-off point K is important and one can use the same rule of thumb as for the first estimation procedure.

4 Empirical Results

4.1 Literature review

Since the introduction of the blockchain technology and bitcoin by the white paper of Satoshi Nakamoto [48] in 2009 and the subsequent spectacular rise in the price, there has been a lot of interest by the academic community. Therefore a lot of papers analyzed

[4]Normally the Stirling formula states $K! = \sqrt{2\pi K}(K/e)^K$ but for K large (e.g. $K > 751$) $\ln(K!) \approx \ln(K^K e^{-K})$ offers sufficient accuracy.

various aspects of crypto-currencies and bitcoin in particular. One extensively explored question is if crypto-assets are speculative assets or money, in the sense that it is a store of value and medium of exchange. To summarize the literature, the answer is mostly that crypto-assets are too volatile and exhibit bubble behavior (Glaser et al (2014) [35], Baek et al (2014) [16], and Chea et al (2015) [23]), which is more characteristic of a speculative asset.The question if there is a bubble component present in the crypto-currencies was further investigated and answered positively (Hafner (2018) [36] and Corbet et al (2017) [26]). Vasek and Moore (2015) [51] report a high prevalence of scams in bitcoin and crypto-assets. Further Dweyer et al (2014) [46] pointed out that the energy consumption of bitcoin is a detriment since the trusted third parties are replaced by the proof of work concept which requires the constant use of lots of CPU power to process transactions (see section 2). Nevertheless, surprisingly, crypto-currencies especially bitcoin can offer lower bid-ask spreads and transaction costs than foreign exchange markets (Kim (2017) [42]). Brauneis et all (2018) [21] concluded that over time the bitcoin is highly liquid and over time the markets get more mature, meaning the arbitrage opportunities between different exchanges vanish. There is a growing consensus that crypto-assets are a new asset class on its own that offers diversification benefits (Baur et al (2018) [17], Corbet et al (2018) [27] and Guesmi et al (2018) [34]), since there seems to exist a crypto-asset risk factor that is uncorrelated to other financial assets. Furthermore, Cretarola et all (2017) [28] modeled the dynamics of bitcoin prices by stochastic differential equations and introduced a sentiment-factor that can influence the price also through lagged values. They also developed pricing formulas for derivatives of bitcoin in the model. Due to the increased popularity of crypto-assets for investment purposes, especially also from institutional investors, the focus has shifted to risk management. Therefore, there is a growing need to model the crypto-asset dynamics and compute risk measures like Value-at-Risk and Expected Shortfall (Borri (2018) [19]). For this task, different types of GARCH models (general auto-regressive conditional heteroscedasticity models where the volatility of the return at time t is modeled by a function of the past) seem to be well suited (Chu et al (2017) [24], Katsiampa (2017) [41] and Guesmi et al (2018) [34]). The most recent research indicates that Markov-switching GARCH models perform best (Caporale et al (2019) [22]) which is possibly due to structural breaks (Ardia et al (2018) [15] and Bouri et all (2018) [20]).
Furthermore the question if bitcoin prices or crypto-asset time series possess the same stylized facts as other financial time series is investigated by Yelowitz et all (2015) [53], Chu et al (2015) [25], Easwaran et al (2015) [33], Zhang et al (2018) [54], Hu et al (2018) [38], and Cunha et al (2019) [29]. The same question viewed in a high-frequency contest is analyzed by Platanakiset et al (2019) [49]. The answer is positive, meaning that the same stylized

facts (see section 3) can be found in crypto-asset time series.
Nevertheless, recently C. Alexander and M. Dakos (2019) [14] pointed out that most of the papers on crypto-assets used flawed data. For example, a lot of internet website offer free to download data which are artificially created and do not represent traded prices. Furthermore, they pointed out that one has to be careful if prices generated by different exchanges are compared in a multivariate statistical analysis since there is no common standard for timestamps of trades. Therefore, in this thesis, the question about stylized facts is revisited using data from traded prices on Kraken, one of the major crypto-currency exchanges.

4.2 Data description

In this section, the methods described in section 3 are applied to analyze the statistical properties of price data of various major stock markets and crypto-currencies. The main focus is on the comparison of the two and whether there are significant differences. All the data that is analyzed is publicly available. First, a short description of the data will be presented.

4.2.1 Crypto-currency data

Four of the main crypto-currencies are considered in the data analysis, the most prominent one being bitcoin. They are:

(1) Bitcoin (BTC), 68.6 % of market capitalisation of crypto-assets

(2) Ethereum (ETH), 7.76 % of market cap

(3) Ripple (XRP), 4.37 % of market cap

(4) Litecoin (LTC), 1.88 % of market cap

The whole market capitalization of crypto-assets on 2019-08-11 is 298 billion USD (see the website coinmarketcap.com for more details) and as can be seen above bitcoin is by far the most dominant one with two-thirds of the whole market cap. Bitcoin has a much higher market capitalisation (see figure 4) and traded volume than bitcoin cash and will therefore be used. In figure 4 the development of the percentage of the market capitalization of the crypto-currency space is displayed for different currencies. In figure 5 the total market cap of the crypto-currency universe is shown, there one sees the boom from 2017-2018. Together these four compromise 82 % of market cap and offer a broad overview of the crypto-assets. The price against the USD for all of them is determined by supply and demand in the market. There are many private exchanges where crypto-currencies can be traded, the price data used are from Kraken, one widely used exchange platform for crypto-currency trading. The data used are from real trades on the exchange and therefore the critique raised by [14] (see bottom of section 4.1) is not applicable here. Further the data analysis, presented in this section, was also done for data from two exchanges bitstamp, bitfinex[5] and data from the internet-site coingecko which is a coin ranking site and not an exchange. C. Alexander and M. Dakos (2019) [14] pointed out that coingecko provides possible flawed data. Nevertheless the results of this thesis stay the same regardless of the source of data. The data were downloaded from the website cryptodatadownload.com on 2019-08-01. The frequency considered is daily and the closing price of the day is used for the analysis. The time-frame considered differs for the four different crypto-assets, since some of them are newer than others with bitcoin being the oldest. The time frame of the analysis for bitcoin is from 2013-10-06 [6] to 2019-07-30, for Ethereum from 2015-08-07 to 2019-07-30, for Ripple from 2017-06-21 to 2019-07-30 and fro Litecoin from 2013-10-24 to 2019-07-30. For the analysis when possible the longest time-frame available for each currency will be used and only if a common time-grid is needed for the analysis (computing correlations,...) will we pick a different window. In figure 6 the price development is displayed. There, everyone sees the spectacular price increases of crypto-currencies in the last few years. Furthermore in table 3 the summary statistics are displayed.

	min	25% qu	median	mean	75 % qu	max
BTC	122.0	414.6	712.0	2846.0	4738.0	19360.0
ETH	0.42	10.9	126.2	206.0	298.0	1359.0
XRP	0.147	0.285	0.363	0.478	0.525	2.780
LTC	1.20	3.69	9.51	37.72	53.48	353.60

Table 3: Summary statistics of crypto-assets

[5]These data are downloaded from Quandl using API. There are two R-packages called Quandl and httr which makes using an API to download data very comfortable.

[6]Bitcoin exists since 2009 but this time frame is chosen due to data quality and availability since at the beginning trades occurred infrequently and not very often.

Figure 4: Development of percentage of market capitalisation for various crypto-currencies.

