Correlation-based algorithmic trading applied to crypto markets

BACHELOR THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Software Engineering and Internet Computing

by

Leonhard Esterbauer
Registration Number 01526782

elaborated at the
Institute of Information Systems Engineering
Research Group for Industrial Software
to the Faculty of Informatics
at TU Wien

Advisor: Thomas Grechenig
Assistance: Dominik Schmelz

Vienna, November 12, 2019

Technische Universität Wien, Forschungsgruppe INSO
A-1040 Wien ▪ Wiedner Hauptstr. 76/2/2 ▪ Tel. +43-1-587 21 97 ▪ www.inso.tuwien.ac.at

Abstract

Over the last years, the price development of cryptocurrency markets encountered noticeable ups and downs. It looks like the dominance of a few big crypto assets like Bitcoin, Ethereum and Litecoin determine the markets and run ahead less popular currencies that behave similarly. In this work, a correlation analysis of major crypto assets is done and dependencies in price developments are searched. For this purpose, data from online sources are collected and compared using statistical measurements. Furthermore, lead-lag properties are investigated and approaches for using the gathered results in an algorithmic trading scenario are given. To get an overview of the presented approaches, performance measurements show the profitability over a selected timespan compared to approaches neglecting correlation values. The results show a high correlated crypto market with small groups of higher correlated assets. Moreover, no relevant lead-lag relationships can be identified in the selected asset combinations which refute the starting thesis that a few leading assets determine the market. Even if some assets look promising and show a higher correlation when compared to a forward shifted signal, the high correlation values are also measured when the same assets are compared to the signal shifted backward. For this reason, the lead-lag investigation does not provide significant values or clear indicators for a real trading scenario. An additional real-world trading simulation with promising signals supports this thesis. The simulation uses common trading approaches with an extension of correlation and lead-lag calculations and comes to the result that none of the presented approaches can outperform a traditional trading approaches with greedy analysis. Moreover, the presented approaches exhibit similar or even worse profits when simulated on historical data using backtesting. As only less research about price development of crypto markets exists, this work acts as an overview of cryptocurrencies and their correlation values and additional investigations are necessary to rate and classify the gathered results.

Keywords

correlation, trading, cryptocurrencies, crypto markets, lead-lag, algorithmic trading

Contents

List of Figures

List of Tables

List of Listings

1 Introduction

1.1 Overview

11 years have passed since Satoshi Nakamoto published the Bitcoin thesis which acts as the foundation of Bitcoin, the most famous cryptocurrency nowadays [25]. Since this event, the public interest in digital assets increased continuously. Especially the new possibility of trading cryptocurrencies online without access restrictions or closing times received great attention. The popularity also reflects in recent estimates which counted 50 million active investors on more than 100 exchanges worldwide [20].

In addition to the high usage of exchanges, the number of tradable cryptocurrencies has grown unmanageable. Services that try to summarize the biggest cryptocurrencies and their trading parameters like price and volume (A.1, A.2), are listing thousands of different cryptocurrencies on hundreds of exchanges reaching from assets with little market capitalization to assets with market dominance like Bitcoin has at the moment. Most popular cryptocurrencies like Bitcoin, Ethereum and Ripple showed high volatilities and price instabilities over the last years and it seems like the main drivers of the crypto prices nowadays are news about development progress, investments, regulations, and frauds.

As an example of measured, high volatility, the Bitcoin price around January 2018 can be used. One month before, at the peak on 17 December 2017 Bitcoin reached a price of over $19,700 at some exchanges before the price dropped to $6,200 on 5 February 2018 (A.3). But not only Bitcoin was affected by this major price drop. All major cryptocurrencies like Ethereum, Ripple and Bitcoin Cash showed similar price developments building groups of seemingly correlated assets (A.4).

Investigations like these count to the area of technical analysis and even though technical analysis is more controversial on stock markets than fundamental analysis, surveys like [21] showed that technical analysis is emphasized by market participants when talking about short time trading. In comparison to stock trading, the crypto market is a relatively young market and shows slight deviations in price movement and development and thus gives new possibilities to use technical analysis. Because of the long existence, the stock market analysis comprises advanced and well-understood techniques that try to predict future stock prices based on historical and present data. Most techniques, like simple chart analysis, can be easily adapted from the stock market to the crypto market and used as time series forecasting indicators. Advanced indicators like correlation-based analysis in stock markets exist in various forms and rely on different mathematical approaches.

A possible application of technical analysis is algorithmic trading. An important factor of this technique is to find suitable rules for a given data source and find a trade-off between data used for calculating the next steps and execution speed. As cryptocurrency data is publicly available at all major exchanges, financial agents can use many different data sources to gain reasonable trading information. The challenging topic for these agents is to filter the data and extract meaningful knowledge to get an information advantage other traders or trading agents do not have.

1.2 Problem description

Considering time-series price developments of cryptocurrencies over the last years, in some timespans it seems like the prices are strongly correlated. A closer examination using correlation analysis gives clear measurement values that can provide information if an asset should be bought or sold in a period. Algorithmic trading can benefit from the results and execute trading decisions depending on qualitative rules and settings. The main task of this work is to investigate if the crypto market exhibits correlation features and how they can be detected and interpreted using algorithmic trading. Additionally, it is interesting if some currencies encounter a lead-lag relationship that can be measured and used to enhance trading decisions. Moreover, a framework for using the gathered results in a real-world scenario is searched which consequently leads to the central and final question if a trader or algorithmic trading bot can use the gathered results to gain an information advantage over other participants in the trading game.

1.3 Motivation

The high volatilities and correlations occurring in cryptocurrency markets facilitate a new starting point in applying correlation-based indicators on chart data and lead to a new research area. As the crypto market is proscribed by most stock market traders, only a little research was done in investigating how well correlation-based indicators perform on the crypto market or if dependencies between the crypto-assets exist. The question is if there is a relationship between single or multiple digital assets and under which circumstances a trader can use the additional information to enhance the trading decision process. Considering volatilities, it is also possible to use correlations and groups of similar prices to observe information about trends in the crypto market and use the gained information to make predictions on future price developments of the whole market.

Another possible use-case is the investigation of correlations to help organizations and exchanges to regulate their cryptocurrencies and markets and reduce negative impacts of volatilities. Taking into account correlations for different groups of assets, also portfolio management is indeed able to reduce risks and give more value in calculating insurances. While correlation and lead-lag analysis give value to traders, the usefulness of an algorithmic trading program applying correlation-based indicators is elicited. As algorithmic trading and computational finance are profitable topics, only little public research of correlation-based algorithmic trading is available and companies keep findings private to get an information advantage if a scenario is applicable.

1.4 Aim of work

The experiment aims to develop a program that simulates a trading scenario based on algorithmic trading. The primary goal of the program is to investigate and measure the correlation and lead-lag relationships of the cryptocurrency market. As input sources different time-scaled data from different sources are fed to the program. Irregularities and incompatibilities of data are compensated with modifications of the software for possible special cases. With this approach, the question "*Are there correlation measurements in the crypto market which can be used in an algorithmic trading scenario?*" is treated. The result is a summary of the tested approaches and the clarification of the hypothesis if a trader can get an information advantage in using correlation analysis or not. Restrictive, any results of the experiment are only applicable to the used data, selected trading approaches and tested scenarios and can not give clear information about future price developments or other currency information. Additionally, the simulations of the trading scenarios are not considered to act optimal and instead represent a possible solution of an algorithmic trading program.

1.5 Structure of work

In the first step, before the experiment starts, fundamentals and state of the art topics are explained in chapter 2. Besides the theoretical and practical fundamentals of crypto markets, the used measurements and methods are explained. Following fundamentals and state of the art, chapter 4 describes how the experiment is done and what strategy is used. In the next step, the data collection process and a description of the data used are given in chapter 5. In this chapter also statistical measurements are examined and compared. The next chapter 6 provides the results of the experiment, where the first subchapters 6.1 and 6.2 investigate correlation and lead-lag properties of the data before subchapter 6.3 makes use of the gathered results and combines the correlation analysis with an algorithmic trading approach. Moreover, the used program and approaches of correlation indicators are described in this chapter and set in contrast to a naive greedy trading algorithm. Following, chapter 7 provides an interpretation and discussion of the gathered results and concludes the experiment before alternative approaches and possible future work not included in the following are treated.

2 Fundamentals

To get an introduction of dependence topics and fundamental methods necessary for the following experiment, this chapter describes prerequisites, additional information, and state of the art topics around finance, cryptocurrencies, and algorithmic trading.

2.1 Cryptocurreny markets

The fundamentals of cryptocurrency markets are similar to stock markets besides a few properties. The mostly free access to exchanges, the restriction to digital assets only, and the treatment of some digital assets as real money can be mentioned, just to name a few. In this context, digital coins are the foundation of crypto markets based on a distributed ledger system which often is immutable and decentralized to gain trust and security. Even though blockchain is only an approach of how a distributed ledger system can be implemented, the terms *blockchain* and *distributed ledger* are often used interchangeably. To ensure a distributed ledger reaches a common consensus, computationally hard problems based on cryptographic puzzles are used. The combinations of how these puzzles are used to create new coins, consume existing coins or reach a common consensus result in a variety of different cryptocurrencies. Two famous methods targeting these tasks are named proof of work and proof of stake. The concept of these approaches is to give every participant in the network a chance to create a new valid block in the blockchain which is accepted by other participants. In general, the block creation process is randomized and is attributable to the fraction of contribution to the network for which participants are rewarded with small amounts of money. While proof of work enforces participants to calculate the next block of transactions with high amounts of computational power, proof of stake puts less effort in the creation of new blocks and transforms the decision process of block creation to a fraction of how much a participant already owns.

A famous example of the proof of work implementation is Bitcoin. Here, proof of work uses the computation power in the network to compute SHA-256 hash functions for transaction blocks [25]. A participant with 50% of the computation power in the network obtains a 50% chance to create the next block and consequently receive the reward. A downside of this method is the high waste of computing power and the possibility of exploiting the network with a 51% attack [34]. In a proof of stake network like Peercoin [16], the fraction of creating a new block is not equal to the computation power but the amount of money a participant already owns. Besides proof of work and proof of stake, several other methods exist like proof of retrievability [24] or tangle, a consensus algorithm used by IOTA [27]. Because of the long and lasting dominance of Bitcoin, the term *Bitcoin* established as a popular example for cryptocurrencies and other, less popular cryptocurrencies are summarized with the term *altcoins* with is an abbreviation for *alternative coins*.

The economical purposes of cryptocurrencies reach from digital payments to smart contracts and decentralized apps. One example of a decentralized app is the social network Cent [1] based on the Ethereum blockchain [33]. Looking at a broader view, blockchain technology looks promising today and envolves in the area of manufacturing, supply chain management, Internet of Things and

[1] https://peepeth.com

healthcare. In cooperation with governments, industries try to enhance the mining process and prevent abuse of systems in developing coins for industrial use like Hyperledger Fabric [2] does. For the private area tutorials and generators exist which make it possible to build and distribute cryptocurrencies for own purposes with the effort of a few clicks. Deceivers use such tools to create cryptocurrencies and distribute them without care about value or backings and try to get money out of initial coin offerings. In this scenario, the support for the currency drops immediately and the creators leave with fiat money like U.S. dollars. Consequently, the price of the created coins drops and the investors lose their money. Another bad reputation of Bitcoin is the usage for illegal activities. Because of the pseudonymity of Bitcoin, many people feel safe and use cryptocurrencies for boodle, but in fact, investigations like [26] show how payments can be related and traced. An example of the impact of illegal activities on the Bitcoin price is the price drop in 2013 when the criminal organization silk road was closed.

Expect a few exceptions like Tether [30], most of the digital assets are unbacked by either physical commodities or sovereign obligation and not regulated or observed by a public institution or other organization resulting in a free market where demand and supply fix the price. Popular platforms like bitfinex [2] and binance [3] are exchanges specialized on online cryptocurrency trading and exchange. All major online exchanges provide their services in common with a public application programming interface 24/7 and make it possible for customers to build their clients and programs like trading bots. The high amount of currencies make it difficult to provide all markets and asset pairs at every exchange. The real price of cryptocurrencies like Bitcoin differs from exchange to exchange. An approach to determine the overall Bitcoin price is discussed in [7]. To provide tradable asset pairs and payouts in fiat currencies like U.S. dollars, exchanges have an extra effort and thus some exchanges like binance do not even provide a fiat currency pair. Instead, payouts of cryptocurrencies are used and single Bitcoin or Theter markets are introduced. This strategy forms new markets and reduces the amount of traded asset pairs so that a few cryptocurrencies like Bitcoin or Theter are used as lending currencies. In this scenario, the Bitcoin price is the dominating factor and is related to all the asset pairs which are not tradable to fiat currencies or other crypto assets. The prior usage of digital lending currencies enhance online-only trading and let users swap exchanges with low effort. The diversity of possible trading pairs and prices facilitate arbitrage trading. Besides the price divergence on different exchange platforms also price divergences at a local level can be observed. [20] investigates arbitrage trading at different levels and found price ratios from 15% to 40% for February 2018. At some point in time these ratios compensate and the different prices approximate prices are close to each other. If one currency leads this behavior and other currencies adjust is investigated in the following.

2.2 Technical analysis

Nowadays, a variety of different approaches exist that are trying to use a kind of technical analysis to get some advantage in gathering information other traders do not have. An example of how many technical analysis approaches exist can be seen in books about the analysis of time series data like [31], [32] and [28]. If technical analysis has value at all and the thesis that technical analysis gains more value when talking about short time trading is investigated in [21]. Correlation-based analysis, which is the target of this experiment, is part of the technical analysis area and can be applied to data in various forms. In general, correlation analysis is trying to find dependencies between assets, for example with autocorrelation like in [4]. Other approaches try to extract important characteristics or try to find relations between single features of financial data. An example

[2] https://www.bitfinex.com
[3] https://www.binance.com

of this proceeding is the evaluation of stock returns and the relation to the trading volume [6]. The possibilities of finding relating parameters are endless and reach from simple comparisons of single assets to finding complex relations between any kind of data like social signals and return values as used in [11].

2.3 Algorithmic trading

Algorithmic trading is a method to support a trader in executing orders while he or she does not need to actively perform the orders. This means the trader does not have to wait for specific events in time and let computers do the work. To act as requested, algorithmic or computer-supported trading needs specific rules to decide how and when a trading decision is suitable for a given scenario. For this purpose, in most cases, a set of technical analysis methods with combinations of simple commands is used as e.g. the instruction to sell an asset if the price hits a specific value. Most exchanges like binance and bitfinex already provide kinds of algorithmic trading approaches to their users as a feature to not being in front of a computer the whole time. Also in stock markets, algorithmic trading is widely used to submit orders and make certain trading decisions. [15] claims that 73% of the trading volume in the USA in 2009 was executed using a kind of algorithmic trading. Additionally [15] found out that algorithmic trading improves liquidity and enhances the informativeness of quotes. As an example of how algorithmic trading is implemented [11] studied the correlation of Bitcoin prices and social signals. In the following experiment, a similar approach is used for simulating the effects of correlation-based indicators.

2.4 Backtesting

To exhibit how well algorithmic trading rules behave, a kind of performance measurement is required. In financial scenarios, the aim of trading rules is to gain better profitability as some other rules provide. For this purpose, some given rules are applied through a trading simulation on historical price data and compared afterward to find the best combination of rules. The resulting procedure is named backtesting and common for measurements of machine learning models or trading strategies. Besides the simple usage of this method, it gives also clear results on how well trading decisions behave on historical or synthetic generated data including special cases and extreme values. The gained overall information on performance measurement helps to improve the trading decisions resulting in better profitability.

2.5 Correlation analysis

Correlation analysis is a statistical approach to measure relationships of two quantitative variables like weight and height of a human. In general, all quantitative variables can be correlated but do not need to be. For example, the experiment in [22] shows the existence of a correlation between babies and storks from a statistical view. With this experiment, a translucent correlation is clarified and the importance of a qualitative interpretation and the question for significance is explained. Besides the common relationship, a causal relationship is possible. In this case, a change of one variable fully or partly causes a change of another variable. In general, changes of variables are caused by various influences. As example for various influences at the stock markets good or bad news of a joint-stock company must be taken into account as well as the actual economic situation etc.

The applications of correlation analysis are spread and reach from investigations in the area of medicine, where consequences of diseases and drugs are measured, to the area of marketing where correlation is used to measure the effects of advertisements in a specific region. A special case of correlation analysis is the time series forecasting approach which is part of regression analysis. In this case, a signal is investigated and a mathematical formula is searched to describe the underlying time series data. In the last years, regression analysis gained much attention as part of many machine learning approaches.

The correlation analysis measurements used in this experiment are limited to mathematical methods that exist in abundance. Two of the most used correlation measures are Pearson (2.1) and Spearman (2.2) correlation. Both calculation methods produce a single value in the range from -1 to 1 that represents the association of two quantitative variables. While the value 1 denotes a perfect monotone increasing relationship of the variables, a value of -1 denotes a perfect monotone decreasing relationship and a value of 0 denotes no relationship. This means if the correlation value is negative, one variable changes its' direction similarly in the opposite direction than the other variable and if the value is positive, the varaibles change their direction similarly in the same direction.

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{2.1}$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{2.2}$$

While Pearson is measuring linear relationships between two continuous variables, Spearman is used to evaluating monotonic correlations based on ranked values for each variable. [14] came to the conclusion one should not overinterpret Spearman's ranked correlation and if linear dependencies are measured Pearson's correlation is sufficient. To get an overview of the correlation values of multiple variables, correlation matrices are a popular way to visualize correlation values with different colors for different correlation ranges.

To investigate a lead-lag relationship the factor of time has to be included in correlation analysis. For this requirement, a shift of one time series variable is applied and the calculation is repeated. Comparing the results before and after leads to the conclusion if a variable is strong or less correlated to future values of another variable. With this approach, a lead-lag relationship of any timespan can be investigated and compared to other time shifts. To get reasonable results, all values must share the same resolution and do not exhibit outliers or missing values as these lead to falsified correlation values. Additionally, lead-lag values must be interpreted and the meaningfulness of the results must be rated. As high correlation values could occur everywhere and even randomly generated values could achieve high correlation values, a reasonable investigation of the lead-lag signal is advisable to reach a higher degree of significance.