4.2.2 Stock price data

We also pick four stock markets for comparison. The frequency is daily and closing prices are used. The data was downloaded from yahoo finance on 2019-08-01. The markets are:

(1) S&P 500 (SP500) in USD

(2) NASDAQ Composite (NAS) in USD

(3) iShares MSCI Emerging Markets ETF (EEM) in USD

(4) DAX Performance-Index (DAX) in EUR

The time frames are from 1999-12-31 to 2019-07-30 for SP500, 1999-12-31 to 2019-07-30 for NAS, 2003-04-14 to 2019-07-30 for EEM and 2000-01-03 to 2019-07-30 for DAX. In figure 7 the price charts are presented and in table 4 the summary statistics are displayed.

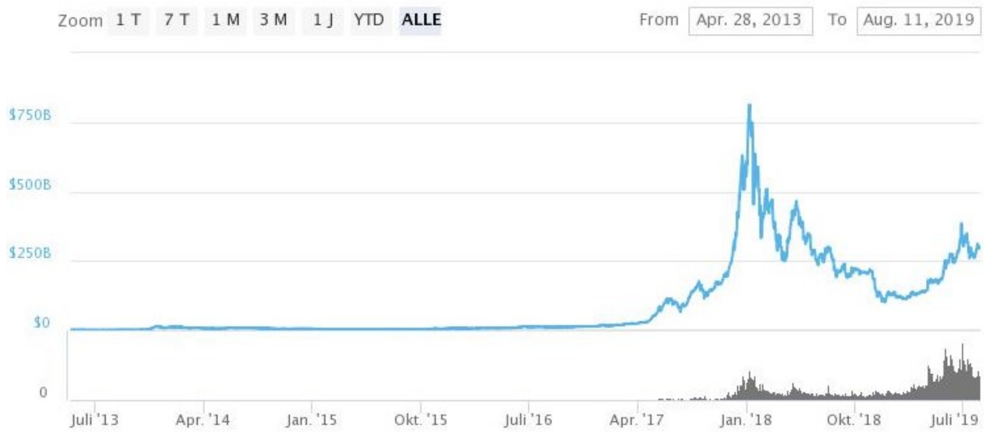

Figure 5: Development of market capitalisation of whole crypto-currency space in USD.

	min	25% qu	median	mean	75 % qu	max
SP500	676.5	1147.0	1351.0	1544.0	1940.0	3026.0
NAS	1114	2075	2607	3366	4512	8330
EEM	11.22	32.27	39.55	36.66	43.03	55.73
DAX	2203	5196	6761	7320	9581	13560

Table 4: Summary statistics of stock indices

4.3 Data analysis

In this section the log-returns of the markets, described above, are analysed, so given the price of an asset P_t we set $r_t = \ln(P_t/P_{t-1})$.

4.3.1 Exploratory data analysis

First, an exploratory data analysis is performed where the mean, quantiles, standard deviation, skewness, and kurtosis are displayed in table 5. All return series are centered around 0. Note that the crypto-currencies have much larger (in absolute terms) min and max values, especially ETH and LTC. Also the interquantile range and the standard deviation are larger across the board for the crypto-currencies than for stocks. Observe that within the four crypto-assets BTC has the lowest volatility and has less extreme values

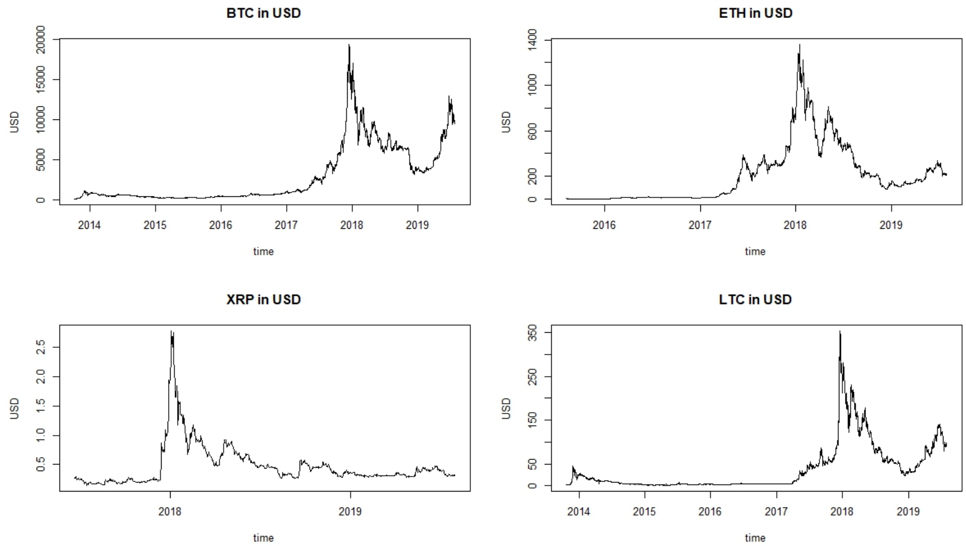

Figure 6: Historic price development of the four crypto-currencies.

than the other three, this is probably due to the maturity of BTC compared to the other ones. Further, it is interesting that XRP and LTC display a large positive skew whereas ETH has a large negative skew. The kurtosis of the crypto currencies is in general higher than for the stock markets except for BTC which has a smaller kurtosis than EEM and SP, which is surprising to the author. For the stock market, the picture is quite similar across the board with SP and DAX being quite similar in terms of volatility, also both have negative skew. Whereas NAS and EEM are a little more volatile and surprisingly EEM has a positive skewness and a quite large kurtosis. All in all, these findings fit to the stylized facts that asset returns posses heavier tails than the normal distribution. If we want to answer wheater returns are in general negative skewed the picture is less clear, they seem to be asymmetrically distributed, but it is not clear to which side they are tilted.

For all eight time series, the Jarque-Bera and Shapiro-Wilk tests strongly reject the H_0 hypothesis (p-values smaller than 10^{-5}) that the data are iid normal, therefore th p-values are not reported. This fits the stylized facts and the analysis of skew and excess kurtosis in table 5. Furthermore, normal qq-plots of the return series are presented in figure 8, if the dots are on the red line this indicates that the data come from a normal distribution. In figure 8 again the normal distribution is strongly rejected due to the tail behavior of the

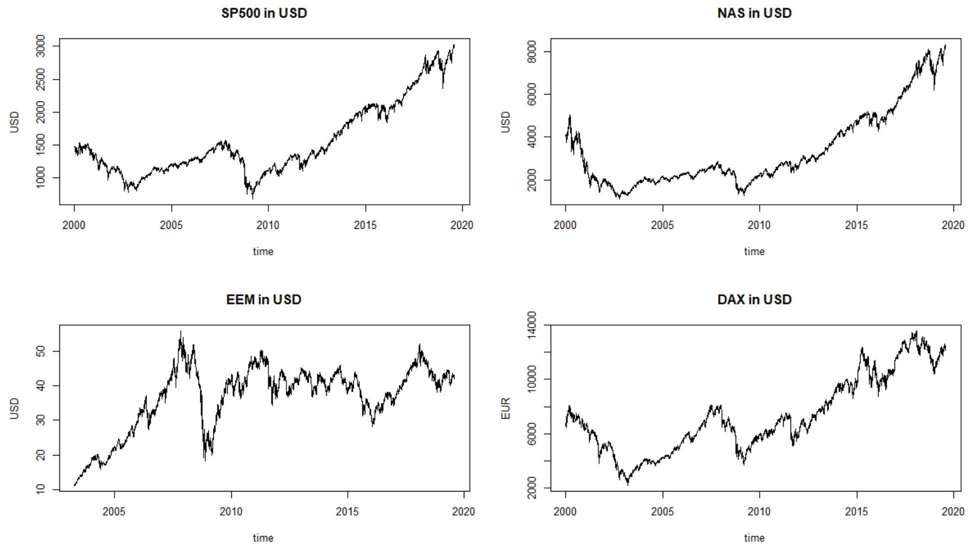

Figure 7: Historic price development of the four stock markets.

returns. In figure 9 the histograms of the returns together with a non-parametric (red) and parametric (blue) estimation of the density are displayed. The parametric density estimation uses the normal density and just inserts the estimated mean and standard deviation. Also here it is visible that the returns are not normal distributed.

In figure 10 the returns of the eight assets are displayed. In figure 10 one sees the higher volatility of crypto-assets. Furthermore, the volatility clustering effect can also be seen in all eight return series since the extreme moves tend to occur in clusters. This is especially visible in the stock market returns where periods with high volatility (large absolute returns) and low volatility can be seen. Similar patterns occur also for the crypto-currency return series, but not as distinct and clear as for stocks.

Next, we take a look at the autocorrelation function of the returns in figure 11. In figure 11 it is visible that no autocorrelation can be found in all return time series since all acf values are by and large inside the blue confidence band. Note the exception of LTC where there seems to be a non-spurious, negative linear relationship between this days return and the next one (lag 1). But all in all, the picture is quite clear that there is no linear relationship between returns and lagged returns.

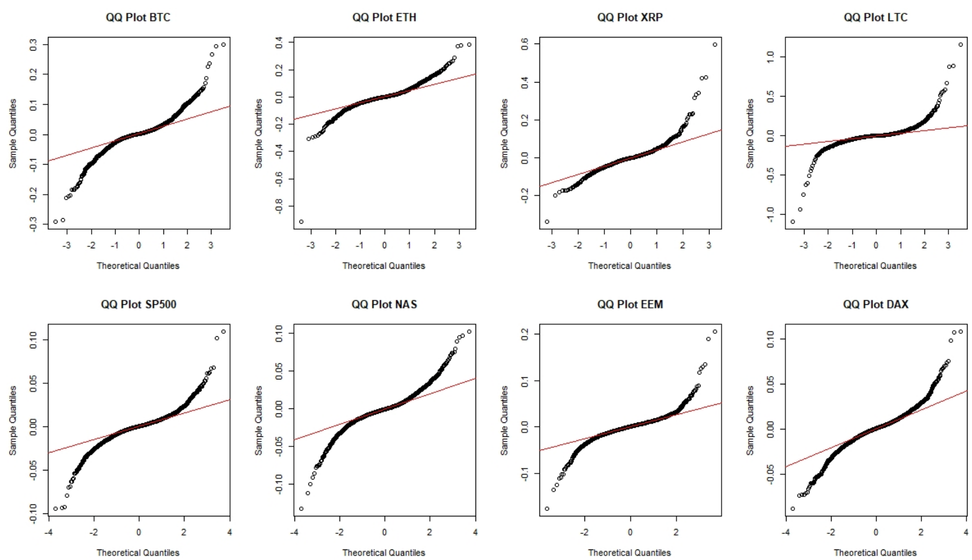

Figure 8: QQ-plots

Next the autocorrelation function (acf) for the squared returns r_t^2 are analyzed in figure 12. There, an interesting pattern emerges. For the stock markets, the typical behavior of slow decay of correlation is exhibited in figure 12 in all stock markets under consideration, this corresponds to the volatility clustering effect. Interestingly, this pattern is much less pronounced for the crypto-currencies, for BTC it is present, but the level of autocorrelation is less than for the stock markets, for ETH there seems to be no correlation of the squared returns after all. For XRP the pattern is present, but seems to follow a somewhat periodic patter and for LTC there is some correlation for the first few lags and around lag 10. This indicates that the volatility clustering effect is not so strong for crypto-currencies. To analyze further we take a look at the acf of the absolute log returns (i.e. $|r_t|$) in figure 13. In figure 13 the same picture is present, the stock markets possess higher acf that is decaying slower than the crypto-assets. Nevertheless figure (13) gives a clearer picture, there is some autocorrelation in the crypto-currencies (also in ETH and LTC where nearly none could be detected using squared returns).

In table 6 the p-values [7] of the Box-Pierce and Ljung-Box test, for testing the H_0 hy-

[7] They are rounded to three digits.

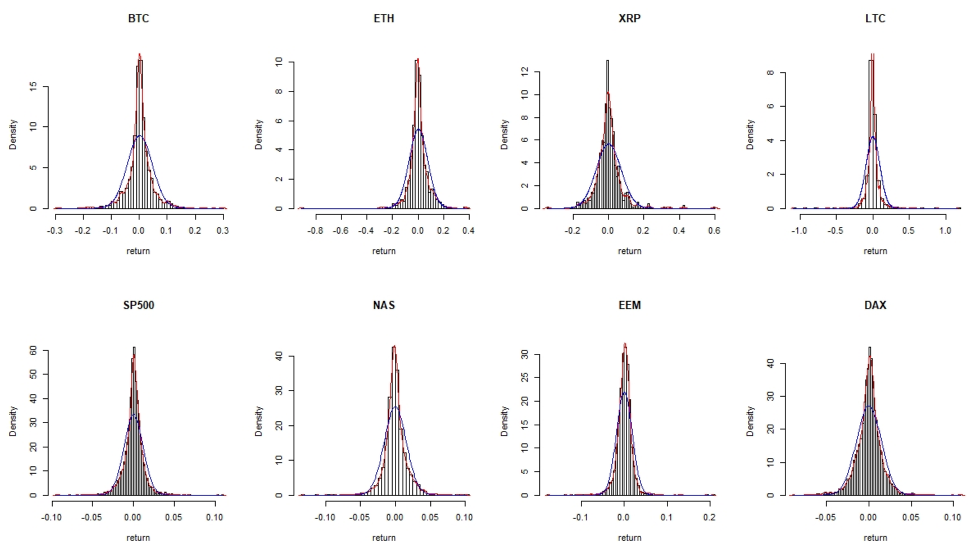

Figure 9: Histograms of asset returns overlaid with non-parametric (red line) and parametric normal (blue line) density estimation

pothesis that returns are iid, (see section3.1) are displayed for different max lag values, i.e. $k = 3, 5, 10, 15, 30$. In table 6 the iid hypothesis is clearly rejected (i.e. p-value below 5% significance level) except for XRP for all but $k = 30$, also DAX has for $k = 3$ high p-values. The behavior of XRP can be explained by the low sample size compared to the other time series. At first glance these results contradict figure 11 where no significant autocorrelation is visible, but interpreting more carefully there are two factors at play. First, the sample size is quite high for all series except XRP and in this situations test tend to find significance quite often [8] and second if one is precise the H_0 states that the returns are iid (which implies 0 acf but not the other way round) and we see in figure 12 and 13 that there are nonlinear dependency's, therefore, they are not iid. The sample sizes are:

- $T = 2111$ for BTC
- $T = 1443$ for ETH

[8] When doing the same tests for sample size $T = 200$ the picture changes and H_0 iid/ no autocorrelation gets not rejected very often (except for LTC where the acf for lag 1 is significantly different from 0 see figure 11). Nevertheless this could be an interesting observation for further analysis but this is beyond the scope of this thesis.

	min	25% qu	median	mean	75 % qu	max	sd	skew	excess kurtosis
BTC	-0.292	-0.014	0.002	0.002	0.019	0.299	0.045	-0.057	6.815
ETH	-0.916	-0.028	0.000	0.003	0.033	0.383	0.074	-1.142	19.277
XRP	-0.341	-0.033	-0.003	0.000	0.025	0.597	0.070	1.729	12.330
LTC	-1.099	-0.025	0.000	0.002	0.023	1.160	0.094	0.678	39.688
SP500	-0.095	-0.005	0.001	0.000	0.006	0.110	0.012	-0.220	8.601
NAS	-0.133	-0.007	-0.001	0.000	0.006	0.102	0.016	0.006	5.909
EEM	-0.176	-0.008	0.001	0.000	0.009	0.205	0.018	0.166	15.468
DAX	-0.089	-0.007	0.001	0.000	0.007	0.108	0.015	-0.058	4.619

Table 5: Summary of returns

- $T = 759$ for XRP
- $T = 2095$ for LTC
- $T = 4925$ for SP500 and NAS
- $T = 4102$ for EEM
- $T = 4968$ for DAX

Furthermore note that in the light of table 6 (i.e. iid and no correlation gets rejected for log returns) we save space by not presenting the Box tests for the squared and absolute returns since there iid-ness gets rejected even more clearly.