An extension of this approach is autocorrelation which investigates the correlation with the signal itself. This means the correlation values of the same signal shifted back and forth in time are calculated and compared. In the financial sector, this approach should not give much information over price development as most of the prices are independent and do not repeat exactly in different periods of time. Another extension of the correlation analysis which is also applicable to the lead-lag analysis is the investigation of how correlation and lead-lag dependencies change over time. For this purpose, the correlation values are calculated for every point in time with a windowing function that calculates the correlation of the values for a certain timespan from the past to the actual point. With this method, a chart of correlation development is produced containing values from -1 to 1 on each specific point in time.

2.6 Price data

Prices of different assets can be evaluated in several ways. One way is to use a given fixed price value with no additional information. Another way is to calculate the price value for a defined scenario from given information like trades data. In this case, the price for an asset exists only when an ask and bid request meets the same conditions and a price for an exchange is found. Furthermore, the actual price of an asset is elicited in taking into account the last successfully executed exchange. This means a change of the actual price is only given when a higher or lower ask and bid request is matched and executed. As an example of how the price can be represented in dependence of time, a time bucket of 1 minute is specified. In this time bucket, the last successful match of an ask-bid request builds the final price and is therefore used at this timestamp t with a resolution of 1 minute. All other trades data in this minute is neglected and does not have an impact on the final price. Even if this is the regular case, it is indeed possible that exchanges adjust their prices because a price change of other assets or exchanges occurs and no trade was executed in a given time bucket. A common use-case for this measurement method is the ohlc (open, high, low, close) data which is often used for visualizing price charts.

2.7 Performance measures in financial markets

The aim of most strategies in financial trading topics is to achieve high profits. To measure how much profit is returned various approaches exist. In (2.3) a simple profit calculation is described where $R(t)$ is the profit which is called return value $P(t)$ is the current value and $P(t-1)$ is the last value of the investment.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.3}$$

When applying this formula on time series data, a function in dependency of time t is produced and can be used to get the profit of the investment at time t. Another approach, based on the profit function $R(t)$, that makes use of logarithmic returns instead of simple returns is shown in the formula (2.4). This kind of modification is widely used to calculate the sum of returns for different time periods as the logarithmic approach makes it possible to sum log-returns across time periods.

$$r_t = \log P_t - \log P_{t-1} \tag{2.4}$$

A benefit of the approaches in (2.3) and (2.4) is the normalization of the data which makes it possible to compare the results among different scaled prices.

Another factor which has to be taken into account is the usage of fees and spreads. In a real-world scenario, a fee must be paid to the exchange platform or trader to execute a trade. This fee is different from platform to platform and range from high rates of 5% or more to a small rates of 0.1% and less. Depending on the currency and amount of trading volume, some exchanges apply different fees on specific use-cases. Additionally to fees, an important aspect of trading is the spread of buy and sell prices. The amount of these spreads can differ from time to time and build a bridge of supply prices to demand prices in the markets. It turns out to be difficult to take into account spreads in real trading scenarios as the market and the supplied and demanded orders can not be predicted and thus are often neglected in simulations. In contrast, fees are a fixed value and are applied depending on rules which are static in advance so simulation is practicable.

3 State of the art

3.1 Correlation analysis

In the financial sector, a variety of different approaches for modeling and abstracting the price development process exist. The aim of these approaches is to predict future trends in the market and make solid trading decisions for better profitability. As this looks like a complex problem, most approaches concentrate on getting little information gain for single assets. For this, investigating the dependence of price changes triggered by external factors like prices or trading volumes from other assets is an obvious starting point for research and also the topic of this thesis. More discrete, the dependence of asset prices and their correlation values, as well as lead-lag performances are investigated similar to the following approaches.

In general, the selection of dependent parameters for a correlation or lead-lag analysis is widespread and can reach from trading-related parameters like volume, as described in [6], to any other external related parameter like the number of employees a joint-stock company exhibits. Taking into account finance-related Google searches, [18] tries to identify cross-correlation with Dow Jones Industrial Average component stocks. The research shows, that there is no evidence for a correlation of online searches for Dow Jones Industrial Average index and the actual price development.

Additionally to basic correlation analysis, there are also existing experiments that use correlation-based analysis to find out lead-lag relations between time-series data like in [8]. In this paper, a method for estimating covariances is developed and applied to the index and futures market on a high-frequency basis of 1 minute time buckets with the data of the S&P 500 index in 1997. The ambiguous results are not quit sure if the futures market leads the index market by at least 10 minutes or the index market leads the futures market by at least 2 minutes. Besides this specific experiment, the question of who leads whom in the relation of the index market and futures market is, in general, a famous field of use treating lead-lag analysis. Other papers that investigate this question can be found in [10] and [5], which both came to the conclusion that the evidence of a leading futures index over the market index is higher than a leading market index. Even if the authors are not sure and admit that the findings should be treated with caution they give insights on how complex and unpredictable financial markets are.

An experiment that takes the comparison even further can be seen in [29] which investigates data of 79 stock market indices and tries to find out lead-lag gaps and dependencies in the time-shifted opening hours of different time zones. The results show the direct effects of price changes between countries in different time zones where the opening hours of the markets are slightly different. Especially, asset prices from eastern countries like China and Taiwan are strongly correlated to the S&P 500 index from the US markets one day before.

3.2 Technical analysis

More modern approaches like [1] use machine learning to find a regression model that is applicable to future data and eventually finds hidden dependencies a human or a simple mathematical approach like linear forecasting does not find. As machine learning and artificial intelligence encountered a big hype in the last years, more and more scientific papers investigate the possible superiority of artificial intelligence like [9] does, where deep learning and cross-correlation are used to forecast trend-changes in financial time series data. Indeed a modern approach, [12] is using machine learning to investigate how critical transmissions can be recognized in a topological way using k-means clustering.

In addition to scientific work, experiments and applications in the open-source sector exist which abstract web-clients or APIs of trading platforms into libraries or frameworks with additional features. Some open-source projects go even further and provide selections of different algorithmic trading or machine learning approaches a user can choose of. Scenarios for applying artificial intelligence exist also in the business area, where programs are used to simulate various possible market trends or support traders executing trading decisions. Alladin [1], a product of BlackRock claims to investigate all important markets to provide solid trading decisions and acts as a big data analysis tool businesses can buy and use for high sums of money.

Considering correlation-based indicators at a more general view, [23] shows opportunities and limitations of correlation trading strategies and how they can be used in getting an information advantage. For this reason, six different correlation methods common in financial scenarios are compared for applicability and performance on stock market data which came to the conclusion that most strategies depend on the levels and volatilities of the underlying assets.

More specialized investigations like [19] try to find the main driver of the Bitcoin price. In this paper, various influences over the Bitcoin price are investigated and analyzed with the wavelet method which leads to the conclusion that Bitcoin does not exhibit typical currency properties and acts similar to shares in the stock market. Another interesting finding of this paper is the price reversal at the end of each Bitcoin mining circle, i.e. when the difficulty of mining gets harder and the participants in the network get less reward for creating a new transaction block. Taking into account other cryptocurrencies, [13] analyzes price data using the Augmented Dickey-Fuller test and found out that the price data of the selected altcoins Ethereum and Monero are not stationary. Furthermore, [13] investigates causal relationships and found contradictory results that require future analysis with more data as we do in the following experiment.

Besides the trading optimization topic, there are questions about the regulation and economical usage of cryptocurrencies as pointed in [3]. Papers like [17] analyze the investment possibilities of Bitcoin and compare volatility, correlation and portfolio performance to more convenient investment assets like gold. How far the new sector in the financial world influences everyday life and what impact the new currencies have is not clear and must be investigated in more detail.

[1] https://www.blackrock.com/risk-management-with-aladdin

4 Methodology

The investigation of how financial markets behave is usually done in an experimental quantitative way as it is not possible at the moment to predict price changing triggers in advance precisely. For this reason, the first step of this work is to collect meaningful market data on which the following measurements and experiments are applied. The second step consists of processing and preparing the collected data for unified usage in the research process. In a third step, the data is analyzed for statistical measurements and a first overview of the data is achieved in a quantitative way before the actual investigation for correlation and lead-lag properties starts. The final step of this work is to apply the data to a trading simulation representing a real-world scenario where trading indicators supported with correlation and lead-lag analysis are examined in a qualitative way.

4.1 Data collection

In a first step, different data sources are collected using online services or crypto exchange platforms. Through the dominance of online exchanges, the data collection procedure is simplified and provides various open data sources to get valuable measurements. Afterward, the collection of different exchanges and assets is qualitatively investigated and checked for validity. Because it is not possible to trade all cryptocurrencies in an equal way at a central place, a comparison of all cryptocurrencies is neither meaningful nor possible. For this reason, an empirical evaluation of the most promising data is used for this experiment. This specification is achieved in taking into account the 20 most important crypto-assets according to their market dominance in trading volume at the moment. As five of the biggest cryptocurrencies already cover a market share of more than 80%, the selection of 20 assets is big enough to provide a first meaningful statement of the whole crypto market.

Additionally, cryptocurrencies that are stable and only act as representation for other fiat currencies like Theter are neglected and will not be investigated in the next steps. The reason for this is the constant trading price of the cryptocurrency which does neither provide volatile prices nor a reason to invest. In addition, the data collection process takes care to obtain all necessary data sources for the respective experiments like high-frequency data and day to day data with wider timespans. For a qualitative investigation of high-resolution data, only a small selection of 3 common cryptocurrencies is examined. This step is introduced because of the limited access to high-resolution data from exchanges. For this smaller selection also data from different exchanges is collected for the purpose of investigating exchanged based correlation and lead-lag analysis. An explanation of what data is used and why it is used concludes the first step. In the case of errors or missing data, a last-value procedure or forward-fill is applied. These procedures replace all faulty or missing data entries with the last available correct value to ensure a continuation of usable data without interruptions. Afterward, the processed data is made publicly available so comprehensibility and repeatability is ensured.

4.2 Data evaluation

A first investigation of the data consists of eliciting and comparing common statistical measurements including mean and standard deviation of the price data to get an overview of the volatility of single assets. In a next step, the data is investigated for correlation using Pearson's and Spearman's correlation values. Pearson's and Spearman's correlation are common methods in statistics to measure how correlated two signals are and therefore usable and valuable for a correlation investigation of time-series price data. Besides investigating the pirce data, correlation values for the common financial measurement values of profit and log-profit are also included which may find a correlation indicator on another scale. To cover a wide range of possible correlation investigations and filter high volatilities which can smaller the correlation values of some comparisons, a qualitative investigation of rolling correlation values with an averaged window is applied in a qualitative way. In addition, correlation values of different selected exchanges are examined to cover the existence of another possible trading indicator from the gathered data. The results of all analysis steps give a first assessment of the correlation of the cryptocurrency market and their dependencies in the selected timespans.

Additionally to the correlation analysis, a lead-lag analysis is performed in this step. For this purpose, the correlation values of shifted asset prices are compared to non-shifted prices. If the correlation values of the shifted price data are higher than the correlation values of non-shifted data, it is possible that one asset leads another asset in case of price development. With the aim of finding lead-lag related signals, it is also possible that a signal indicates its own price development which is also investigated in this step. Concluding, this means if a price signal is higher related to the past of the same or another signal, it acts as a promising candidate as an indicator signal in the algorithmic trading approach in the following experiment.

4.3 Experiment

After performing the first analysis steps, an algorithmic trading approach is implemented in a prototypical way which only fulfills the purpose of representing a real trader or an algorithmic trading agent. Because of the testing-environment and the fact, that the experiment can be simply repeated in case of failure, only little error and crash handling are implemented and a simple source code is achived that is able to process different time-scaled data with little modifications. Additionally, important parts and procedures of the program are highlighted and explained in detail to ensure readability and reproducibility of the experiment.

In the next step, the program is used to perform simulations. In order to do so, tests with a greedy solution are executed to get a baseline for the amount of profit an algorithmic trading program is able to achieve. For this purpose, a simple moving average technique is used with qualitative investigated parameters and time windows. Afterward, the extended analysis of lead-lag strategies is applied to the program using promising signals and indicators from the analysis step before. In the next step, supporting correlation based indicators are added to the simple moving average approach and a strategy to enhance the indicators is investigated. The simulation is done using the backtesting of lead-lag trading indicators on the test data. To cover a wide range of lead-lag relationships, the different time-scaled data sources are investigated in time-buckets of 1 day, 1 hour, 1 minute, 10 seconds, 5 seconds and 1 second and afterward compared in a qualitative way using a small number of assets. Furthermore, performance measurements of the selected approaches and assets are compared with different manually adjusted algorithm parameters.

The procedure of the experiment and parameter adjustments during the simulation runs are explained in detail to ensure reproducibility and comprehensibility. A confrontation of the gathered results and an additional explanation in common with an interpretation follow. In the last step, a conclusion of the experiment and a discussion about other fields of work and possibilities completes the work.

4.4 Tooling

For the data collection process, data fetchers written in JavaScript are executed in a nodeJS environment for rapid prototyping reasons. In sum, three different data collection programs are created, whereby one fetches data from an online portal and two others make use of an API provided by the selected exchange platforms. While data from the online portal does only consist of small sizing day data measurements, the high frequency trades data needs to be managed and preprocessed for the next steps. For this reason, the trades data fetchers feed the gathered data directly into a timescaledb [1] database instance which is a modified PostgreSQL database optimized for time-series data like the price data of this experiment is. Some additional advantages of this database are the provided functionalities for calculating prices of time buckets and contained features to forward-fill missing values on a high-performance level. Another step to boost up performance is to outsource all common aggregate functions like min, max, count, etc. to the database where common SQL functions are implemented in an optimized way. To process the day data sources which are available as .csv-files, the performance of the python programming language with additional libraries like pandas and numpy is sufficient. For the simulation part also the python programming language in a combination with a jupyter notebook instance optimized for data science purposes is used. The chosen environment is a collection of common data investigation tools that provide online support and out of the box functionalities for data investigations and thus is suitable for the following experiment.

[1] https://www.timescale.com

5 Data

The biggest publicly available exchanges are online exchanges and thus, all data for this experiment is collected from online sources. To get an overview and assessment of the different cryptocurrencies and exchanges, data from the online platform coingecko is used [1]. On this platform, exchanges are compared about relevance and popularity which is achieved in comparing day trading volume in USD. In further consideration, a rise and fall of big cryptocurrencies exchanges and their changing trading volume over the last years can be observed.

5.1 High resolution data selection

The first exchange selected for quality investigations with high-resolution data is the Kraken exchange platform [2]. Kraken is existing since 2011 and has always been a popular and stable exchange platform with the possibility of trading fiat currencies. Even other exchanges outdated Kraken in trading volume (A.7), the feature of deposit and withdraw fiat money makes it an important player on the crypto market. Sorting the 24h Bitcoin table for trading volume, Kraken is holding two positions in the top 10 with asset pairs in EUR and USD (A.5) which further highlights its importance. The second exchange selected for this experiment is Bitfinex exchange platform [3]. The importance of this platform can be seen at the incidents of the last years where negative rumors of Bitfinex lead to an immediate drop of the Bitcoin price. Besides the high trading volume in USD (A.5) and the interconnectedness to Theter, coingecko lists Bitfinex at position two in their popularity factor (A.6). As a matter of fact, the binance exchange platform [4] showed higher trading volumes at the moment but misses comparable asset pairs which can be traded at other exchanges in an equal way. Theter, which is the most traded cryptocurrency nowadays, comes close to this requirement but the price developments still show slight deviations of a few cents over some timespans and thus makes it hard to calculate a real profit in USD or get comprehensible comparisons.

In the next step, the cryptocurrencies from the selected exchanges are chosen. For this purpose, all cryptocurrencies at coingecko are sorted by their 24h trading volume (A.8). The first cryptocurrency in the list is Theter which is neglected due to the stable representation of USD. For a qualitative investigation, the next three cryptocurrencies are chosen for the accurate part of the experiment, which are Bitcoin, Ethereum and Litecoin. These cryptocurrencies are tradeable to USD at both, Bitfinex and Kraken, which make it possible to compare the prices for the same assets at different exchanges. As these platforms provide chart data only to a rough level like 1min, the trading data is collected in a different way. For this purpose, a data grabber is created which collects all trades data using the API Endpoints of the platforms. The actual price data at a specific time is calculated in taking the last value a trade was successfully matched at that moment.

[1] https://coingecko.com
[2] https://www.kraken.com
[3] https://www.bitfinex.com
[4] https://www.binance.com

5.2 High resolution data overview

The data collected comprise of values from June in 2019 and reach from 01.06.2019 to 30.06.2019. In table 5.1 a summary of the gathered trading data shows higher trading volume at Bitfinex while Kraken only examines 0.27-0.55 times the amount of Bitfinex. Also, the different statistical measurements like average price and min/max prices differ to a certain extent at both exchanges visible in table 5.2. As a measurement for volatility, the price standard deviation can be used. Here, all investigated currencies exhibit high standard deviations and price changes. Taking into account the price change of Bitcoin, the minimum value and maximum value are far apart and differ more than 6000 USD in a single month which gives a good starting point for correlation investigations while the price data is volatile. Even the price data includes only one month of trading data, the high-resolution data is eligible for investigating a lead-lag relationship between the two exchanges Kraken and Bitfinex.

Asset pair	Trades	Price volume	Trading volume
Kraken Bitcoin/USD	952445	95977099503,50	270480,5017
Kraken Ethereum/USD	113000	14210346,65	1271428,6336
Kraken Litecoin/USD	378000	106540005,73	2517450,6440
Bitfinex Bitcoin/USD	2080000	20980570133,63	487682,3823
Bitfinex Ethereum/USD	882530	250365265,83	4698383,1409
Bitfinex Litecoin/USD	552793	70645590,50	5322314,2948

Table 5.1: Summary of high-resolution trades data

Asset pair	Average price	Price standard deviation	Min price	Max price
Kraken Bitcoin/USD	10076,91	1718,59	7425	13875,70
Kraken Ethereum/USD	125,75	12,12	98	145,36
Kraken Litecoin/USD	281,85	31,91	226,56	363,29
Bitfinex Bitcoin/USD	10086,81	1737,64	7455	13764
Bitfinex Ethereum/USD	283,69	31,38	226,01	363,30
Bitfinex Litecoin/USD	127,79	11,26	98,16	146,95

Table 5.2: Statistical measurements of high-resolution trades data

5.3 Day data selection

For a quantitative investigation, the averaged and weighted data from coingecko is used. The price values are calculated by taking into account the percentage share of an asset pair on an exchange of all listed asset pairs for the cryptocurrency listed at coingecko [5]. For example, the effects of a price change of Bitcoin/USD at Kraken have less impact on the overall price because this asset constitutes only less than a half percent of all trades listed at coingecko. With this scenario, a stable and general price for a cryptocurrency is ensured and because of the multitude of exchanges and asset pairs listed at coingecko these values are reliable and meaningful.