All in all figure 11, 12, and 13 tell the story that there is no linear dependence between returns, but that there is nonlinear dependence, especially the volatility clustering effect. This fits the stylized facts [39], [45] but it seems that the nonlinear dependence and volatility clustering is weaker for crypto-currencies than for stock prices, which is surprising but fits the observations for figure 10.

4.3.2 Unit root test

Here we test if the prices , log prices and returns are stationary (i.e. possess no unit root) by using the Dickey Fuller, Augmented Dickey Fuller and KPSS tests described in section 3.3. To keep the analysis simple [9] the ADF is used in the most general form (type

[9] And since that is the way how it is implemented in R.

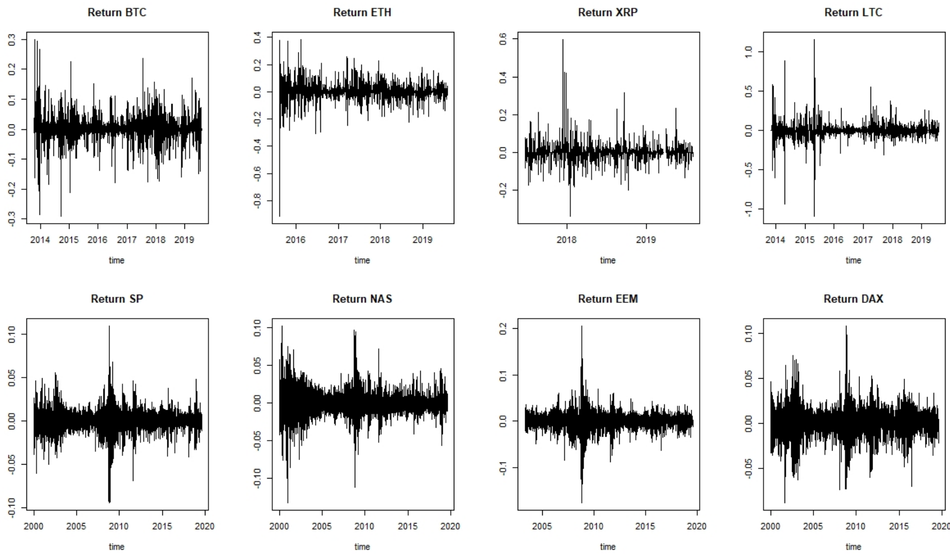

Figure 10: Returns

C) of the tests, since they are nested (i.e. the type C ADF incorporates the type B ADF if one tests with the t-test-statistic) and the sample sizes are large enough that we do not have troubles by including a few more parameters that need to be estimated. Further, the number of lags included in the model. The KPSS will be used for level stationary (type A) and trend stationary (type B).

In table 7 the p-values of the DF, ADF (H_0 the process is non-stationary) and KPSS (H_0 the process is stationary) tests are displayed. Since the critical values for the KPSS test are calculated by R using a table of stored values for just a certain range, for the p-values in the range $0.01 - 0.1$ the values are accurate (interpolation is used) whereas otherwise it will just be reported if the values are below 0.01 (i.e. < 0.01 the H_0 gets rejected for the significance level 0.01) or above 0.1 (i.e. > 0.1 the H_0 is not rejected for significance level 0.1). For the DF and ADF, the p-values are reported in the range $0.01 - 1$ or < 0.01. Further the ADF will be used for number of lags $p = 3, 6, \lfloor (T_1)^{1/3} \rfloor$. The hypothesis is that levels and log levels are non-stationary and returns are stationary.

In table 7 it is visible that levels and log levels are non-stationary for crypto-assets and stocks. EEM-log-levels seem to be stationary according to the DF and ADF test when using a 0.1 confidence level but are also non-stationary when using 0.01 as testing level.

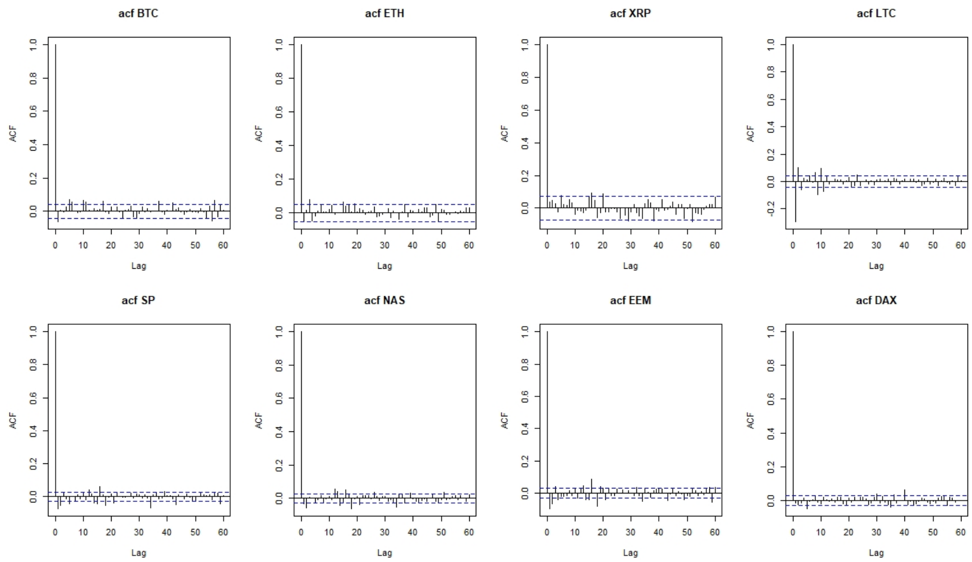

Figure 11: Autocorrelation (4) of returns r_t

Also, NAS and EEM-returns are non-stationary according to KPSS A, BTC and ETH are non-stationary according to KPSS B when using 0.1 as testing level. But when a 0.01 testing level is used, all tests agree that all levels and log levels are non-stationary and all returns are stationary. This is what we expected and fits perfectly to the stylized facts.

$k = 3$	BTC	ETH	XRP	LTC	SP500	NAS	EEM	DAX
Box-Pierce	0.033	0.003	0.315	0	0	0	0	0.165
Ljung-Box	0.033	0.003	0.313	0	0	0	0	0.165
$k = 5$ B-P	0.001	0.003	0.133	0	0	0	0	0.002
L-B	0.001	0.003	0.130	0	0	0	0	0.002
$k = 10$ B-P	0	0.019	0.235	0	0	0.001	0	0.006
L-B	0	0.019	0.227	0	0	0.001	0	0.006
$k = 15$ B-P	0	0.011	0.290	0	0	0	0	0.023
L-B	0	0.011	0.277	0	0	0	0	0.023
$k = 30$ B-P	0	0.039	0.043	0	0	0	0	0.003
L-B	0	0.036	0.034	0	0	0	0	0.003

Table 6: p-values of Box-Pirece and Ljung-Box test for lags $k = 3, 5, 10, 15, 30$