[5] https://www.coingecko.com/en/methodology

5.4 Day data overview

In sum, the data collected from coingecko represents asset prices from 20 cryptocurrencies which contain daily price data from 01.01.2018 to 01.08.2019. In table 5.3 statistical measurements for the day data are calculated. In contrast to the short time high-resolution data, the spread and price standard deviation of the day data is even higher when comparing the prices of Bitcoin, Ethereum and Litecoin. The high volatility can also be observed when looking at Neo, Dash or Qtum which provide a max/min ratio of up to 64. In contrast, less popular altcoins like Paxos-Standard and Huobi-Token provide a low max/min ratio of 5x or even less.

5.5 Data processing

Both data sets contain errors and missing values that are replaced for the experiment. The raw data sets can be downloaded from Github [6]. While the trades data is fixed using the locf method from timescaledb which forwards the last observed value, the day data is treated with the ffill method of pandas in python which does the same. For initial missing data, like it is the case at Paxos-Standard, the value of NaN (Not a Number) in python was set as no price value existed before.

Currency name	Average price	Price standard deviation	Min price	Max price
bitcoin	7506.05	2569.12	3216.63	18343.66
ethereum	474.66	301.68	83.79	1448.18
litecoin	104.01	62.3	23.06	289.93
eos	7.8	4.0	1.73	21.27
bitcoin-cash	852.71	560.13	76.93	2988.64
ripple	0.66	0.46	0.26	3.4
tron	0.04	0.03	0.01	0.23
ethereum-classic	16.24	8.76	3.49	44.34
bitcoin-cash-sv	93.01	25.16	42.02	209.19
paxos-standard	1.01	0.01	0.98	1.04
zcash	224.88	146.96	49.07	880.2
stellar	0.28	0.13	0.09	0.88
dash	332.84	250.12	58.51	1220.51
neo	49.83	41.74	5.61	198.38
qtum	13.78	14.23	1.56	100.22
binancecoin	11.09	3.43	4.47	24.37
cosmos	0.17	0.06	0.07	0.28
cardano	0.21	0.2	0.03	1.18
monero	166.53	95.87	38.4	542.33
huobi-token	2.33	1.13	1.04	5.93

Table 5.3: Statistical measurements of day data

[6] https://github.com/estallio/crypto-correlation-analysis

6 Results

6.1 Correlation

The following chapter investigates correlation behavior and shows measurements for the high-resolution trades data and the low-resolution day data. In figure A.9 and A.10 Pearson's and Spearman's correlation values of the high-resolution trades data are investigated with different time bucket sizes and the results are visualized in correlation matrices. The calculated correlation values show a strong correlation between the Bitfinex and Kraken exchange platforms and a high correlation between Bitcoin and Ethereum while Litecoin's correlation is significantly lower. Even if a high correlation between the exchange platforms was expected before, the actual measurement values do not provide clear signs for trading decisions in the chart and a closer examination is necessary in the algorithmic trading process.

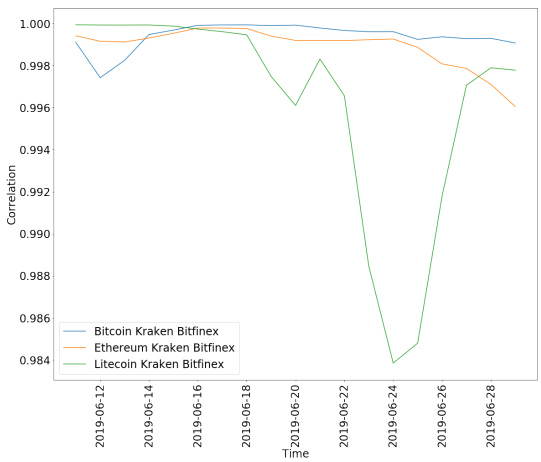
Looking at the correlation of Bitcoin and Litecoin in figure A.9 and A.10, the correlation values are smaller than 0.55 in all cases which do not look promising for future usages. Besides the correlation values of specific assets, the similarity of Pearson's and Spearman's correlation matrices can be observed, whereby Pearson's correlation shows finer gradation which is recognized in the Litecoin columns and is attributable to the different methods of correlation calculation. Looking at the matrices at a general view, there is no difference in the correlation of different time buckets visible. This means the correlation is equally spread among time buckets and provides the same values when comparing the last executed trade of every second, minute or day.

Besides the trades data, the day data visualized in figure A.13 and 6.3, show similar values when comparing Pearson's and Spearman's correlation coefficients. Although the data cover different timespans and resolutions, similarities between matrices and value ranges can be observed like at the trades data which shows again the affinity of Pearson's and Spearman's correlation values. When looking at single assets, the broader selection of different currencies shows a wider correlation range from less than -0.4 to 1, but in fact, the majority of the correlations are located in the area from 0.4 to 0.9 or even higher looking at Bitcoin-Cash and Ethereum. This behavior can also be seen when comparing the correlation values from Dash, Neo, and Qtum where Spearman's correlation shows nearly the same correlation values among the currencies while Pearson's correlation shows slight gradations that are also attributable to the different calculation steps of the correlation values. Even if there are differences between Pearson's and Spearman's correlation values, they are too close to favoring one method over the other in this case. Additionally, some currencies like Zcash or Dash show high correlation values with both calculation methods while other currencies like Paxos-Standard and Cosmos do not. A reason for this observation is the novelty of Paxos-Standard and Cosmos which results in fewer price observations than established assets exhibit. For simplicity reasons and the equality of the correlation measurements in this and further experiments, Spearman's correlation is neglected in the following results and measurements for Pearson's correlation are preferred.

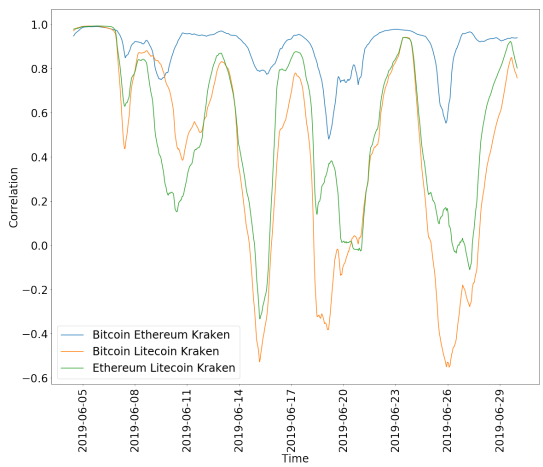
A comparison of the correlation of profit and log-profit calculations for the high-resolution data is visualized in A.11 and A.12. Similar to the Pearson and Spearman comparison, the profit and log-profit calculations are not different enough to prefer one calculation method over another as the measurements even include the same value ranges. In contrast to the default correlation measurements in figure A.9 and A.10, the profit and log-profit correlation values are different when looking at different time-scaled data points which is attributable to the price ups and downs not considered when calculating 1 day time buckets. For the measurements in the minute area and below, the correlation drops to less than 0.2 in most of the cases while the profit correlation at 1-day buckets is nearly the same at the end of each day. These investigations show slight differences at different exchanges and give a possible way to gain valuable data for trading indicators.

In contrast to the profit and log-profit calculations for trades data, the investigations of day data are less valuable for future indicator usages. Even though the correlation values in figure A.14 and A.15 look similar to the default correlation investigations of the day data in A.13, the value ranges are a little lower and about 0.2 lower across the board. The reason for this observation is the lost standard deviation in the calculation of profit and log-profit which leads to a worse result in this case.

The next comparison in figure 6.1 and 6.2 shows measurements of how the correlation changes over time when investigating high-resolution trades data and day data respectively. In fact, the correlation changes all the time and only stays in a similar range with other currencies for a small period of time. For example, the Bitcoin correlations in figure 6.2b are in a similar range from April 2018 to June 2018, after and before a drop occurred and the correlation diverges. A similar behavior can be seen in figure 6.2a from December 2018 to January 2018. Besides the small timespans with high correlation values, similarities in the drop behavior can be seen. This is the case in figure 6.2a around February 2019 or in figure 6.2b around August 2018 where all selected altcoins drop their correlation with Bitcoin in a similar way. The frequency and strength of these price developments look like good and bad rumors of cryptocurrencies brought the crypto market out of balance. In contrast to the high similarities of the day correlations, the investigations of the high-resolution data in figure 6.1b exhibit higher volatility. The only constant correlation level in this investigation is the Bitcoin price between the exchanges which diverges sometimes but corrects instantly to the price of the other exchange. In figure 6.1a the maximum measured deviation can be observed in June 2019 where the Litecoin price at Kraken and Bitfinex diverges to a correlation value of 0.984 which gives only a little chance to find lead-lag indicators when comparing different exchanges.

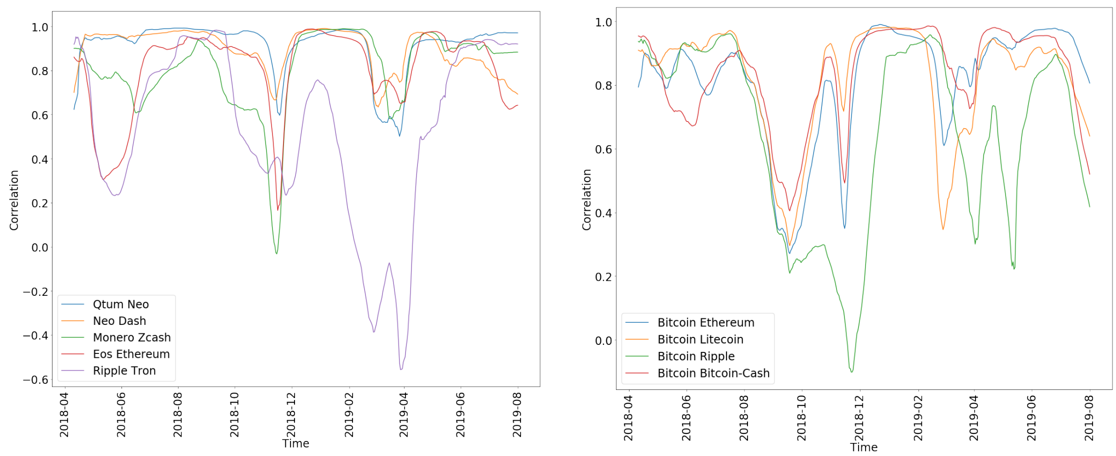


(a) Rolling Pearson correlation of Bitcoin, Ethereum and Litecoin at Kraken exchange platform compared to the same assets at Bitfinex, time buckets: 1 day, window: last 10 values



(b) Rolling Pearson correlation of Bitcoin, Ethereum and Litecoin among each other at Kraken exchange platform, time buckets: 1 minute, window: last 5000 values

Figure 6.1: Rolling Pearson correlations of trades data selection



(a) Rolling Pearson correlation of day data selection, window size: 100

(b) Rolling Pearson correlation of day data selection, window size: 100

Figure 6.2: Rolling Pearson correlation of day data selection

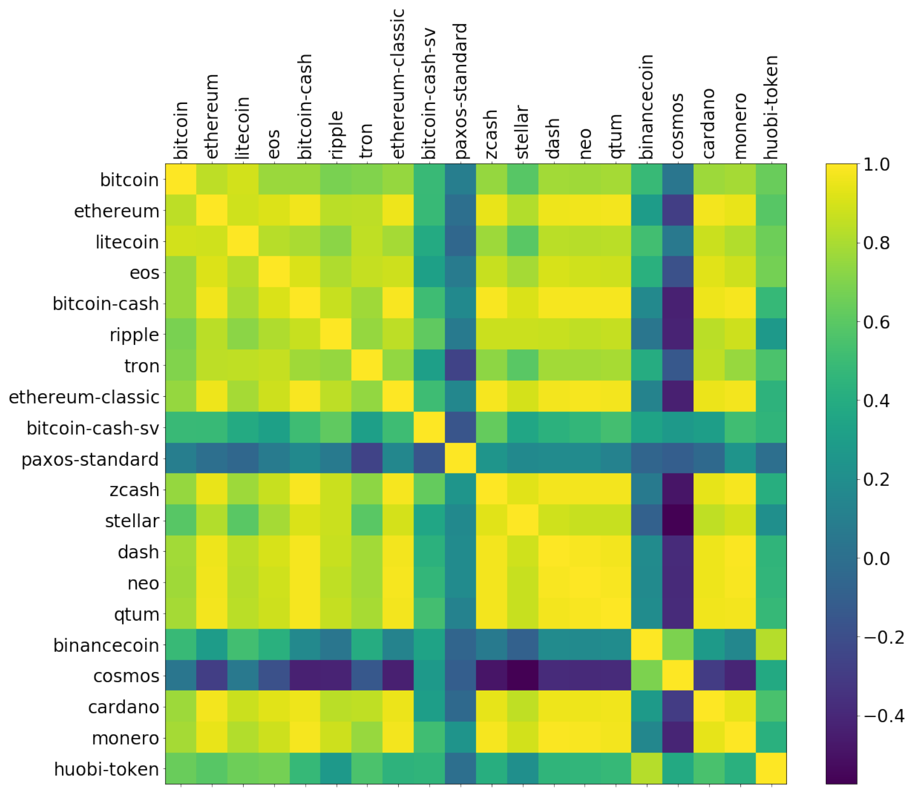


Figure 6.3: Spearman correlation of day data

6.2 Lead-Lag

In the next step, the lead-lag effect of different cryptocurrencies is investigated. For this purpose, the results from chapter 6.1 are compared to slightly shifted price data. In figure 6.4 two correlation matrices of high-resolution trades data with different time resolutions and shift-ranges are shown. Looking at a shift of 1-day for a resolution of 1-day time buckets in figure 6.4a, a strong correlation of the original prices between the exchanges can be observed while the shifted prices correlate in a lower level of 0.95. The higher resolution of the time buckets in figure 6.4b with a

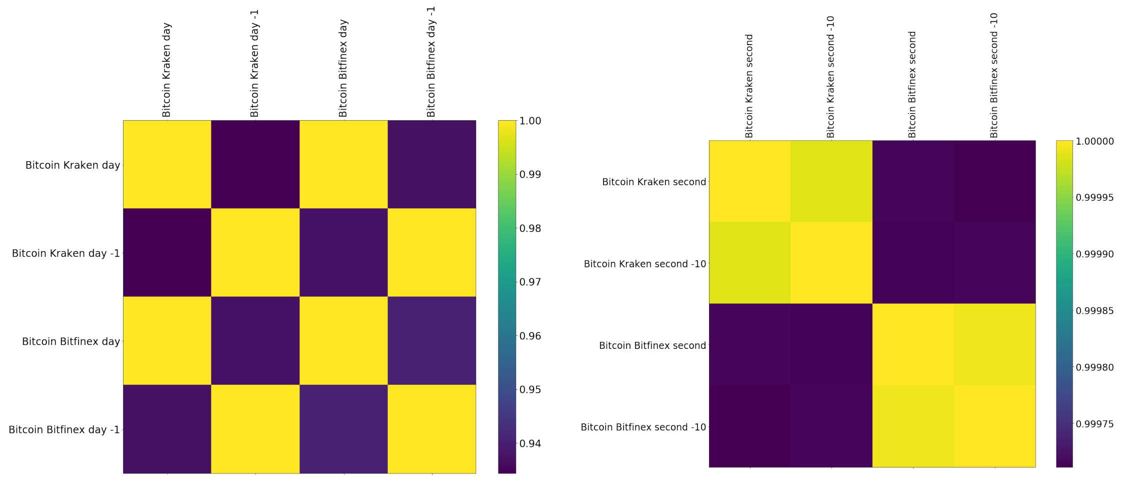
10-second shift shows fewer differences and a slight gradation of the shifted signals which indicates a slow price development in the 10s-area and a fast price development over the day. Even if the correlation values are high and indicate a dependency of the selected signals, a shift does not enhance the correlation of the different cryptocurrencies and exchanges and is therefore not suitable for future lead-lag indicators. The findings of these graphics are representative for the comparisons of other trades data which result in similar matrices that are neglected for simplicity reasons.

While the high-resolution data provides no useful lead-lag relationship, the day data in figure 6.7 exhibits a wider range of correlation values for shifted signals. To investigate the shifted correlation coefficients, two correlation matrices with subset selections of the selected 20 cryptocurrencies are compared. The first comparison in figure 6.7a shows correlation values of popular und dominating cryptocurrencies like Bitcoin and Ethereum while the second comparison in figure 6.7b contains values for less dominant coins like Monero and Neo. Except for Bitcoin-Cash and Ethereum, no pair of assets in figure 6.7a shows a higher correlation than 0.9. Neglecting the high correlation values, even the higher correlated Bitcoin-Cash and Ethereum measurements do not provide a clear sign that one signal acts as a possible forecasting indicator for another signal. Nevertheless, the correlation of these signals exhibit the the most promising values and thus Ethereum and Bitcoin-Cash are selected for the simulation part in the next section.

In the second comparison in figure 6.7b, a similar behavior can be observed looking at the altcoins Monero, Qtum, Neo and Zcash. The higher correlation values of the data one day before with the data from another asset one day before is a negative observation for using these signals as forecasting indicators. This case can be seen at the correlation values of Monero and Neo, where the default data of Neo and the data of Monero one day before show a high correlation, but the correlation of Neo one day before and Monero one day before is even higher.

To include a wide range of different correlation investigations in this experiment, in figure A.17 the unlikely existence of an autocorrelated indicator is investigated. For this, the price data of the different currencies are compared to the currencies themselves with using autocorrelation. The subfigures provide information about a subset of the day data assets and show how the correlation changes when the signal of the same signal is shifted from 10 days in the past to 10 days in advance. In fact, no lead-lag relationship can be observed and all currencies show a high degree of correlation with past and future signals and indicate a default uncorrelated price development. The only valubale observation of this method is the higher correlation values of Eos, Monero, Ethereum, and Neo while Qtum shows a faster correlation change over time.

The last comparison in this chapter shows shifted correlation plots in a 3D graph in figure 6.5 and A.16 where a correlation of high-resolution data with different time buckets is investigated. In the first figure 6.6, Bitcoin is compared with shifted signals from -5 to 5 shifted time buckets of other currencies in a 1-day time bucket resolution. The results show the highest correlation with the unshifted prices in the middle of the figure and a decrease at the edges. While Ethereum's correlation is dropping in an equal way, the correlation of Litecoin is even lower and negative with a shift of 5 days. These results show a normal and unpredictable price development looking at correlation and no clear lead-lag indication can be observed. The comparison of the high-resolution time buckets in figure A.16 gives even less information for further lead-lag analysis because the signal does not change enough in the higher resolution timespans. Equal behavior can be observed in figure 6.6 and A.18. Here, Bitcoin and Monero are compared to the other currencies and show little changes of about 0.05 in correlation at the edges.



(a) Shifted Pearson correlation of Bitcoin trades data, time buckets: 1 day

(b) Shifted Pearson correlation of Bitcoin trades data, time buckets: 1 second

Figure 6.4: Shifted Pearson correlation of Bitcoin trades data

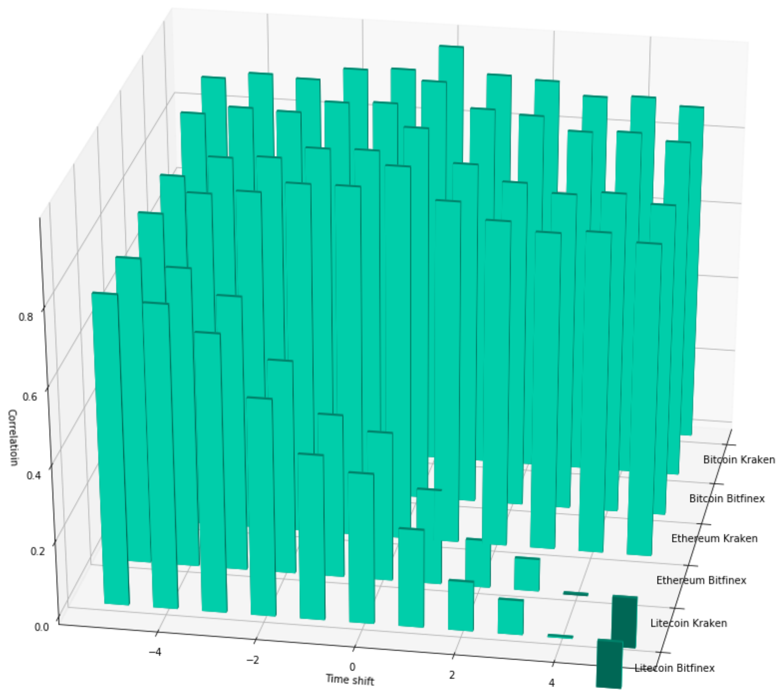


Figure 6.5: Shifted Pearson correlation of Bitcoin vs others, time buckets: 1 day - 3d

6.3 Algorithmic trading

In this chapter, different approaches that use correlation measurements to enhance algorithmic trading decisions are expressed. For a comprehensible investigation, the simulation is first performed with a greedy approach using moving averages. The moving average technique is often used as a simple technical indicator with different window sizes to get a lagged direction of the

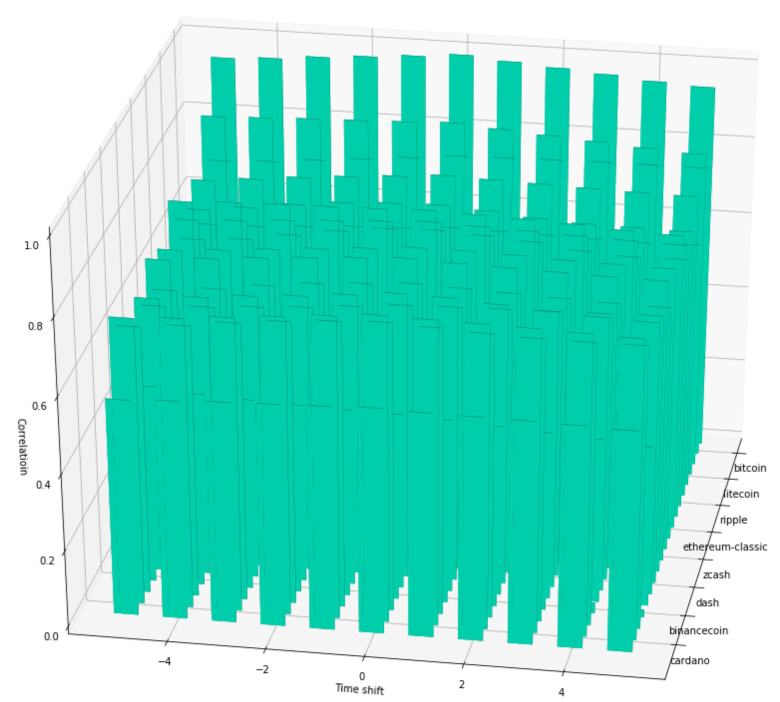
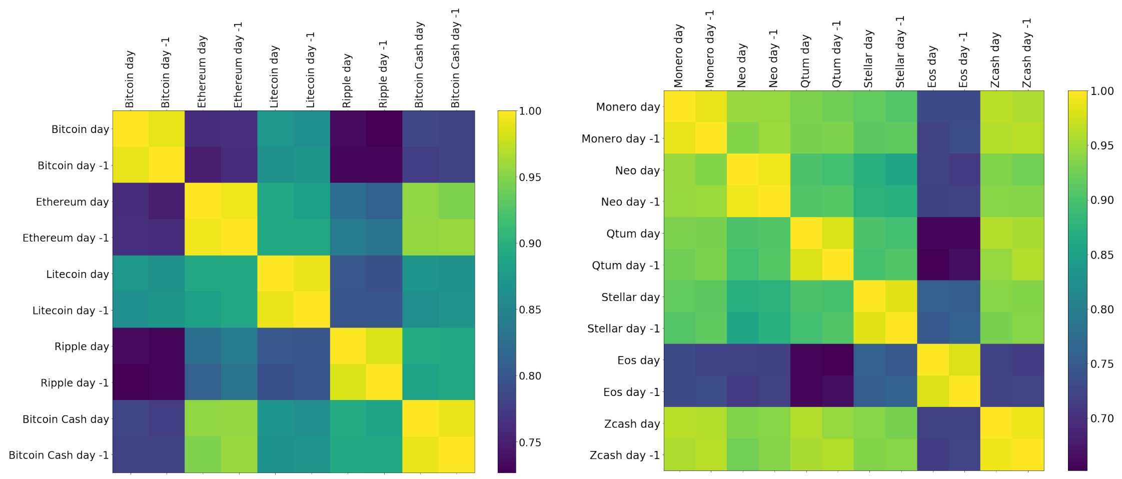


Figure 6.6: Shifted Pearson correlation of day data Bitcoin vs others - 3d



(a) Shifted Pearson correlation of day data selection **(b)** Shifted Pearson correlation of day data selection

Figure 6.7: Shifted Pearson correlation of day data selection

current price. The procedure of this technique is shown in the listings A.1 and A.2 where a moving average approach is implemented. The shown code fragments cover a section of the used algorithm with two different approaches using the python programming language.

The first implementation in listing A.1 makes use of the fast array and listing manipulation bindings to lower-level languages which speed up the performance of the calculations. This technique is used because of the high amount of trades data which has the disadvantage of providing only end results and no timestamp analysis. In contrast, the second implementation in listing A.2 is implemented using a for-loop and thus provides data entries at every point in time. This approach makes it possible to investigate profit development over the whole timespan and analyze the behaviour of ups and downs of the actual asset balance.

To get an impression of the approaches' trading performance, the profit values of the simulation are compared to a buy and hold strategy. For this, a spending of 100 U.S. dollars at the beginning of the available timespan is simulated. To get the actual performance of this strategy, the amount of cryptocurrency is traded back into U.S. dollar with the price available at the needed point in time. Looking at the full timespan of available data, the result of the buy and hold strategy in the first row of table 6.2 represents the actual value in US dollar at the end of the investigated timespan. As this strategy results in a reduction of the investment of more than 71%, a real trader would have made big losses which the following strategies try to prevent.

The first of three strategies tries to use a lead-lag relationship between the two exchanges Kraken and Bitfinex. The results of using the moving average technique with different scaled window sizes and an initial profit amount of 100 U.S. dollar are shown in table 6.1. The focus of this experiment is to investigate the possibilities of a lead-lag relationship between the exchanges and not force a scenario of high-profit values. For this, 1 second time buckets with a window size from the last 100 entries to the last 604800 entries are investigated which represents a timespan of the 7 days. Additionally, a higher time bucket size of 5 seconds with window sizes from the last 100 and 1000 entries are calculated to get a more reasonable comparison of the applied trading technique.

The results in table 6.1 show mostly low values which is due to the applied fees for every trade and the high amount of trades. In fact, this experiment demonstrates only the effect of lead-lag shifted data used as an indicator for trading decisions. The low-profit values do not have an impact on the performance between the shifted data entries and can be compared among themselves. Even the profit values are in similar ranges, it is interesting that the profit value of the -1 shifted Bitcoin Bitfinex data provides the highest amount in all tested scenario while the other measurements are up to a factor of 10 worse. Additionally to the signals of different exchanges, the Bitcoin price data from Kraken is simulated with Ethereum prices from Kraken as an indicator signal in table 6.1. The results are similar to the Bitcoin comparison of Kraken and Bitfinex, whereby every shift of -1 results also in a higher profit than a shift of 0 or 1.

For analyzing day data, the second of three strategies is used in which a lead-lag effect between different assets is searched. In general, all following approaches are implemented using a starting budget of 100 U.S. dollars that represents a 100% budget and the percentage change at the end, and market maker fees (selling-fees) of 0.1% and market taker fees (buying-fees) of 0.2%. In addition, the approaches 01-03 use the price data of Ethereum and Bitcoin-Cash which examined the most promising correlation values in the lead-lag investigation before.

In listing A.2 a greedy approach is used which implements a strategy for buying Ethereum when the current price is higher than the average price of the averaged value of some historical values. In the first run, the length of the history is set to the last 30 entries before the parameter is increased to a length of 100 entries in a second run which has the effect of smoothing volatile timespans. The gained overall profit can be seen in the second row of table 6.2 at approach 01 of the day data. To get a closer look at the performance and possible strengths and weaknesses of this approach, the price development over time is plotted in figure 6.8a and 6.9a. The first comparison to a buy and hold strategy shown in table 6.2 indicates a higher profit of this strategy with both parameters. Nevertheless, this strategy acts only as a simple greedy strategy with is extendend in the next step and the results are only used to get an additional comparison for the following algorithms besides

Trades data source	Signal data source	Time bucket size	Shift	Window size	Profit
Bitcoin Kraken	Bitcoin Bitfinex	1sec	0	100	1.51412509799e-23
Bitcoin Kraken	Bitcoin Bitfinex	1sec	1	100	8.90638892236e-24
Bitcoin Kraken	Bitcoin Bitfinex	1sec	-1	100	2.59568498298e-23
Bitcoin Kraken	Bitcoin Bitfinex	1sec	0	1000	0.0057240704883
Bitcoin Kraken	Bitcoin Bitfinex	1sec	1	1000	0.00448588911875
Bitcoin Kraken	Bitcoin Bitfinex	1sec	-1	1000	0.00702878495739
Bitcoin Kraken	Bitcoin Bitfinex	1sec	0	86400	38.1226817749
Bitcoin Kraken	Bitcoin Bitfinex	1sec	1	86400	35.9162492894
Bitcoin Kraken	Bitcoin Bitfinex	1sec	-1	86400	39.6380769392
Bitcoin Kraken	Bitcoin Bitfinex	1sec	0	259200	81.3486367754
Bitcoin Kraken	Bitcoin Bitfinex	1sec	1	259200	78.035536048
Bitcoin Kraken	Bitcoin Bitfinex	1sec	-1	259200	81.6579989704
Bitcoin Kraken	Bitcoin Bitfinex	1sec	0	604800	114.247955822
Bitcoin Kraken	Bitcoin Bitfinex	1sec	1	604800	110.719328738
Bitcoin Kraken	Bitcoin Bitfinex	1sec	-1	604800	115.679416525
Bitcoin Kraken	Bitcoin Bitfinex	5sec	0	100	0.000530275674393
Bitcoin Kraken	Bitcoin Bitfinex	5sec	1	100	0.000201765518792
Bitcoin Kraken	Bitcoin Bitfinex	5sec	-1	100	0.00176582346424
Bitcoin Kraken	Bitcoin Bitfinex	5sec	0	1000	4.10289131621
Bitcoin Kraken	Bitcoin Bitfinex	5sec	1	1000	2.76956538868
Bitcoin Kraken	Bitcoin Bitfinex	5sec	-1	1000	6.44631023846
Bitcoin Kraken	Ethereum Kraken	1sec	0	100	4.04780526835e-16
Bitcoin Kraken	Ethereum Kraken	1sec	1	100	3.04071833099e-16
Bitcoin Kraken	Ethereum Kraken	1sec	-1	100	5.96365338283e-16
Bitcoin Kraken	Ethereum Kraken	1sec	0	1000	0.00924284201197
Bitcoin Kraken	Ethereum Kraken	1sec	1	1000	0.00830657013504
Bitcoin Kraken	Ethereum Kraken	1sec	-1	1000	0.0105524446972
Bitcoin Kraken	Ethereum Kraken	5sec	0	100	0.000210175253561
Bitcoin Kraken	Ethereum Kraken	5sec	1	100	0.000123082790688
Bitcoin Kraken	Ethereum Kraken	5sec	-1	100	0.000497758499934
Bitcoin Kraken	Ethereum Kraken	5sec	0	1000	2.51561453919
Bitcoin Kraken	Ethereum Kraken	5sec	1	1000	2.05832113687
Bitcoin Kraken	Ethereum Kraken	5sec	-1	1000	3.44391896212

Table 6.1: Profit measurements of trades data

the simple buy and hold strategy.

A possible expansion of this strategy is to take into account another signal as a trading indicator. For this, a moving average indicator signal is introduced which replaces the moving average of the current currency. The main procedure of this method is implemented in listing A.3 where the variable *cryptoCurrencyData* is replaced by a variable called *signalData*, which represents the price data of another currency in this case. The stated procedure is used to simulate the profit values over time for a different signal. In figure 6.8b and 6.9b, the Ethereum price was investigated for a lead-lag relationship with Bitcoin-Cash as selected in the lead-lag investigation before, but as table 6.2 shows, the performance of this approach, called approach 02, is worse than executed with the default moving average approach 01. An even worse result is gained with a shifted Bitcoin-Cash signal when compared to the Ethereum price in figure 6.8c and 6.9c. The gathered results are labled as approach 03 and documented in table 6.2 to provide a comparison to the other approaches. Looking at these values, a loss of 26% can be determined at the 30 days moving-average column, but even if this parameter leads to a bad investment strategy in this case, the 100 days moving average shows a higher end-profit than approach 02. This means the performance of the trading approaches reflects the measurements from the lead-lag investigation which exhibits no clear indicator signals.

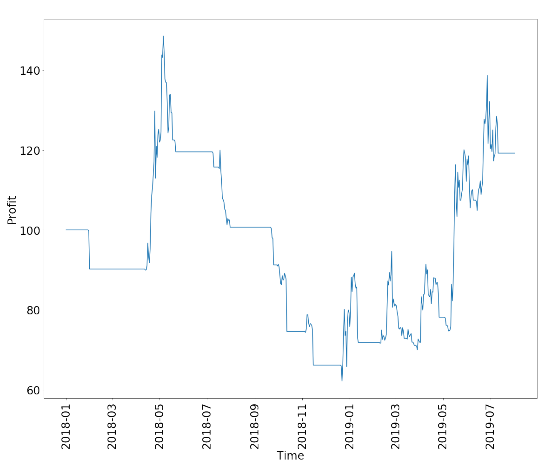
An extended version of the last approach can be seen in listing A.4 where all selected currencies of the day data are taken into account for a collective amount of profit. For this, every currency got a reserved and independent budget which equally sums up to an investment of 100 U.S. dollars for trading it separately without indicators of another currency as in approach 03. The performance of this implementation is listed as approach 04 in table 6.2 and is used as a greedy solution for a comparison to the next approach. The regarding price development plot of this approach is visualized in figures 6.8d and 6.9d where a constant price drop from November 2018 to February 2019 can be observed in both charts. In fact, the volatilities and losses of the investment of the 100-day moving average are less and result in higher overall values especially at the end of the simulation.

To extend approach 04, approach 05 tries to find a lead-lag relationship of every possible pair of assets which also results in a longer algorithm shown in listing A.5. For this purpose, the correlation value of the shifted signal is calculated and if the shifted correlation is higher than the default correlation a buy action is executed. The performance values of this approach in table 6.1 show similar values than the greedy approach 04. The main difference can be discovered in the simulation history plot in figure 6.9e, in which a high price change in February 2018 can be observed.

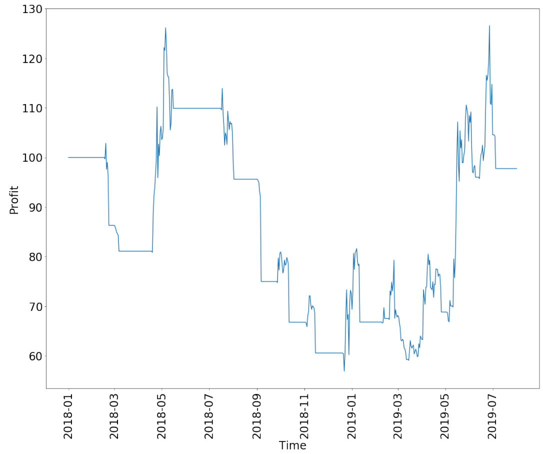
The last of the three approaches try to enhance the overall profit with the method of adding a percentage of the amount which is available to buy, dependent on the current correlation of an indicator signal. As only one signal is requeired for this approach, the correlation of each signal to the Bitcoin price is used as this seems to be the dominating factor in the crypto market. As shown in listing A.6, all selected currencies are iterated and the simulation process is applied in sequential order. A closer look of the results and a comparison to the simulation of the other approaches is printed in the summarized table in 6.2.