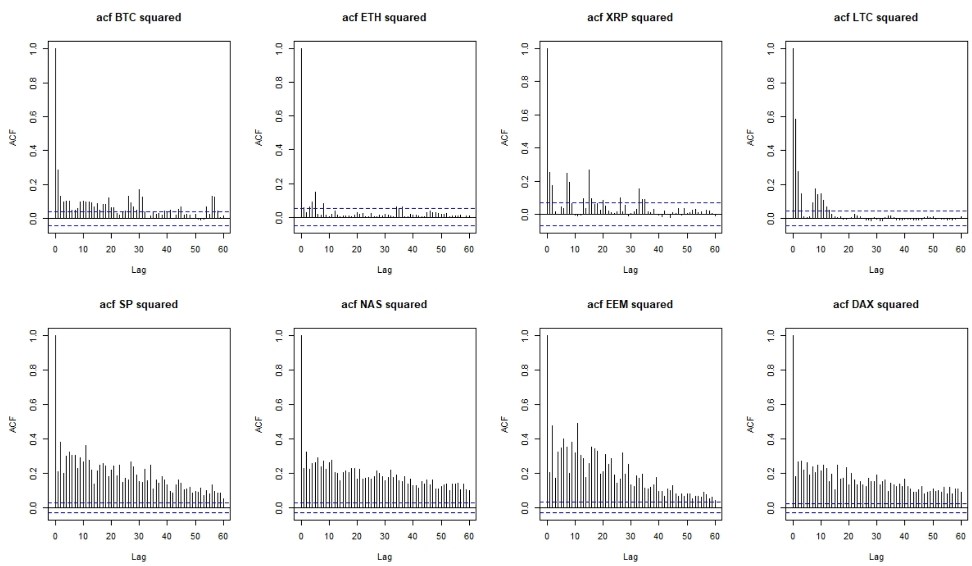

Figure 12: Autocorrelation (4) of squared returns r_t^2

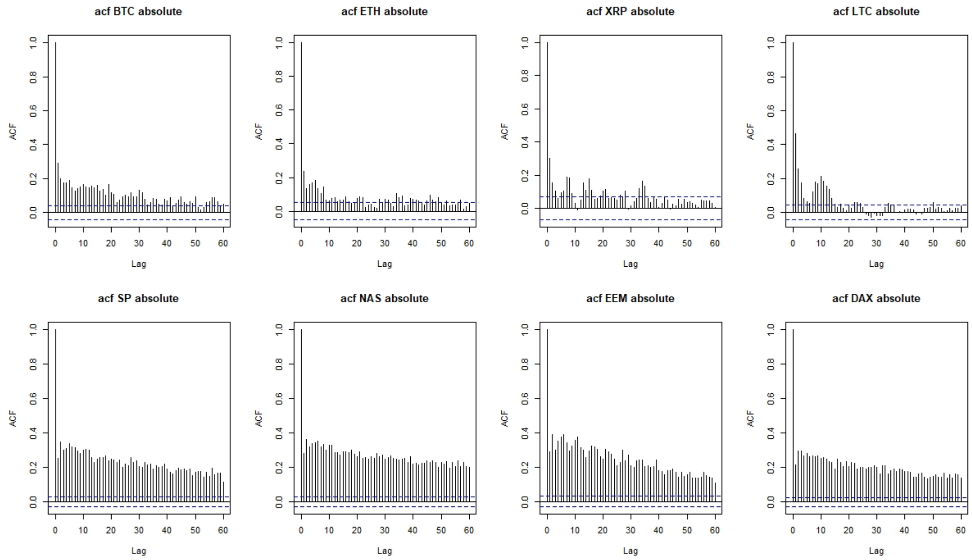

Figure 13: Autocorrelation (4) of absolute returns $|r_t|$

	DF	ADF $p = 3$	ADF $p = 6$	ADF $p = \lfloor (T_1)^{1/3} \rfloor$	KPSS A	KPSS B
BTC-level	0.4897	0.4421	0.4335	0.2365	< 0.01	< 0.01
BTC-log level	0.8028	0.8411	0.7436	0.6787	< 0.01	< 0.01
BTC-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	0.07625
ETH-level	0.7387	0.7073	0.5138	0.7084	< 0.01	< 0.01
ETH-log level	0.9685	0.9621	0.972	0.9519	< 0.01	< 0.01
ETH-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	0.03703
XRP-level	0.4213	0.3033	0.08506	0.08454	< 0.01	< 0.01
XRP-log level	0.6788	0.5787	0.5228	0.4183	< 0.01	< 0.01
XRP-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	> 0.1
LTC-level	0.3273	0.2019	0.0433	0.2304	< 0.01	< 0.01
LTC-log level	0.5842	0.808	0.7646	0.727	< 0.01	< 0.01
LTC-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	> 0.1
SP500-level	0.6876	0.7924	0.8072	0.835	< 0.01	< 0.01
SP500-log level	0.3827	0.504	0.5054	0.5179	< 0.01	< 0.01
SP500-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	> 0.1
NAS-level	0.9285	0.9499	0.9606	0.9542	< 0.01	< 0.01
NAS-log level	0.9778	0.99	0.998	0.9832	< 0.01	< 0.01
NAS-returns	< 0.01	< 0.01	< 0.01	< 0.01	0.02839	> 0.1
EEM-level	0.1248	0.2595	0.3114	0.2738	< 0.01	< 0.01
EEM-log level	0.05092	0.08547	0.08881	0.09643	< 0.01	< 0.01
EEM-returns	< 0.01	< 0.01	< 0.01	< 0.01	0.09963	> 0.1
DAX-level	0.3129	0.2926	0.3081	0.2636	< 0.01	< 0.01
DAX-log level	0.269	0.277	0.2824	0.2137	< 0.01	< 0.01
DAX-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	> 0.1

Table 7: p-values of unit root tests

4.3.3 Tail index estimation

Now we will estimate the tail index for the left, right and absolute tail (i.e. using the estimator on $|r_t|$ and thus collapsing the left and right tail to one single tail) of the return distributions of the different assets. The estimation is performed by applying method 1 (44) and the hill estimator (45). The threshold K for the left tail is set such that 20% of the data are below the threshold (i.e. the index of the empirical 20% quantile is used for K and for the right and absolute tail equivalently the 80% quantile is used). Therefore, all estimators use 20% of the data to estimate the tail index α. Furthermore, the hill estimator that is implemented in R in the function hillplot() is also included in the analysis for the absolute tails. This estimator agrees with my implementation of the hill estimator but chooses the threshold differently by an asymptotic rule and therefore displays different values.

Furthermore, for comparison also a random draw of a standard normal and a two-sided Pareto distribution [10] with tail index 1.5 and location parameter 1 with sample size 4000 is included in the comparison to get a feeling for the accuracy of the estimator (i.e. comparison with Pareto-distribution) and what happens if there are no Pareto-like tails present (i.e. comparison with normal distribution).

In table 8 the results are displayed. Method 1 will be indicated by a 1, e.g. left 1 is the estimate of the left tail index using method 1, and left hill denotes the estimate by the Hill estimator. Further, the R implementation of the hill estimator will be indicated by abs hill R.

In table 8 one sees that the left and right tail of the distribution are roughly on the same level. Further, the Hill estimate using 20% of the data gives lower estimates than method 1 using the same data. Also by collapsing both tails to one the estimated indexes get larger than the individual ones. We also see in the last row that all estimator work well for estimating the tail index 1.5 of the Pareto distribution and for the normal distribution [11] (second last row) higher tail indexes are estimated.