Approach	**30 days moving average**	**100 days moving average**
Buy and hold	28.08	28.08
01	119.21	138.34
02	97.75	136.33
03	73.93	149.32
04	121.94	128.62
05	109.90	117.99
06	101.11	133.96

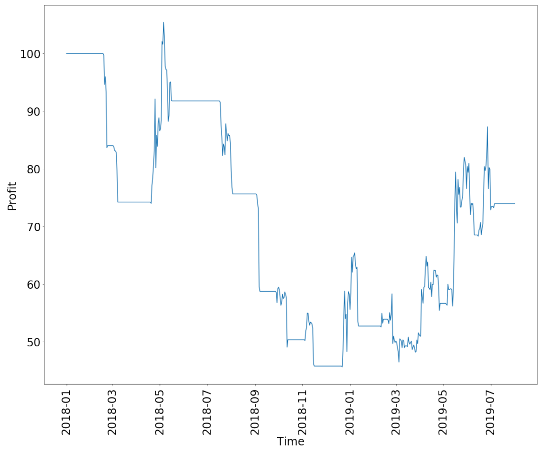
Table 6.2: Profit measurements of day data



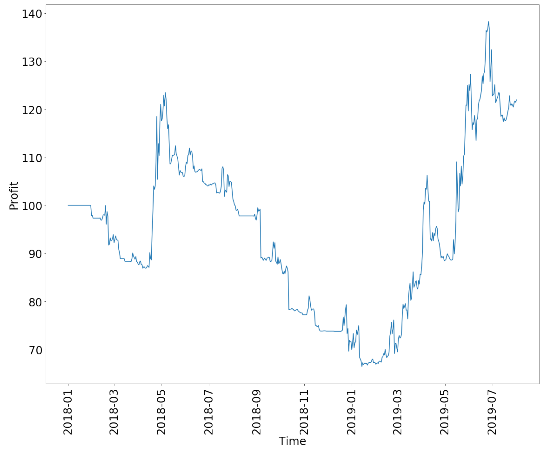
(a) Results of greedy approach 01 applied on day data with 30 days moving average



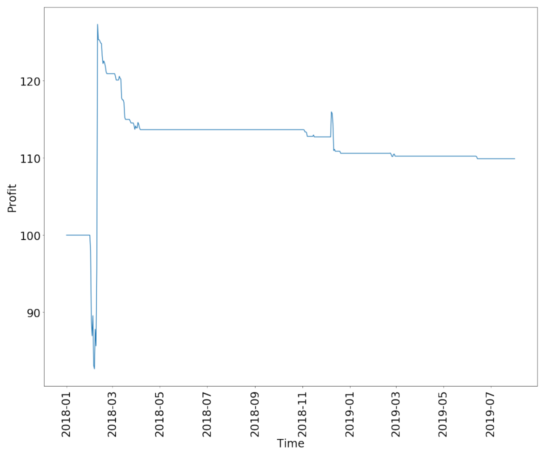
(b) Results of greedy approach 02 applied on day data with taking another currency as trading indicator and 30 days moving average



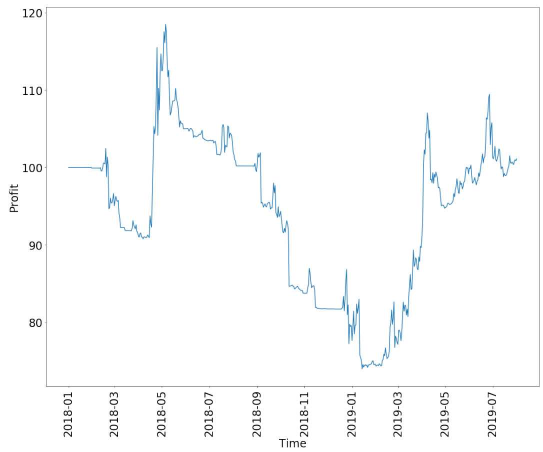
(c) Results of greedy approach 03 with taking another, shifted currency as a trading indicator and 30 days moving average



(d) Results of greedy approach 04 with taking into account all currencies with 30 days moving average

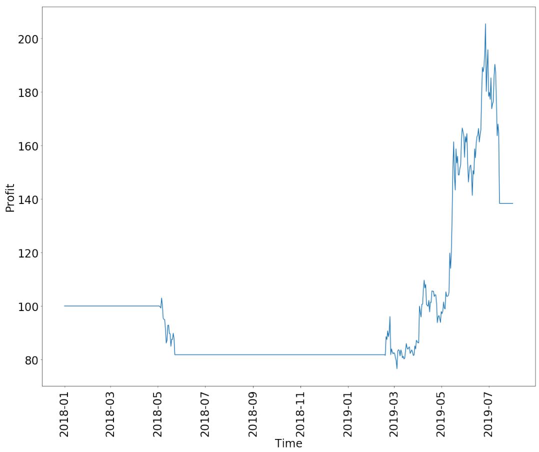


(e) Results of greedy approach 05 with taking into account all possible lead-lag combinations with 30 days moving average

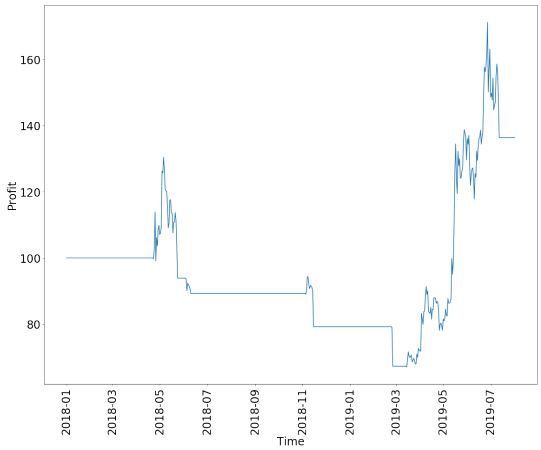


(f) Results of greedy approach 06 buying percentage part of the correlation to Bitcoin with 30 days moving average

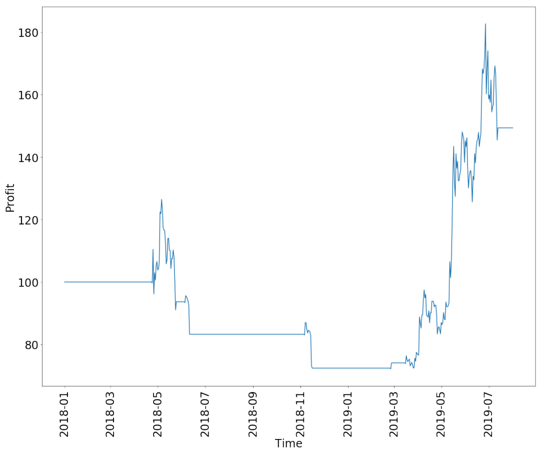
Figure 6.8: Results of the profit measurements of different algorithmic approaches with 30 days moving average



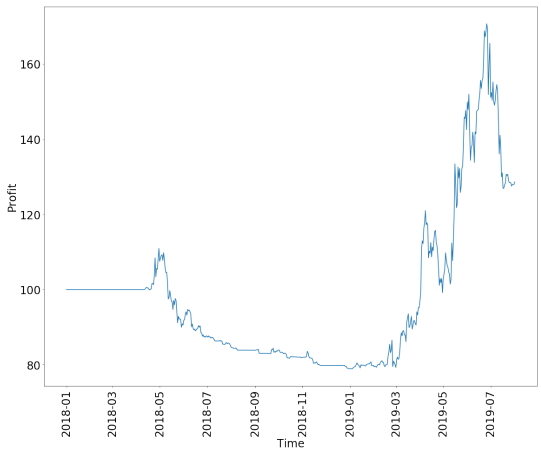
(a) Results of a greedy approach 01 with 100 days moving average



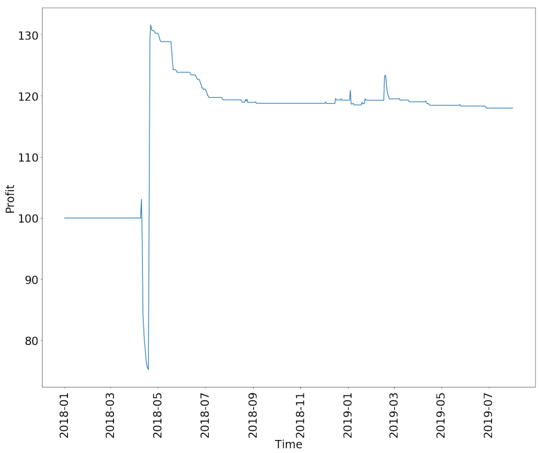
(b) Results of a greedy approach 02 with taking another currency a trading indicator with 100 days moving average



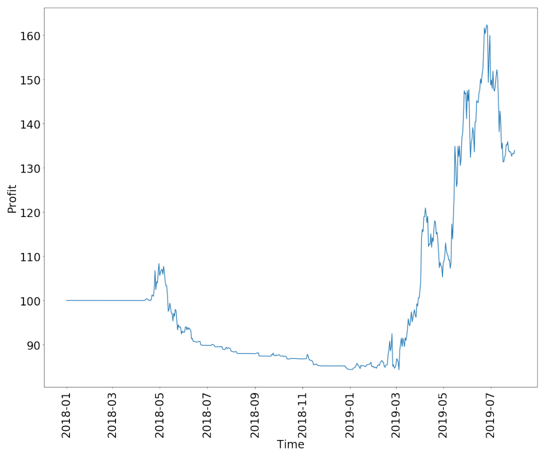
(c) Results of a greedy approach 03 with taking another, shifted currency a trading indicator with 100 days moving average



(d) Results of a greedy approach 04 with taking into account all currencies with 100 days moving average



(e) Results of a greedy approach 05 with taking into account all possible lead-lag combinations with 100 days moving average



(f) Results of a greedy approach 06 buying percentage part of the correlation to Bitcoin with 100 days moving average

Figure 6.9: Results of the profit measurements of different algorithmic approaches with 100 days moving average

7 Conclusion

The gathered results give a first impression about the correlation behavior of the crypto market and show algorithmic trading approaches for applying correlation values in real world scenarios. In a first investigation, different measurements show a high correlation between the selected exchanges and assets which implies a strong relationship of the crypto market. The direct dependency of exchanges is attributable to the undesired opportunity for traders to take advantage of arbitrage trading. A divergence of exchange prices would lead to the possibility of switching to another exchange with the high transfer speed of cryptocurrencies instead of waiting for bank transfers. As this case is bad for the reputation of exchanges with lagged prices, countermeasures are taken so it is not possible to benefit from this scenario.

Beside arbitrage trading, the correlation measurements of different cryptocurrencies show high correlated price developments of 0.8 of Pearson's correlation and above. Especially small groups of altcoins like Qtum, Neo and Dash show high correlation values when investigated on a daily price resolution. Also popular crypto assets like Bitcoin, Ethereum and Litecoin show strong correlation values of 0.7 and above for small time buckets of 1 second up to 1 day time buckets. The uncertain and restricted usage of cryptocurrencies forms a dependent market with a dominant ratio of Bitcoin while big price drops and hikes of single assets are rare and only occur in special cases.

Even though most of the assets show high correlation values, it is not possible to observe lead-lag effects for crypto markets as [29] found for stock markets. In fact, this circumstance is attributable to the missing closing times and the international availability of online exchanges which update price developments instantly and prevent the abuse of arbitrage or lead-lag indication so even high correlated cryptocurrencies provide no lead-lag effect and change their prices instantly. Additionally, the correlation values show a slight divergence from time to time so a stable lead-lag indicator for future price developments would have low relevance anyway. This thesis give a continuation of the early and basic analysis in [13] and refute the opportunity of predicting crypto assets with other crypto assets.

Despite the poor view of measurements, algorithmic trading approaches that make use of the gathered results are presented. A comparison of the performance charts gives no clear results if the algorithms result in more or less profit applied to correlated data. A well performing algorithm like in the preceding experiment [11] which only uses price data instead of social signals can not be found with the presented approaches. At the end, it looks like the decisive factor is the parameterized moving average technique which makes the difference, and further comparisons of different technical analysis methods must be investigated. A promising application of correlation based trading is portfolio management and trading decisions based on correlation ratios to spread risk and cohesion of the market which is already used in similar ways for asset management of stock markets.

To get further considerations about relevance and profitability of correlation based trading, investigations of performance measurements on other markets are necessary. An example for this is to transform experiments like [23] that investigate the influence of correlation to other financial products like Swaps. To propose a more precise hypothesis, additional technical indicators and parameters must be taken into account and compared to the presented indicators and approaches.

Bibliography

References

[1] Suni Ajaykumar. „A regression model for crypto-currency price“. 2018. URL: https://www.researchgate.net/publication/325403172_A_regression_model_for_crypto-currency_price (visited on 10/13/2019).

[2] Elli Androulaki et al. „Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains“. 2018. URL: https://arxiv.org/abs/1801.10228 (visited on 10/13/2019).

[3] Susan Athey et al. „Bitcoin Pricing, Adoption, and Usage: Theory and Evidence“. In: Stanford University Graduate School of Business Research Paper No. 16-42, 2016. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2826674 (visited on 06/21/2018).

[4] António Cerquerira. „Autocorrelation in Daily Stock Returns“. In: Porto: PFN – Portuguese Finance Network, 2006. URL: https://sigarra.up.pt/faup/en/PUB_GERAL.PUB_VIEW?pi_pub_base_id=50372 (visited on 04/22/2018).

[5] Kalok Chan. „A Further Analysis of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market“. In: Arizona State University: The Review of Financial Studies Vol. 5, No. 1, 1992, pp. 123–152. URL: https://www.jstor.org/stable/2962016 (visited on 10/13/2019).

[6] Shiu-Sheng Chen. „Revisiting the empirical linkages between stock returns and trading volume“. In: Stanford University Graduate School of Business Research Paper No. 16-42, 2012. URL: https://www.sciencedirect.com/science/article/pii/S0378426612000453 (visited on 07/09/2019).

[7] Pavel Ciaian, Miroslava Rajcaniova, and d’Artis Kancs. „The economics of BitCoin price formation“. In: Applied Economics, Vol. 48, No. 19, 2016, 1799–1815. URL: https://doi.org/10.1080/00036846.2015.1109038 (visited on 10/13/2019).

[8] Frank De Jong and Theo Nijman. „High frequency analysis of lead-lag relationships between financial markets“. In: Journal of Empirical Finance, Vol. 4, Issues 2-3, 1997, pp. 259–277. URL: https://www.sciencedirect.com/science/article/pii/S0927539897000091 (visited on 10/13/2019).

[9] Wen-Qi Duan and H. Eugene Stanley. „Cross-correlation and the predictability of financial return series“. In: Physica A: Statistical Mechanics and its Applications, Vol. 390 Issue 2, 2011, 290–296. URL: https://www.sciencedirect.com/science/article/pii/S0378437110007934 (visited on 10/13/2019).

[10] Alex Frino and Andrew West. „The Lead–Lag Relationship Between Stock Indices and Stock Index Futures Contracts: Further Australian Evidence“. In: Abacus, Vol. 35, No. 3, 1999. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6281.00049 (visited on 10/13/2019).

[11] David Garcia and Frank Schweitzer. „Social signals and algorithmic trading of Bitcoin“. In: R. Soc. open sci. 2: 150288, 2015. URL: http://dx.doi.org/10.1098/rsos.150288 (visited on 10/13/2019).

[12] Marian Gidea et al. „Topological recognition of critical transitions in time series of cryptocurrencies“. 2019. URL: https://arxiv.org/abs/1809.00695 (visited on 10/13/2019).

[13] Dave Hagemann. „A Time Series Analysis of Crypto Currency Price Data“. In: Unergraduate Research Followship Working Paper Series, 2018. URL: https://coles.kennesaw.edu/econopp/docs/Hagemann-Paper-WEBSITE.pdf (visited on 10/13/2019).

[14] Jan Hauke and Tomasz Kossowski. „Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data“. In: Quaestiones Geographicae, Vol. 30: Issue 2, 2011. URL: https://doi.org/10.2478/v10117-011-0021-1 (visited on 10/13/2019).

[15] Terrence Hendershott, Charles M. Jones, and Albert J. Menkveld. „Does Algorithmic Trading Improve Liquidity?“ In: Journal of Finance, Vol. 66, 2011, pp. 1–33. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1100635 (visited on 10/13/2019).

[16] Sunny King and Scott Nadal. „PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake“. 2012. URL: https://www.semanticscholar.org/paper/PPCoin%3A-Peer-to-Peer-Crypto-Currency-with-King-Nadal/0db38d32069f3341d34c35085dc009a85ba13c13 (visited on 10/13/2019).

[17] Tony Klein, Hien Pham Thu, and Thomas Walther. „Bitcoin is not the New Gold – A Comparison of Volatility, Correlation, and Portfolio Performance“. In: International Review of Financial Analysis, Vol. 59, 2018, pp. 105–116. URL: https://www.sciencedirect.com/science/article/abs/pii/S105752191830187X (visited on 10/13/2019).

[18] Ladislav Kristoufek. „Power-law correlations in finance-related Google searches, and their cross-correlations with volatility and traded volume: Evidence from the Dow Jones Industrial components“. In: Physica A: Statistical Mechanics and its Applications, Vol. 428, 2014, pp. 194–205. URL: https://www.sciencedirect.com/science/article/pii/S0378437115001636 (visited on 10/13/2019).

[19] Ladislav Kristoufek. „What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis“. In: PLoS ONE 10(4): e0123923, 2015. URL: https://doi.org/10.1371/journal.pone.0123923 (visited on 10/13/2019).

[20] Igor Makarov and Antoinette Schoar. „Trading and arbitrage in cryptocurrency markets“. In: Journal of Financial Economics, 2019. URL: https://doi.org/10.1016/j.jfineco.2019.07.001 (visited on 10/13/2019).

[21] Ben R. Marshall, Rochester H. Cahan, and Jared M. Cahan. „Does intraday technical analysis in the U.S. equity market have value?“ In: New Zealand: Journal of Empirical Finance, 2008, pp. 199–210. URL: https://www.sciencedirect.com/science/article/pii/S0927539807000588#sec4 (visited on 06/20/2018).

[22] Robert Matthews. „Storks Deliver Babies (p = 0.008)“. In: Teaching Statistics, Vol. 22, 2000, pp. 36–38. URL: https://doi.org/10.1111/1467-9639.00013 (visited on 10/13/2019).

[23] Gunter Meissner. „Correlation Trading Strategies – Opportunities and Limitations“. In: The Journal of Trading Fall 2016, 11 (4), 2016, pp. 14–32. URL: https://doi.org/10.3905/jot.2016.11.4.014 (visited on 10/13/2019).

[24] Andrew Miller et al. „Permacoin: Repurposing Bitcoin Work for Data Preservation“. In: 2014 IEEE Symposium on Security and Privacy, 2014, pp. 475–490. URL: https://ieeexplore.ieee.org/document/6956582 (visited on 10/13/2019).

[25] Satoshi Nakamoto. „Bitcoin: A Peer-to-Peer Electronic Cash System“. 2008. URL: https://bitcoin.org/bitcoin.pdf (visited on 10/13/2019).

[26] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. „Ransomware payments in the Bitcoin ecosystem“. In: Journal of Cybersecurity, Vol. 5, Issue 1, 2019, 1–11. URL: https://doi.org/10.1093/cybsec/tyz003 (visited on 10/13/2019).

[27] Serguei Popov. „The Tangle“. 2018. URL: https://assets.ctfassets.net/r1dr6vzfxhev/2t4uxvsIqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf (visited on 10/13/2019).

[28] K. Rehfeld et al. „Comparison of correlation analysis techniques for irregularly sampled time series“. In: Nonlin. Processes Geophys. 18, 2011, pp. 389–404. URL: https://doi.org/10.5194/npg-18-389-2011 (visited on 04/22/2018).

[29] Leonidas Sandoval Jr. „To lag or not to lag? How to compare indices of stock markets that operate on different times“. In: Physica A: Statistical Mechanics and its Applications, Vol. 403, 2013, pp. 227–243. URL: https://doi.org/10.1016/j.physa.2014.02.039 (visited on 10/13/2019).

[30] „Tether: Fiat currencies on the Bitcoin blockchain“. 2016. URL: https://tether.to/wp-content/uploads/2016/06/TetherWhitePaper.pdf (visited on 10/13/2019).

[31] Ruey S. Tsay. *Analysis of Financial Time Series*. 3rd ed. New Jersey: John Wiley & Sons, Wiley Series in Probability and Statistics, 2010. ISBN: 9780470414354. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470644560 (visited on 10/13/2019).

[32] P. Whittle. „The Analysis of Multiple Stationary Time Series“. In: 1st ed. Journal of the Royal Statistical Society. Series B (Methodological) Vol. 15 No. 1, 1953. Chap. 15, pp. 125–139. URL: http://www.jstor.org/stable/2983728.

[33] Gavin Wood. „Ethereum: A secure decentralised generalised transaction ledger“. 2017. URL: https://gavwood.com/paper.pdf.

[34] Jennifer J. Xu. „Are blockchains immune to all malicious attacks?“ In: Financial Innovation, Vol. 2, 2016. URL: https://www.researchgate.net/publication/311568850_Are_blockchains_immune_to_all_malicious_attacks (visited on 10/13/2019).

A Appendix

Figure A.1: Cryptocurrency summary from Coingecko (https://www.coingecko.com, visited on 03.09.2019 at 15:40)

Kryptowährungen: 5589 Börsen: 370 Marktkapitalisierung: 270.928.963.557 $ 6,4% 24-Stunden-Volumen: 61.778.425.681 $ Dominanz: BTC **70%** ETH **7,1%** XRP **4,2%**

Figure A.2: Cryptocurrency summary from CoinMarketCap (https://coinmarketcap.com, visited on 03.09.2019 at 15:40)

Cryptocurrencies: 2579 • **Markets:** 20454 • **Market Cap:** $269.926.428.772 • **24h Vol:** $62.450.932.549 • **BTC Dominance:** 70.5%

Figure A.3: Chart showing high volatility of Bitcoin (https://coinmarketcap.com, visited on 03.09.2019 at 15:40)

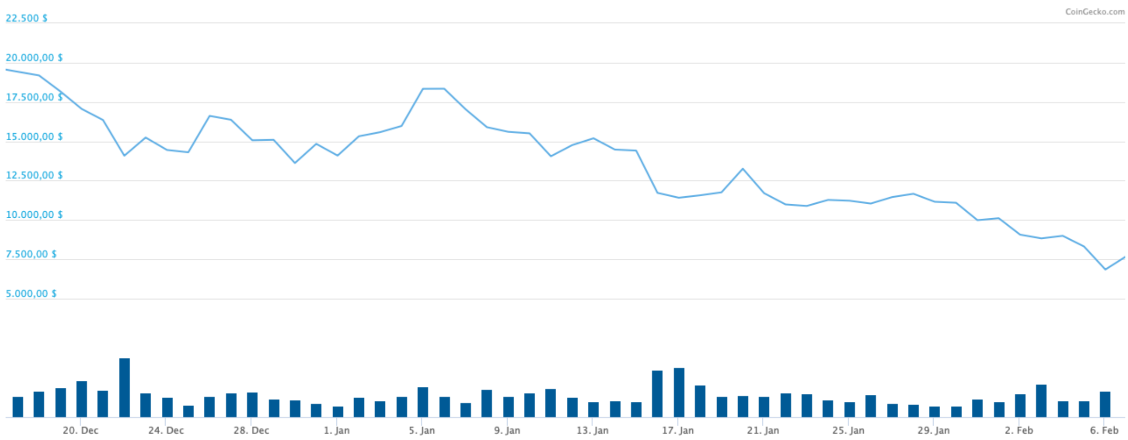


Figure A.4: Chart showing the similarity of the cryptocurrency prices (https://www.coingecko.com, visited on 12.09.2019 at 10:30)

☆	2	Ethereum	ETH	177,73 $	0.4%	-0.5%	1.7%	7.604.370.084 $	107.708.236	19.154.881.739 $	
☆	3	XRP	XRP	0,252726 $	0.0%	-1.5%	-2.4%	1.269.589.664 $	43.024.433.511	10.873.207.380 $	
☆	4	Bitcoin Cash	BCH	296,41 $	0.4%	-1.4%	0.9%	1.631.353.699 $	17.998.689	5.334.936.584 $	
☆	5	Litecoin	LTC	68,77 $	0.3%	-2.5%	2.6%	2.565.785.108 $	63.237.191	4.345.968.703 $	
☆	6	EOS	EOS	3,71 $	0.6%	-0.6%	12.3%	2.329.849.517 $	1.026.112.264	3.804.055.717 $	

Figure A.5: Exchanges and provided asset pairs sorted by 24h trading volume (https://www.coingecko.com, visited on 19.09.2019 at 18:00)

#	Exchange	Pair	Price	Spread	+2% Depth	-2% Depth	24h Volume	Volume %	Last Traded	Trust Score
*	eToro Sponsored									
*	Crypto.com Sponsored									
1	Binance	BTC/USDT	$9,855.82	0.01%	$2,074,459	$2,482,560	$328,098,778	1.57%	Recently	
3	Coinbase Pro	BTC/USD	$9,869.29	0.01%	$1,320,504	$940,042	$117,639,100	0.56%	Recently	
8	Bitfinex	BTC/USD	$9,888.34	0.0%	$127,198	$273,240	$67,392,575	0.32%	Recently	
11	Upbit	BTC/KRW	$9,855.16	0.02%	$95,835	$84,364	$56,589,092	0.27%	Recently	
9	Binance	ETH/BTC	$9,863.99	0.03%	$418,505	$501,348	$54,587,100	0.26%	Recently	
2	Bitstamp	BTC/USD	$9,871.80	0.05%	$4,918,361	$3,904,756	$54,005,086	0.26%	Recently	
5	Kraken	XBT/EUR	$9,894.02	0.0%	$1,508,704	$1,867,643	$47,219,517	0.23%	Recently	
4	Kraken	XBT/USD	$9,871.76	0.01%	$2,365,873	$2,356,988	$47,545,737	0.23%	Recently	
17	Binance	XRP/BTC	$9,863.99	0.07%	$180,482	$290,709	$39,102,012	0.19%	Recently	
6	bitFlyer	BTC/JPY	$9,877.80	0.03%	$1,620,120	$1,690,129	$33,501,117	0.16%	Recently	

Figure A.6: Exchanges sorted by Coingecko's popularity value (https://www.coingecko.com, visited on 19.09.2019 at 18:00)

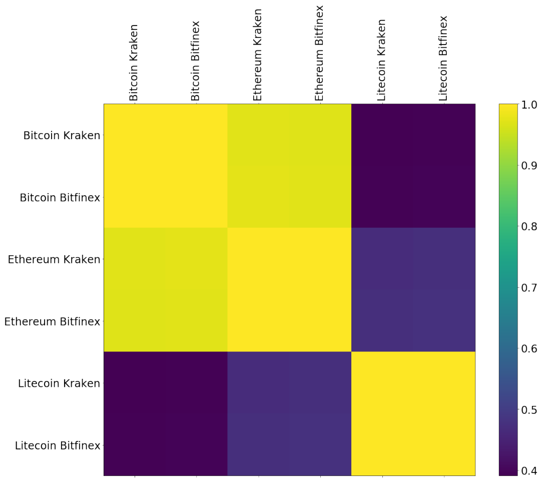
#	Exchange	Trust Score Beta	24h Volume (Normalized)	24h Volume	Visits (SimilarWeb)	# Coins	# Pairs	Last 7 Days
1	Binance Centralized	10	$1,047,738,196	$1,349,084,503	20,530,000	168	573	
2	Bitfinex Centralized	10	$150,815,906	$150,815,906	3,063,000	34	106	
3	KuCoin Centralized	10	$85,804,860	$85,804,860	2,332,000	209	462	
4	Bittrex Centralized	10	$38,106,707	$38,106,707	3,325,000	250	364	
5	Poloniex Centralized	10	$30,892,356	$30,892,356	2,513,000	56	101	
6	Coinbase Pro Centralized	9	$291,764,212	$303,321,990	5,717,000	22	53	
7	Upbit Centralized	9	$261,874,942	$261,874,942	6,315,000	204	338	
8	Huobi Global Centralized	9	$225,514,474	$953,857,132	4,439,428	224	538	
9	Kraken Centralized	9	$165,275,322	$170,059,477	3,248,000	22	87	
10	Bitstamp Centralized	9	$103,855,199	$115,462,617	2,035,000	5	14	

Figure A.7: Exchanges sorted by normalized 24h trading volume (https://www.coingecko.com, visited on 19.09.2019 at 18:00)

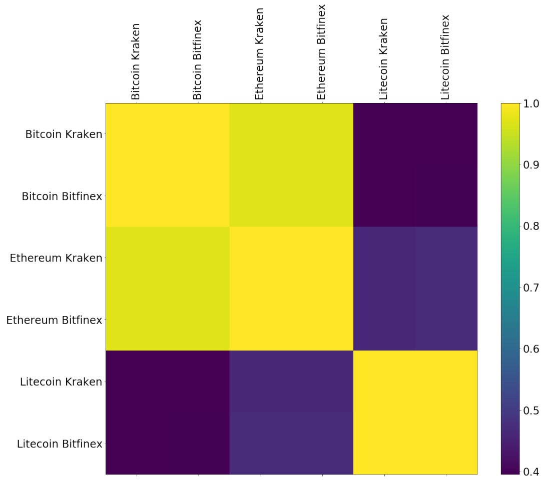
#	Exchange	Trust Score Beta	24h Volume (Normalized)	24h Volume	Visits (SimilarWeb)	# Coins	# Pairs	Last 7 Days
1	Binance Centralized	10	$1,047,738,196	$1,349,084,503	20,530,000	168	573	
6	Coinbase Pro Centralized	9	$291,764,212	$303,321,990	5,717,000	22	53	
7	Upbit Centralized	9	$261,874,942	$261,874,942	6,315,000	204	338	
8	Huobi Global Centralized	9	$225,514,474	$953,857,132	4,439,428	224	538	
9	Kraken Centralized	9	$165,275,322	$170,059,477	3,248,000	22	87	
2	Bitfinex Centralized	10	$150,815,906	$150,815,906	3,063,000	34	106	
10	Bitstamp Centralized	9	$103,855,199	$115,462,617	2,035,000	5	14	
3	KuCoin Centralized	10	$85,804,860	$85,804,860	2,332,000	209	462	
17	HitBTC Centralized	8	$82,369,676	$667,614,450	1,614,000	366	882	
55	CoinTiger Centralized	6	$80,701,812	$950,806,626	1,588,678	112	181	

Figure A.8: Cryptocurrencies sorted by 24h trading volume (https://www.coingecko.com, visited on 19.09.2019 at 18:10)

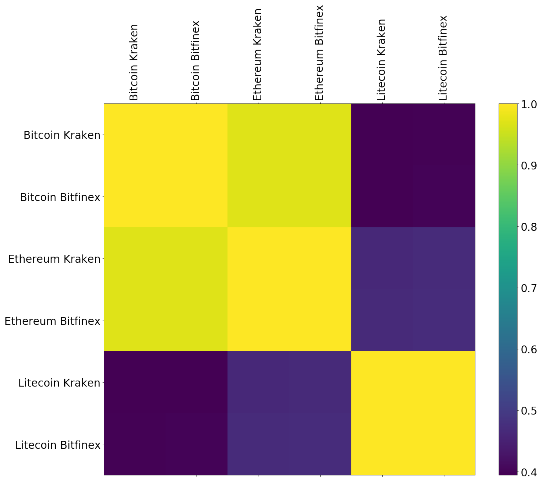
#	Coin		Price	1h	24h	7d	24h Volume	Circulating Supply	Mkt Cap	Last 7 Days
7	Tether	USDT	$1.00	0.3%	0.5%	0.3%	$24,940,300,260	3,561,089,500	$3,584,428,155	
1	Bitcoin	BTC	$9,991.11	0.9%	-1.4%	-1.8%	$21,547,223,412	17,945,387	$179,294,317,564	
2	Ethereum	ETH	$218.25	2.8%	2.6%	20.7%	$9,700,723,451	107,807,936	$23,495,240,695	
5	Litecoin	LTC	$75.62	1.5%	-3.8%	8.1%	$3,953,739,514	63,290,029	$4,789,160,810	
6	EOS	EOS	$3.93	1.6%	-4.7%	4.5%	$3,105,640,361	1,027,130,578	$4,035,129,287	
3	XRP	XRP	$0.300742	1.2%	-3.9%	16.6%	$2,360,389,306	43,055,012,634	$12,872,638,423	
4	Bitcoin Cash	BCH	$316.14	1.8%	-2.6%	5.7%	$2,102,121,897	18,012,052	$5,706,190,796	
13	TRON	TRX	$0.01740575	2.8%	-3.4%	16.6%	$846,315,692	66,140,232,427	$1,152,722,463	
21	Ethereum Classic	ETC	$6.15	1.9%	-4.6%	-2.7%	$726,592,216	113,597,861	$701,085,184	
10	Stellar	XLM	$0.081557	-0.4%	5.3%	35.7%	$574,572,533	20,084,644,643	$1,632,698,323	



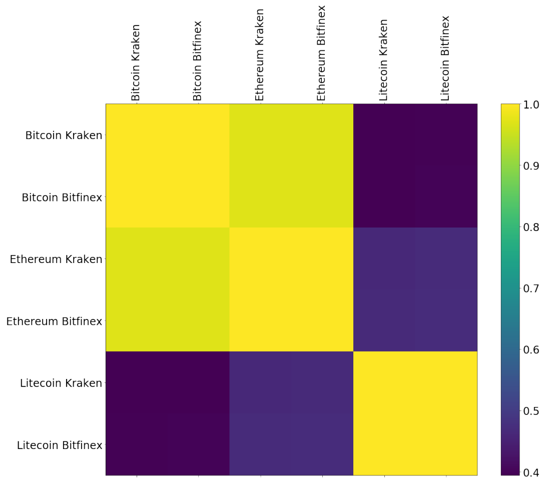
(a) Pearson correlation of trades data, 1 day time buckets



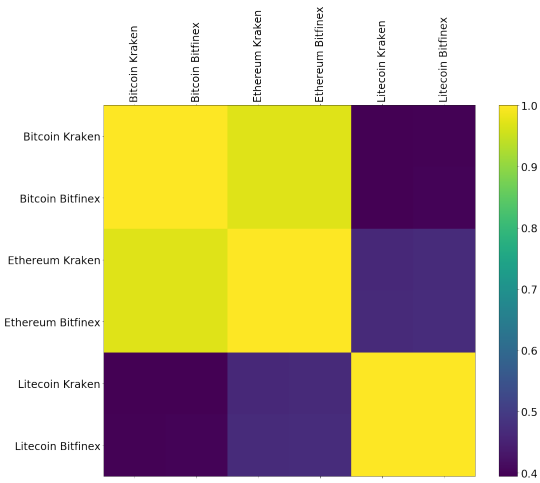
(b) Pearson correlation of trades data, 1 hour time buckets



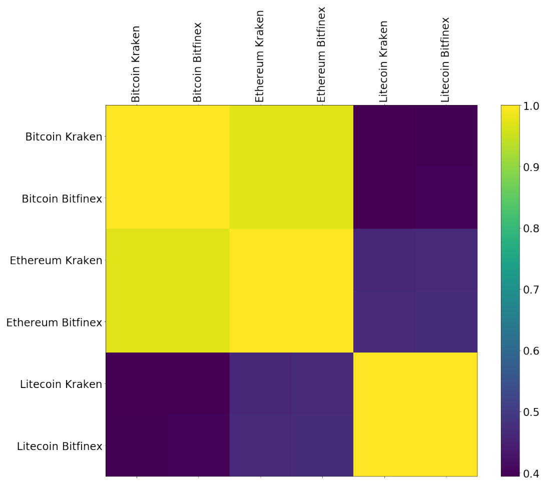
(c) Pearson correlation of trades data, 1 minute time buckets