All in all the tail indexes of the stock returns are a bit higher than the ones of the crypto

[10] The density is given by $f(x) = 1/(4|x|^{3/2})\mathbf{1}_{\{|x|\geq 1\}}$

[11] The normal distribution tails decay with exponential speed and all moments exist, therefore the tail index is large or ∞

returns but the differences are not extreme. Furthermore, table 8 conveys the message that for return series at most the second moment exists but not even the third (skewness) or fourth (kurtosis).

	left 1	left hill	right 1	right hill	abs 1	abs hill	abs hill R
BTC	1.88	1.44	1.62	1.16	2.46	1.93	2.44
ETH	1.95	1.38	1.78	1.41	2.52	1.99	2.77
XRP	1.59	1.23	2.17	1.74	2.24	1.93	2.56
LTC	1.44	1.18	1.63	1.24	1.82	1.60	1.85
SP500	2.08	1.62	1.90	1.38	2.54	2.14	2.56
NAS	1.94	1.38	2.03	1.60	2.65	2.11	2.7
EEM	2.14	1.85	1.96	1.54	2.36	2.16	2.38
DAX	2.26	1.69	2.11	1.53	2.79	2.34	2.86
$N(0,1)$	2.99	2.14	3.01	2.15	4.23	3.32	4.35
Pareto$(1,1.5)$	1.52	1.52	1.46	1.42	1.47	1.48	1.49

Table 8: Tail index estimation

4.3.4 Correlation analysis

In this section, the correlation between the eight different assets is estimated. This is of interest due to the diversification effect, i.e. including uncorrelated assets in a portfolio can reduce risk (i.e. volatility) without diminishing the expected return. Since crypto-assets are new, we try to answer the question if they move together with the stock markets or not.

Here we will just make a crude estimation of the correlation by the usual formula [12]. Here we have just one complication since not all eight time-series are on the same time grid [13] and for some, there are missing values. Further, the time-frame is different for the different assets, e.g.. some crypto-currencies are newer than others. Therefore when

[12]Let $\mathbf{X}_t$ and $\mathbf{Y}_t$ be two uni-variate samples on the same grid with sample size T. Then the empirical covariance is given by:

$$\hat{\text{cov}}(\mathbf{X},\mathbf{Y}) = \frac{1}{T}\sum_t (\mathbf{X}_t - \bar{\mathbf{X}}_T)(\mathbf{Y}_t - \bar{\mathbf{Y}}_T)$$

and the estimated correlation is then given by $\frac{\hat{\text{cov}}(\mathbf{X},\mathbf{Y})}{\sqrt{\hat{\text{cov}}(\mathbf{X},\mathbf{X})\hat{\text{cov}}(\mathbf{Y},\mathbf{Y})}}$

[13]For example, BTC is also traded on the weekend whereas stocks are not

computing the correlation between two assets the time grid is chosen such that for each time there are values for both assets and such that the grid spans the longest possible time-span. In table 9 the correlation matrix for the eight assets is presented. The picture is clear, the crypto-currencies are correlated, i.e. they have correlation around 0.5, but they are uncorrelated with stock market returns.

	BTC	ETH	XRP	LTC	SP500	NAS	EEM	DAX
BTC	1	0.42	0.52	0.39	0.01	-0.05	0.00	0.00
ETH		1	0.67	0.37	0.03	-0.04	0.04	0.04
XRP			1	0.56	0.07	-0.07	0.15	0.05
LTC				1	0.01	-0.01	0.00	-0.01
SP500					1	0.89	0.85	0.60
NAS						1	0.81	0.55
EEM							1	0.58
DAX								1

Table 9: Correlation matrix

4.4 Robustness with respect to data source

C. Alexander and M. Dakos (2019) [14] pointed out that most of the academic literature so far was carried out using problematic data in the sense that a lot of the freely available data are not based on real traded prices or that when combining data of crypto-assets from different sources it is not clear if the timestamps are comparable. For further details of the potential problems with the crypto-currency data see [14].
In this section, we demonstrate the robustness of the findings above with respect to different data sources. The analysis will only be presented exemplary for bitcoin prices since the picture is the same across the board and reporting the same tables and pictures that are presented above, for different data sources do not add value. This is done by using two additional data sources, then repeating the same analysis and see if the results differ substantially. The first source is bitstamp, a major exchange platform for crypto-currencies and the second is the coin ranking website coingecko. Especially coingecko is criticized in [14] for the data quality since it is itself not an exchange platform and therefore their prices are not based on real trades. In figure 14 we report the percentage deviation of the bitcoin prices of bitstamp and coingecko with respect to the time series from Kraken in the time-frame 2014-04-15 to 2019-07-30. In figure 14 we also find large deviations of the prices

from the three different sources like C. Alexander and M. Dakos (2019) [14] pointed out. Nevertheless, the analysis of the stylized facts, for the three different time series, gives the

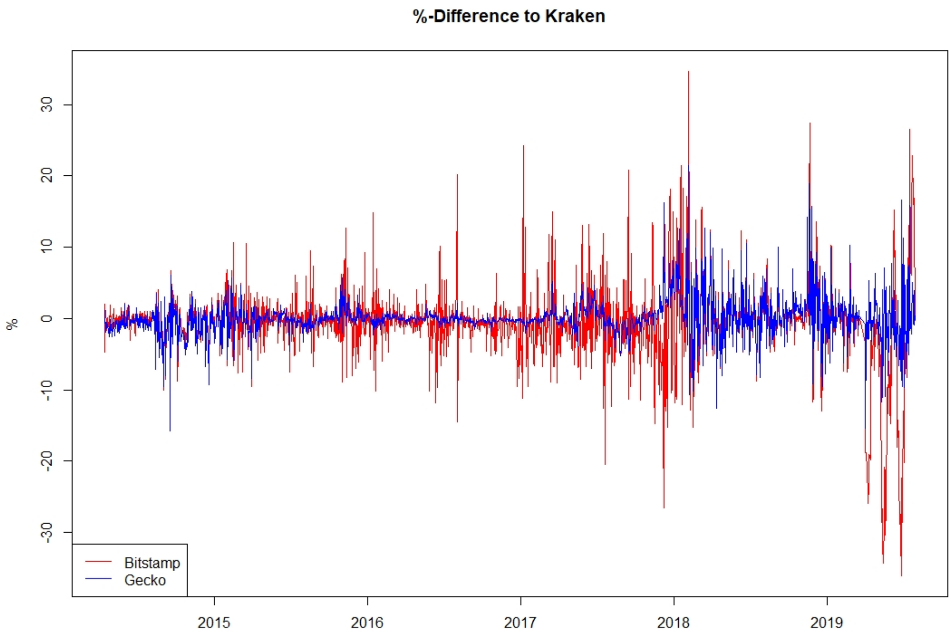

Figure 14: Deviation of bitcoin prices from bitstamp and coingecko to prices from Kraken

same answer. For all, the conclusions of the statistical tests are from a qualitative point of view the same and the differences are minor. This will be demonstrated by revisiting table 7 and presenting the p-values of the tests for stationarity (DF and KPSS tests) for the different data sources. In table 10 the p-values for the three different sources of bitcoin prices are presented and they correspond to the first three rows of table 7. In table 10 one sees that the conclusion, that levels and log-levels are non-stationary whereas the returns are, are robust with respect to the three different data sources. The different data sources, therefore, have only a minor influence on the p-values and this picture is seen for all tests that are done in section 4, but we will not report them here. Further exemplary evidence that the statistical analysis carried out in this thesis is robust with respect to the data source can be seen in figure 15 where the acf of the returns and squared returns from the three different data sources are compared. The qualitative behavior is seen in figure 15 stays the same as in figure 13.