(d) Pearson correlation of trades data, 10 seconds time buckets

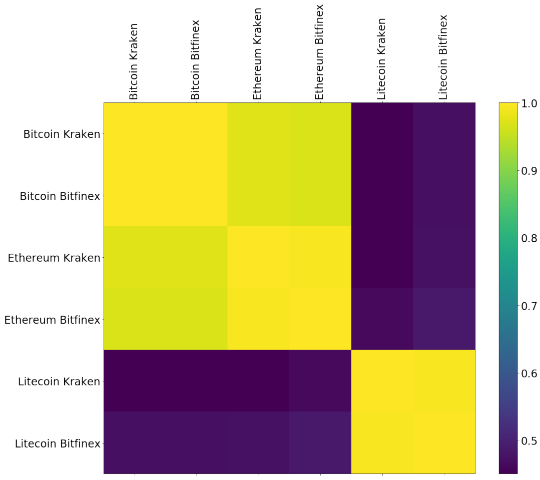


(e) Pearson correlation of trades data, 5 seconds time buckets

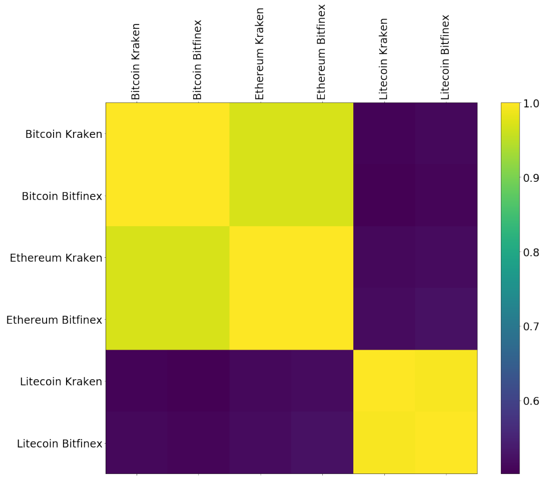


(f) Pearson correlation of trades data, 1 second time buckets

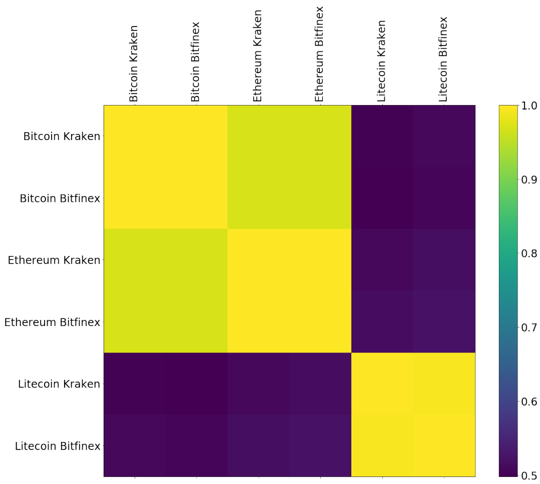
Figure A.9: Pearson correlation of trades data, different time buckets



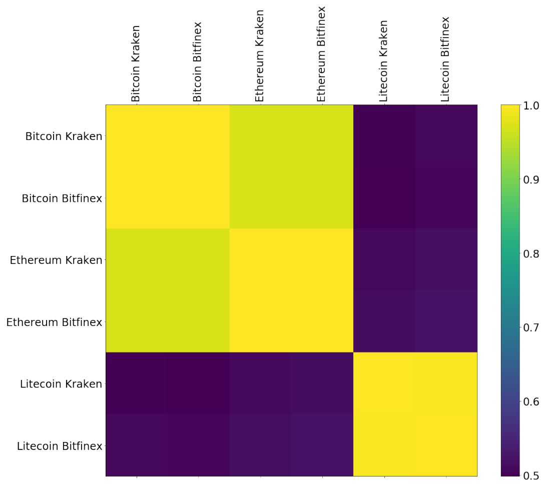
(a) Spearman correlation of trades data, 1 day time buckets



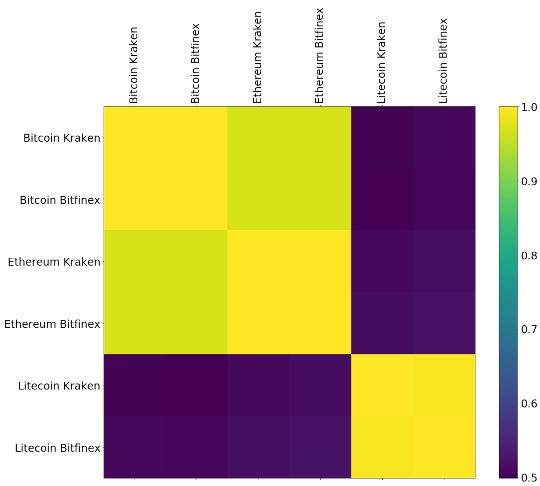
(b) Spearman correlation of trades data, 1 hour time buckets



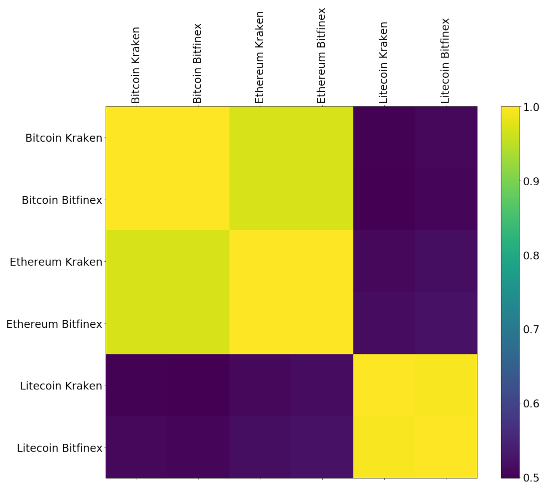
(c) Spearman correlation of trades data, 1 minute time buckets



(d) Spearman correlation of trades data, 10 seconds time buckets

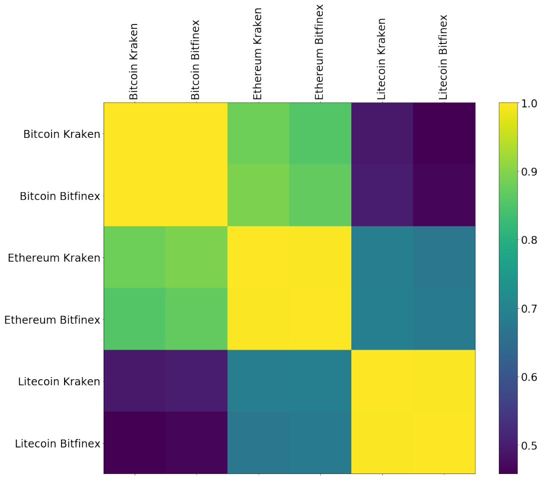


(e) Spearman correlation of trades data, 5 seconds time buckets

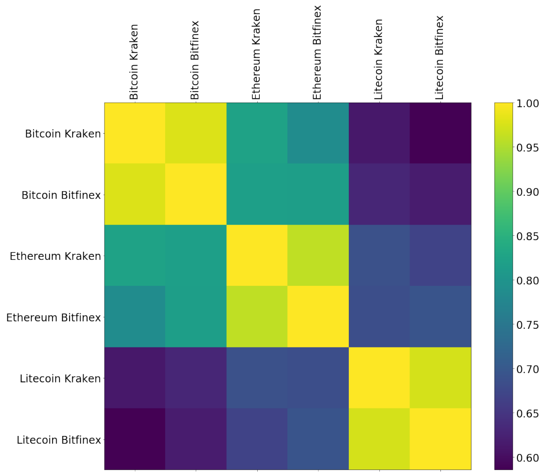


(f) Spearman correlation of trades data, 1 second time buckets

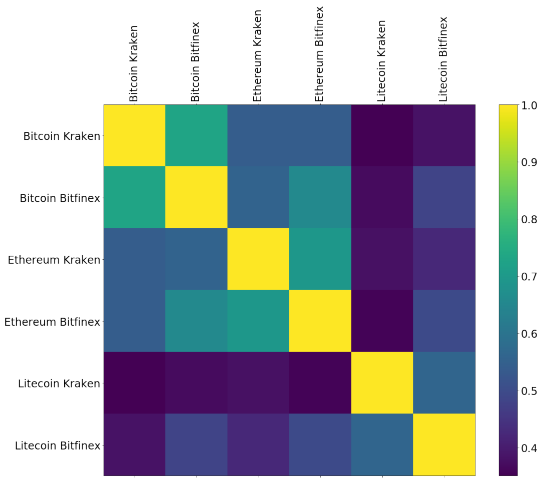
Figure A.10: Spearman correlation of trades data, different time buckets



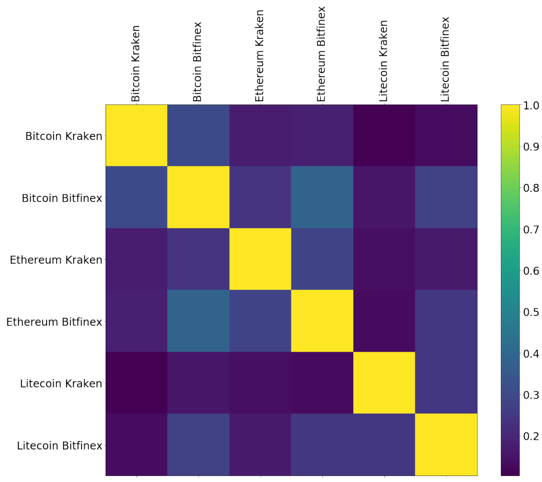
(a) Profit Pearson correlation of trades data, 1 day time buckets



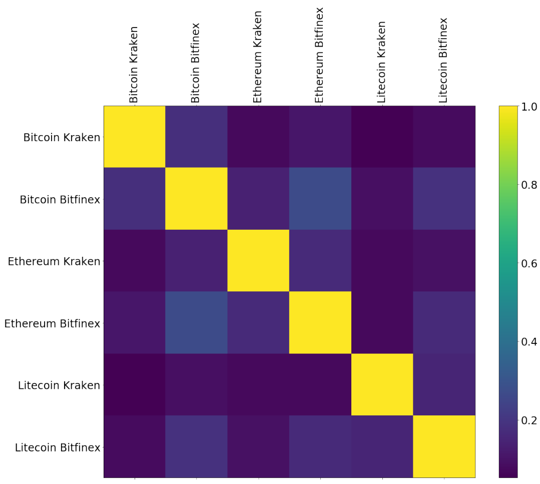
(b) Profit Pearson correlation of trades data, 1 hour time buckets



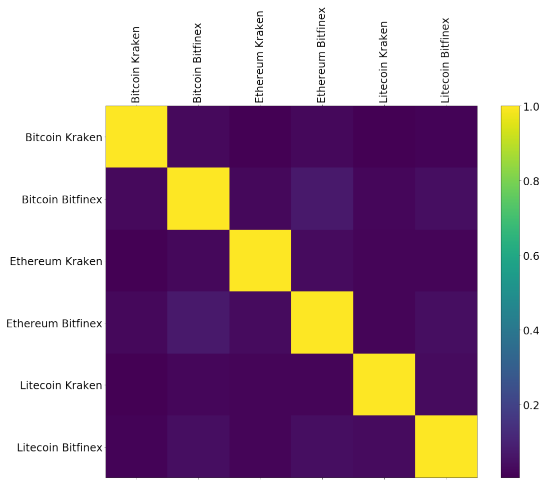
(c) Profit Pearson correlation of trades data, 1 minute time buckets



(d) Profit Pearson correlation of trades data, 10 seconds time buckets

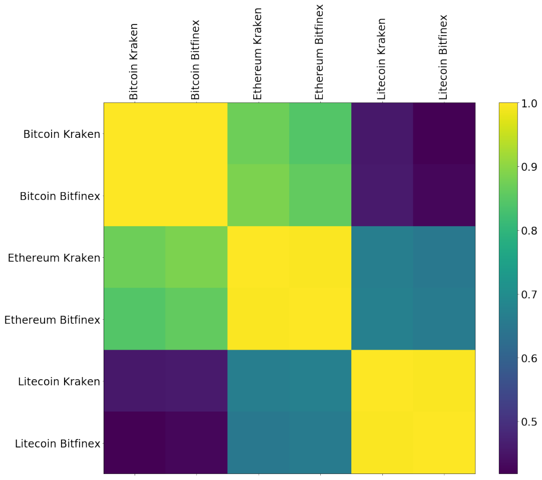


(e) Profit Pearson correlation of trades data, 5 seconds time buckets

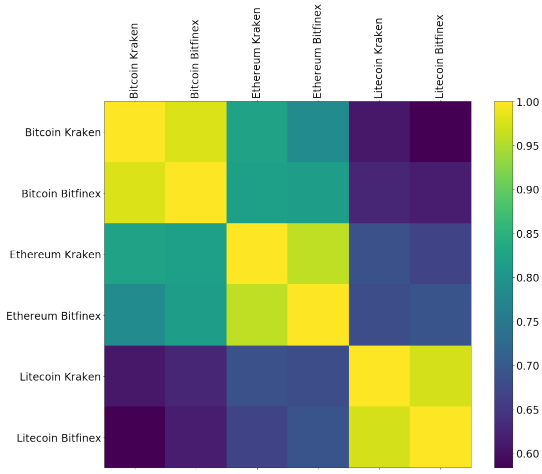


(f) Profit Pearson correlation of trades data, 1 second time buckets

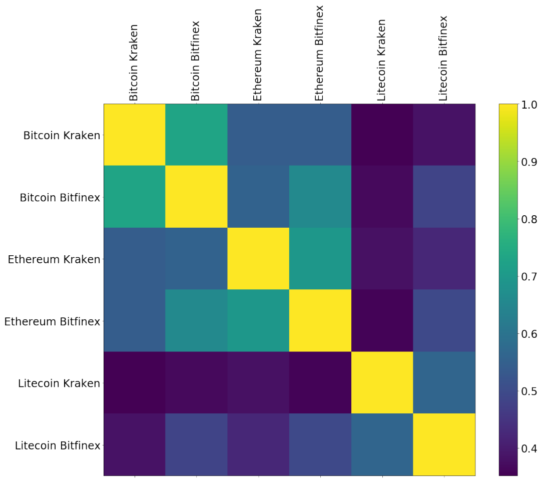
Figure A.11: Profit Pearson correlation of trades data, different time buckets



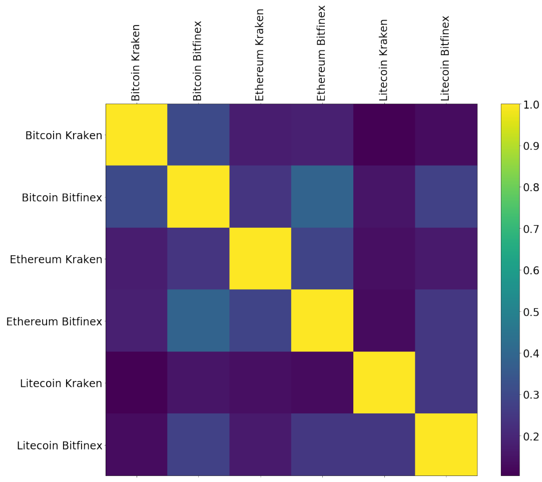
(a) Log-Profit Pearson correlation of trades data, 1 day time buckets



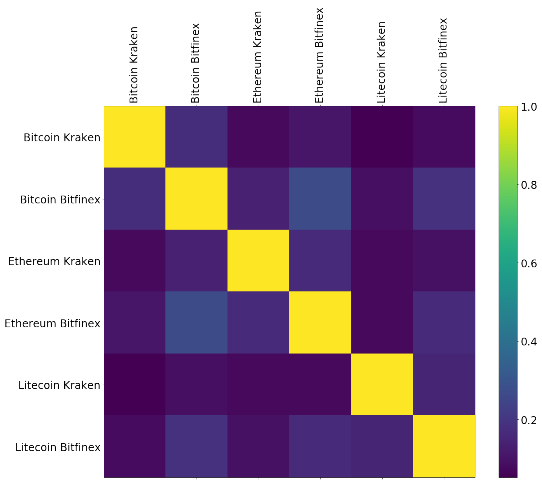
(b) Log-Profit Pearson correlation of trades data, 1 hour time buckets



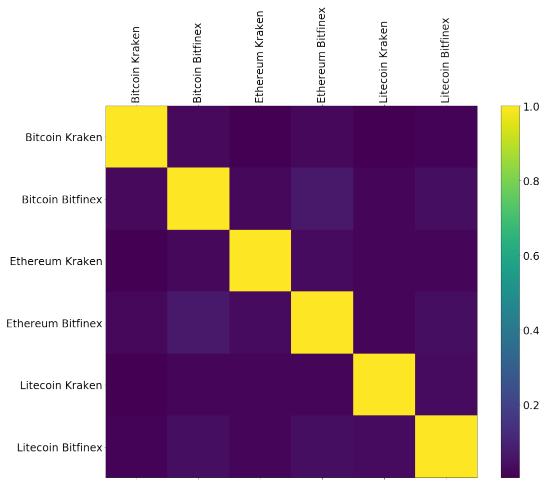
(c) Log-Profit Pearson correlation of trades data, 1 minute time buckets



(d) Log-Profit Pearson correlation of trades data, 10 seconds time buckets



(e) Log-Profit Pearson correlation of trades data, 5 seconds time buckets



(f) Log-Profit Pearson correlation of trades data, 1 second time buckets

Figure A.12: Log-Profit Pearson correlation of trades data, different time buckets