Kraken	DF	ADF $p=3$	ADF $p=6$	ADF $p=\lfloor (T_1)^{1/3} \rfloor$	KPSS A	KPSS B
BTC-level	0.4897	0.4421	0.4335	0.2365	< 0.01	< 0.01
BTC-log level	0.8028	0.8411	0.7436	0.6787	< 0.01	< 0.01
BTC-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	0.07625
bitstamp						
BTC-level	0.4915	0.4408	0.4306	0.251	< 0.01	< 0.01
BTC-log level	0.6217	0.6263	0.5937	0.6427	< 0.01	< 0.01
BTC-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	0.03454
coingecko						
BTC-level	0.4068	0.4071	0.4556	0.1732	< 0.01	< 0.01
BTC-log level	0.5999	0.62791	0.5973	0.6316	< 0.01	< 0.01
BTC-returns	< 0.01	< 0.01	< 0.01	< 0.01	> 0.1	0.03017

Table 10: p-values of unit root tests in bitcoin prices for the three different data sources (Kraken, bitstamp, and coingecko)

4.5 Summary of data analysis

In this section, the returns of crypto-assets have been compared to returns from major stock market indices by statistical analysis. The question if crypto-assets behave differently than normal asset returns and if they possess the same stylized facts have been studied and the answer is positive. Crypto-assets conform to the same statistical properties as stocks albeit there are slight variations. The crypto-currency prices are non-stationary and the returns are stationary (so they are integrated of order one) like for stock prices. The crypto returns are non-normal distributed and possess heavier tails and much larger standard deviations than stock returns. Furthermore in terms of linear relationship over time both behave similarly and display no autocorrelation. For both the volatility clustering effect is present but it is of interest that this effect is less pronounced for crypto-assets. Further, crypto-assets are uncorrelated to the major stock markets. Furthermore the robustness of the findings with respect to different data sources is demonstrated by repeating the same statistical analysis for three different data sources.

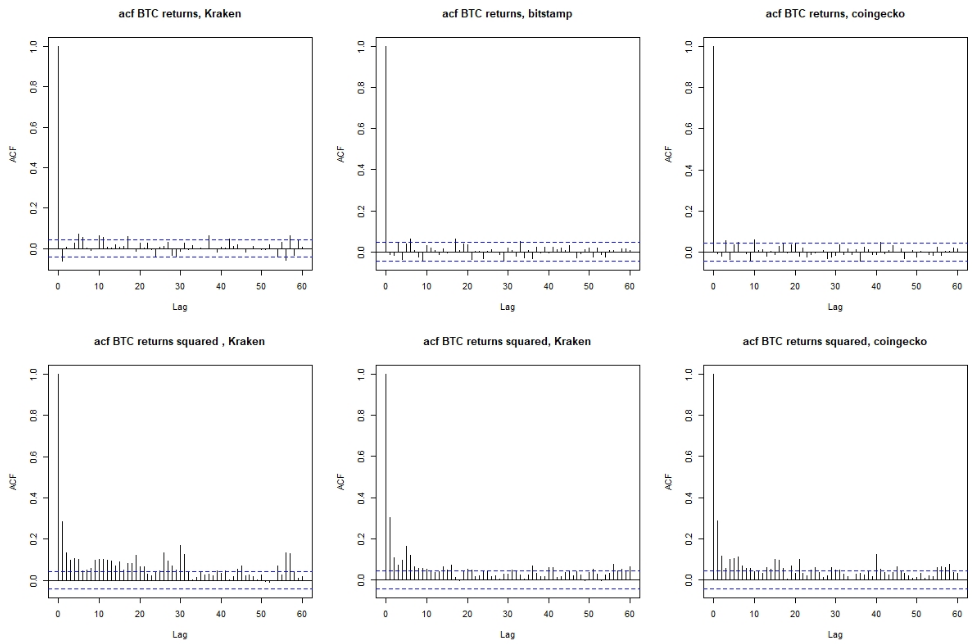

Figure 15: autocorrelation of returns and squared returns for different data sources (Kraken, bitstamp, and coingecko)

5 Conclusion

In this thesis, the blockchain technology and the most prominent application, i.e. bitcoin is described in detail in section 2. The blockchain is, in essence, a distributed ledger of the transaction history that is open to the public. There is no need for a trusted third party (i.e. banks and central banks) that clears the transaction. The integrity of the transaction history is protected by cryptography, i.e. proof of work, instead of a central clearing party. Therefore no single party has the power to make changes, like reversing transactions, increasing the money supply, denying access and so on ..., to the system. This can be seen either as an advantage (i.e. your money will not be diluted by a central bank that decides to print money, e.g. quantitative easing) or a disadvantage (e.g. you accidentally type in the wrong amount or wrong address). Nevertheless, since its creation the blockchain technology and especially bitcoin gathered a lot of attention that resulted in a spectacular rise of the price of one bitcoin (BTC) against the USD from 0 to almost 20000 USD that was in large part driven by excitement and speculation. This resulted in extremely volatile prices which discourage its use as a stable and reliable payment system. The supply of bitcoin is organized by the bitcoin protocol, an open-source code, which caps the total number of BTC at 21 million units. Nevertheless, there are also concerns or disadvantages associated with the technology. They include the electricity consumption by the technology, also at the moment, the system is not capable of handling the large volumes of transaction that come by large scale adoption of BTC as a means for everyday payments and transactions. Furthermore, there are concerns, due to no or light regulation, that a lot of illegal transactions are facilitated through the bitcoin network.

Nevertheless, the blockchain technology underpinning bitcoin is an exciting innovation where no one knows where it will be deployed in the future. There has been an explosion of crypto-currencies, often with fraudulent intents, that use the technology and only time will tell which one is the best and survives. From an investor's point of view, it is an exciting new asset class. Therefore four of the most dominant crypto-currencies have been analyzed and compared to stock market assets in this thesis.

Stock prices, as an asset class, are known to display certain statistical properties that are quite universal. These are called stylized facts and include that prices are non-stationary but the returns are stationary, that returns are not normal distributed especially have heavier tails and asymmetry compared to the normal distribution. Furthermore, there is the so-called volatility clustering effect, which states that volatility is not constant over time, but instead if volatility is high than it is more likely that the volatility on the next day is also high. Therefore volatile periods and relatively calm periods can be observed in asset returns. This feature can, for example, be captured by the dependency of squared or absolute returns over time. It was widely observed that returns possess no linear relationship over time (i.e. no autocorrelation) but the squared returns display a quite strong correlation due to the volatility clustering effect.

Therefore the thesis aimed to compare crypto-assets to stock markets and see if the same stylized facts can be observed for crypto returns. Therefore the theoretical foundations and statistical tools needed for answering this question were explored in this thesis in section 3. With this tools at hand, the question was answered by comparing four of the main crypto-currencies (i.e. bitcoin, ethereum, ripple, and litecoin) to four major stock markets (i.e. S&P500, NASDAQ, MSCI Emerging Markets Index, and DAX) in section 4. The answer is positive and in general crypto-assets conform to the same stylized facts as stock markets albeit with slight variations. It is no surprise that crypto-currencies are more volatile than major and liquid stock markets. Nevertheless, all the stylized facts seem to hold also for crypto-assets albeit the volatility clustering effect is less distinct than for the stock market. This was a surprising finding of this thesis and warrants further research. The prices of crypto-assets are non-stationary, but the returns are. Furthermore, there is no linear relationship (i.e. autocorrelation) in the return series. Crypto returns are not normally distributed and possess heavier tails than the normal distribution and even heavier tails than stock market returns, but this is not so surprising due to the extreme price swings that can be observed. The returns seem to be distributed asymmetrically, but there is no clear picture in which direction they are skewed. Last but not least, volatility is much higher than in the stock markets that were considered in this thesis. In section 4.1 a short overview of the existing literature is presented, where recently in 2019 C. Alexander and M. Dakos [14] raised the concern that most previous papers used flawed data that originate not from real trades. In this thesis data from real traded prices where used and the findings of this thesis conform to the literature. Furthermore the analysis of the stylized facts in this thesis were repeated using also data from bitstamp and coingecko, the latter being a source that [14] described as potentially flawed, but the findings were robust with respect to the source of data origin.

All in all, one can say that the crypto-currencies are a new and exciting assets class underpinned by an innovative technology that has a lot of potential. From a statistical viewpoint the returns they offer do not behave so different than stock prices although they are much more volatile. Furthermore, the crypto returns are not strongly correlated with stock markets and therefore it can be advantageous, from an investor's point of view, to have exposure to them. Nevertheless, only time will tell if they are meant for the long term or if they will disappear again. Therefore strict risk management is needed when investing due to the extreme volatility and heavy tails of crypto returns.

References

[1] Bitcoin cash. https://de.wikipedia.org/wiki/Bitcoin_Cash. Accessed: 2019-08-14.

[2] bitcoin wiki. https://en.bitcoin.it/wiki/Clients. Accessed: 2017-09-30.

[3] bitcoin wiki. https://en.bitcoin.it/wiki/Hashcash. Accessed: 2017-09-30.

[4] bitcoin wiki. https://en.bitcoin.it/wiki/Protocol_documentation#Transaction_Verification. Accessed: 2017-09-30.

[5] bitcoin wiki. https://en.bitcoin.it/wiki/Difficulty. Accessed: 2017-09-30.

[6] bitcoin wiki. https://en.bitcoin.it/wiki/Block_chain. Accessed: 2017-09-30.

[7] bitcoin wiki. https://en.bitcoin.it/wiki/Block_chain_browser. Accessed: 2017-09-30.

[8] bitcoin wiki. https://en.bitcoin.it/wiki/Transaction. Accessed: 2017-09-30.

[9] bitcoin wiki. https://en.bitcoin.it/wiki/Confirmation. Accessed: 2017-09-30.

[10] bitcoin wiki. https://en.bitcoin.it/wiki/Proof_of_work. Accessed: 2017-09-30.

[11] bitcoin wiki. https://en.bitcoin.it/wiki/Target. Accessed: 2017-09-30.

[12] Sha256. http://www.quadibloc.com/crypto/mi060501.htm. Accessed: 2017-09-30.

[13] Wikipedia. https://de.wikipedia.org/wiki/SHA-2. Accessed: 2017-09-30.

[14] C. Alexander and M. Dakos. A critical investigation of cryptocurrency data and analysis. *Quantitative Finance*, 0(0):1–16, 2019.

[15] D. Ardia, K. Bluteau, and M. Rüede. Regime changes in bitcoin garch volatility dynamics. *Finance Research Letters*, 29, 08 2018.

[16] Baek and Elbeck. Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters*, 22, 01 2015.

[17] D. Baur, T. Dimpfl, and K. Kuck. Bitcoin, gold and the us dollar – a replication and extension. *Finance Research Letters*, 25, 10 2017.

[18] P. Billingsley. *Convergence of probability measures.* John Wiley and Sons, Inc, 1968.

[19] N. Borri. Conditional tail-risk in cryptocurrency markets. *SSRN Electronic Journal*, 01 2018.

[20] E. Bouri, L. Gil-Alana, R. Gupta, and D. Roubaud. Modelling long memory volatility in the bitcoin market: Evidence of persistence and structural breaks. *International Journal of Finance & Economics*, 10 2018.

[21] A. Brauneis, R. Mestel, R. Riordan, and E. Theissen. A high-frequency analysis of bitcoin markets. *SSRN Electronic Journal*, 01 2018.

[22] G. M. Caporale and T. Zekokh. Modelling volatility of cryptocurrencies using markov-switching garch models. *Research in International Business and Finance*, 48, 12 2018.

[23] J. E.-T. Cheah and J. Fry. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics Letters*, 130, 02 2015.

[24] J. Chu, S. Chan, S. Nadarajah, and J. Osterrieder. Garch modelling of cryptocurrencies. *Journal of Risk and Financial Management*, 10:17, 10 2017.

[25] J. Chu, S. Nadarajah, and S. Chan. Statistical analysis of the exchange rate of bitcoin. *PloS one*, 10:e0133678, 07 2015.

[26] S. Corbet, B. Lucey, and L. Yarovya. Datestamping the bitcoin and ethereum bubbles. *Finance Research Letters*, 26, 12 2017.

[27] S. Corbet, A. Meegan, C. Larkin, B. Lucey, and L. Yarovaya. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165, 04 2018.

[28] A. Cretarola, G. Figà-Talamanca, and M. Patacca. A sentiment-based model for the bitcoin: Theory, estimation and option pricing. *SSRN Electronic Journal*, 09 2017.

[29] C. Da Cunha and R. Silva. Relevant stylized facts about bitcoin: Fluctuations, first return probability, and natural phenomena. 05 2019.

[30] L. de Haan and S. Resnick. On asymptotic normality of the hill estimator. *Communications in Statistics. Stochastic Models*, 14(4):849–866, 1998.

[31] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 1979.

[32] M. Donsker. *An invariance principle for certain probability limit theorems.* Memoirs of the American Mathematical Society, 1951.

[33] S. Easwaran, M. Dixit, and S. Sinha. *Bitcoin Dynamics: The Inverse Square Law of Price Fluctuations and Other Stylized Facts*, pages 121–128. 01 2015.

[34] Z. Ftiti, k. gusemi, S. Saadi, and I. Abid. Portfolio diversification with virtual currency: Evidence from bitcoin. *International Review of Financial Analysis*, 63, 03 2018.

[35] F. Glaser, K. Zimmermann, M. Haferkorn, M. Weber, and M. Siering. Bitcoin - asset or currency? revealing users' hidden intentions. *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*, 01 2014.

[36] C. Hafner. Testing for bubbles in cryptocurrencies with time-varying volatility. *SSRN Electronic Journal*, 01 2018.

[37] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[38] A. Hu, C. Parlour, and U. Rajan. Cryptocurrencies: Stylized facts on a new investible instrument. *SSRN Electronic Journal*, 01 2018.

[39] M. K. Jilla, S. C. Nayak, and M. A. Bathula. Stylized facts of financial time series: A comprehensive analysis. pages 302–307, 2018.

[40] G. Judge, W. Griffiths, H. Lütkepohl, and T.-c. Lee. *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons Inc, 1988.

[41] P. Katsiampa. Volatility estimation for bitcoin: A comparison of garch models. *Economics Letters*, 158:3–6, 06 2017.

[42] S. Kim. On the transaction cost of bitcoin. *Finance Research Letters*, 23, 08 2017.

[43] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1):159 – 178, 1992.

[44] M. Loève. *Probability theory. Vol. II, 4th ed.* Springer Verlag, 1978.

[45] H. Malmsten and T. Terasvirta. Stylized facts of financial time series and three popular models of volatility. 3, 09 2004.

[46] D. Malone and K. O'Dwyer. Bitcoin mining and its energy footprint. pages 280–285, 01 2014.

[47] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2015.

[48] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Cryptography Mailing list at https://metzdowd.com*, 03 2009.

[49] E. Platanakis, C. Sutcliffe, and A. Urquhart. Cryptocurrency portfolios using heuristics. 08 2019.

[50] S. E. Said and D. A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[51] M. Vasek and T. Moore. There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams. pages 44–61, 01 2015.

[52] W. Whitt. Some useful functions for functional limit theorems. *Mathematics of Operations Research - MOR*, 5:67–85, 02 1980.

[53] A. Yelowitz and M. Wilson. Characteristics of bitcoin users: an analysis of google search data. *Applied Economics Letters*, 22:1–7, 01 2015.

[54] W. Zhang, P. Wang, X. Li, and D. Shen. Some stylized facts of the cryptocurrency market. *Applied Economics*, pages 1–16, 06 2018.