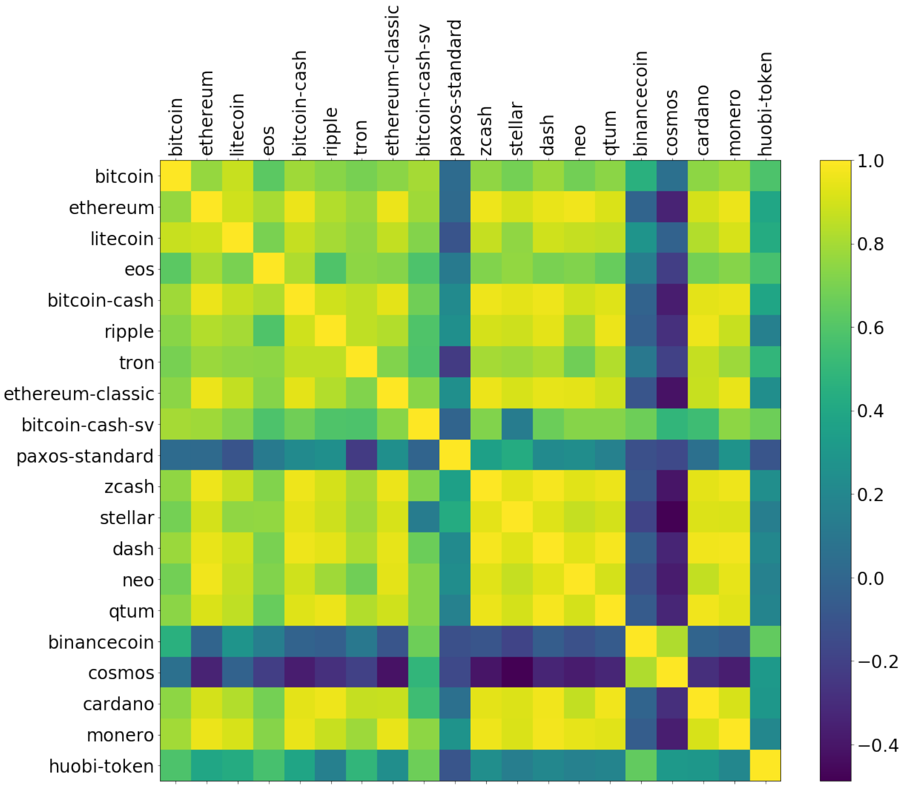


Figure A.13: Pearson correlation of day data

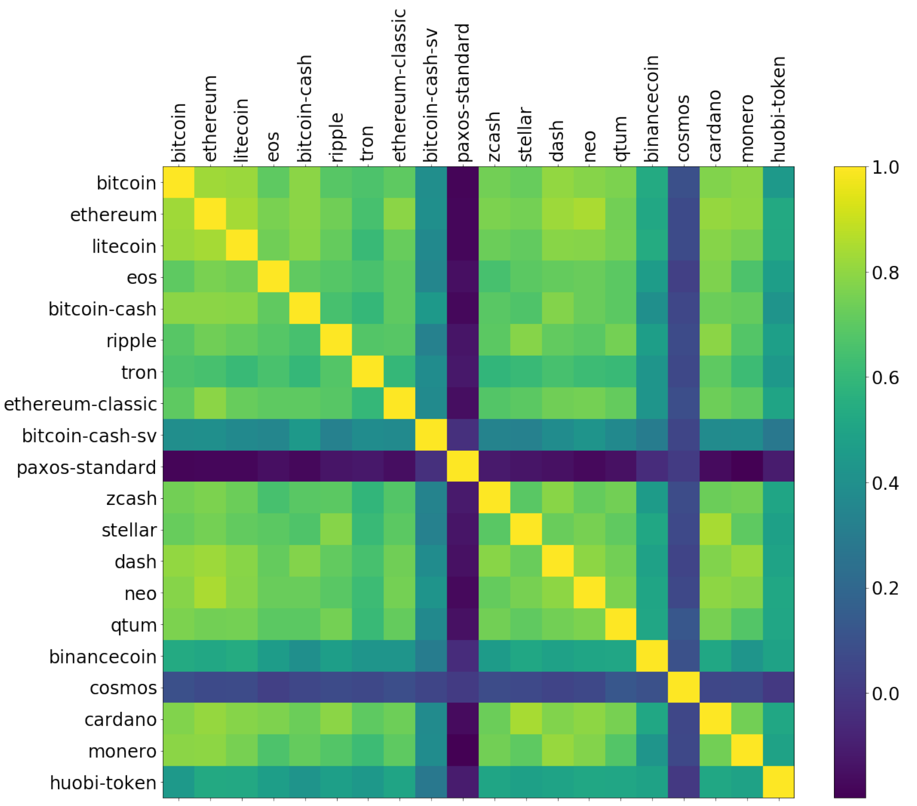


Figure A.14: Profit Pearson correlation of day data

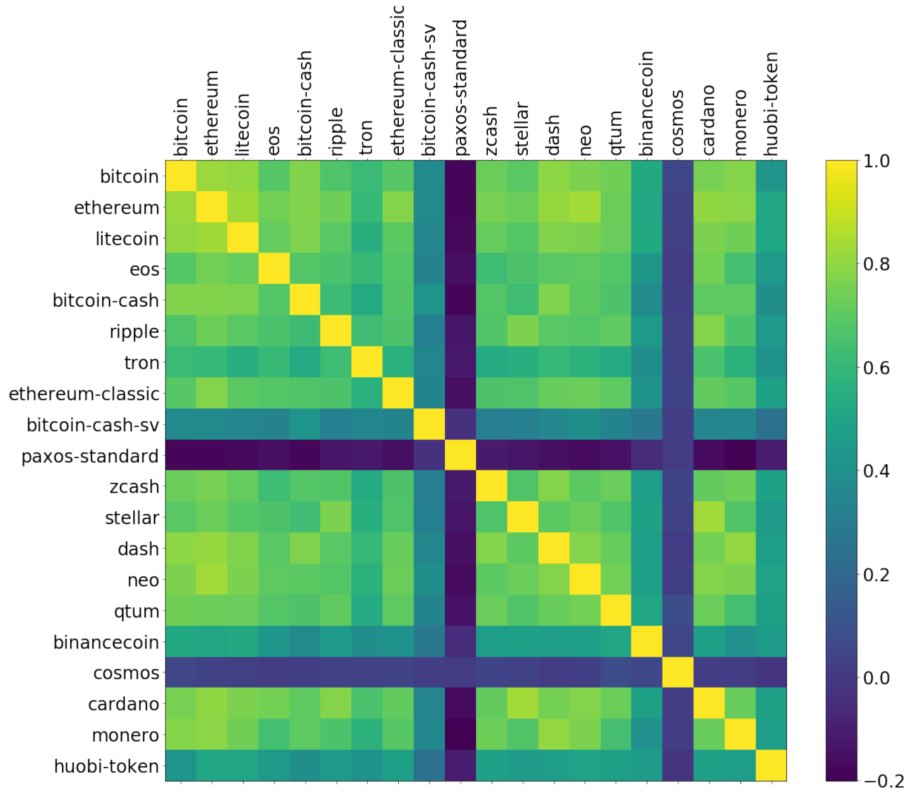


Figure A.15: Log-Profit Pearson correlation of day data

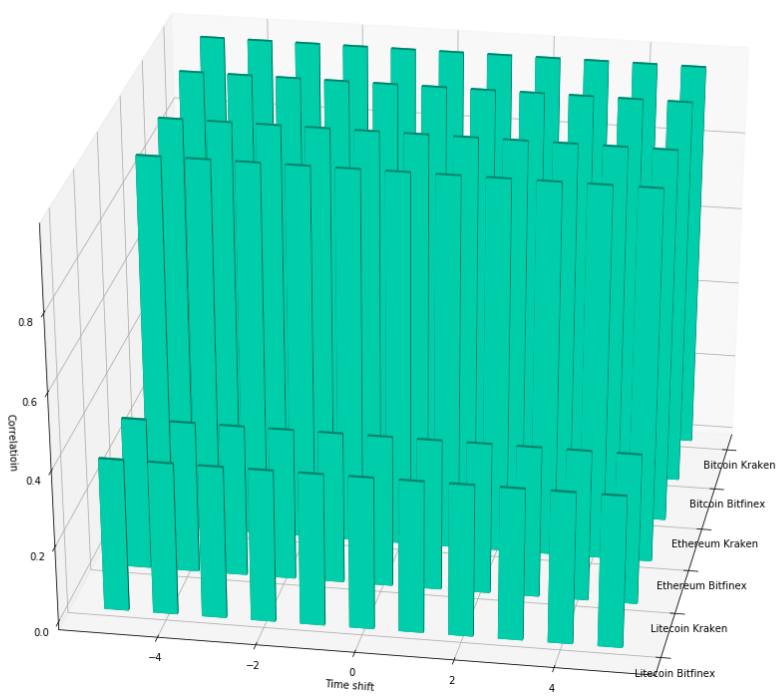
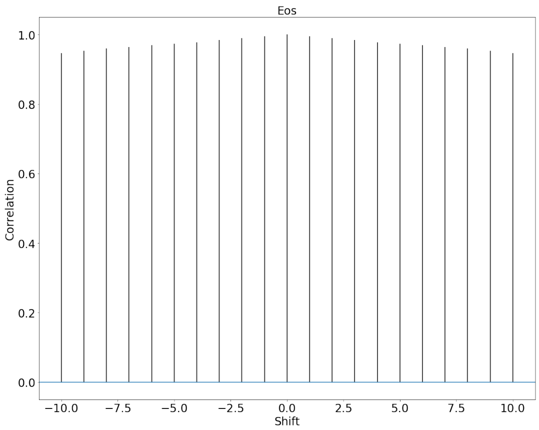
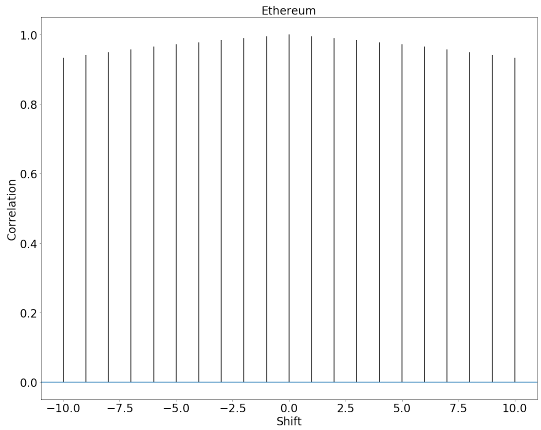


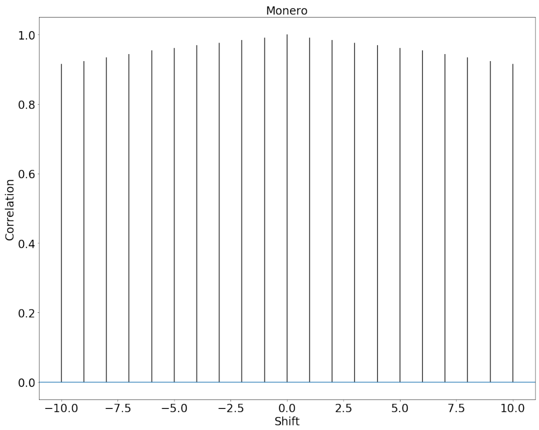
Figure A.16: Shifted Pearson correlation of Bitcoin vs others, time buckets: 1 second



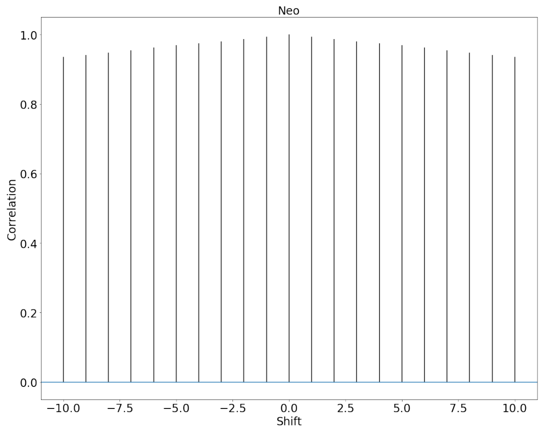
(a) Autocorrelation of Eos day data



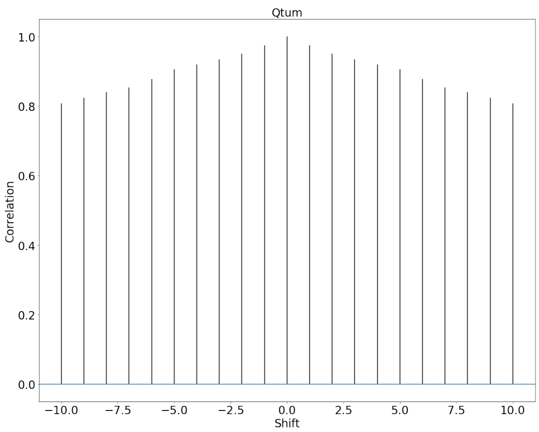
(b) Autocorrelation of Ethereum day data



(c) Autocorrelation of Monero day data



(d) Autocorrelation of Neo day data



(e) Autocorrelation of Qtum day data

Figure A.17: Autocorrelation selection of day data

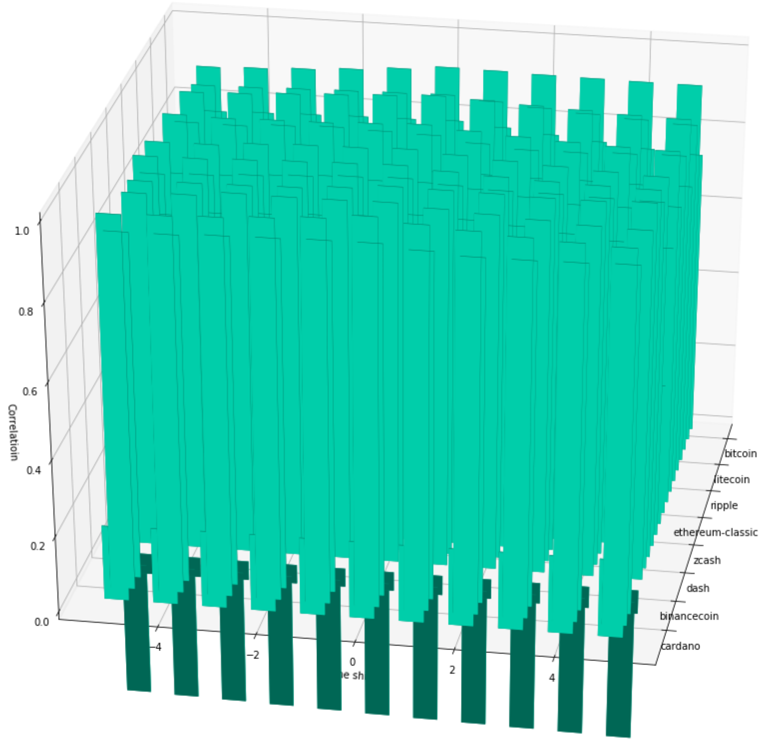


Figure A.18: Shifted Pearson correlation of day data Monero vs others - 3d

```
rollingSignalBoolean = signalData > signalData.rolling(movingAverageWindow).mean()
rollingSignalInt = 1 * rollingSignalBoolean
rollingSignalStep = rollingSignalInt - rollingSignalInt.shift(1)

combinedDataFrame = pd.DataFrame(index=tradingData.index)
combinedDataFrame['tradingData'] = tradingData
combinedDataFrame['rollingSignalStep'] = rollingSignalStep

combinedDataFrame = combinedDataFrame[combinedDataFrame['rollingSignalStep'] != 0]

def simulateTrading(row):
    if (row['rollingSignalStep'] > 0 and dollar != 0):
        crypto = dollar / row['tradingData'] * 0.998
        dollar = 0

    if (row['rollingSignalStep'] < 0 and crypto != 0):
        dollar = row['tradingData'] * crypto * 0.999
        crypto = 0

    return row

combinedDataFrame = combinedDataFrame.apply(lambda row: simulateTrading(row), axis=1)
```

Listing A.1: Moving average trades algorithm

```
for index in cryptoCurrencyData.index:
    if (cryptoCurrencyData[index] > cryptoCurrencyData.rolling(windowsSize).mean()[index] and dollar != 0):
        crypto = dollar / cryptoCurrencyData[index] * 0.998
        dollar = 0

    if (cryptoCurrencyData[index] < cryptoCurrencyData.rolling(windowsSize).mean()[index] and crypto != 0):
        dollar = cryptoCurrencyData[index] * crypto * 0.999
        crypto = 0
```

Listing A.2: Moving average day algorithm greedy approach 01

```
for index in tradingData.index:
    if (signalData[index] > signalData.rolling(windowsSize).mean()[index] and dollar != 0):
        crypto = dollar / tradingData[index] * 0.998
        dollar = 0

    if (signalData[index] < signalData.rolling(windowsSize).mean()[index] and crypto != 0):
        dollar = tradingData[index] * crypto * 0.999
        crypto = 0
```

Listing A.3: Moving average day algorithm greedy with different signal approach 02, shifted approach 03

```
for currency in data.columns:
    selection = data[currency]
    for index in selection.index:
        if ((selection[index] > selection.rolling(windowsSize).mean()[index]) and (dollar[currency] != 0)):
            crypto[currency] = dollar[currency] / selection[index] * 0.998
            dollar[currency] = 0

        if ((selection[index] < selection.rolling(windowsSize).mean()[index]) and (crypto[currency] != 0)):
            dollar[currency] = selection[index] * crypto[currency] * 0.999
            crypto[currency] = 0
```

Listing A.4: Moving average day algorithm greedy with taking into account all currencies approach 04

```
rollingSignalCorrelation = data.rolling(windowsSize).corr()

for index in data.index:
    for currency1 in rollingSignalCorrelation.columns:
        for currency2 in rollingSignalCorrelation.columns:
            if rollingSignalCorrelation[currency1][index][currency2] < rollingSignalCorrelation[currency2].shift(1)[index][currency1] and rollingSignalCorrelation[currency1][index][currency2] > 0 and (correlationBookDollar[currency1][currency2] != 0):
```

```
                    if data[currency2][index] > data[currency2].shift(1)[index]:
                        correlationBookCrypto[currency1][currency2] = correlationBookDollar[currency1][currency2] / data[currency1][index] * 0.998
                        correlationBookDollar[currency1][currency2] = 0
                elif correlationBookCrypto[currency1][currency2] != 0:
                    correlationBookDollar[currency2][currency1] = correlationBookCrypto[currency2][currency1] * data[currency2][index] * 0.999
                    correlationBookCrypto[currency2][currency1] = 0
```

Listing A.5: Moving average day algorithm greedy with taking into account all currencies and their lead-lag combinations approach 05

```
for index in data.index:
    for currency in data.columns:
        selection = data[currency]
        if ((selection[index] > selection.rolling(windowsSize).mean()[index]) and (crypto[currency] == 0)):
            corr = selection.rolling(windowsSize).corr(data['bitcoin'].rolling(windowsSize))[index]
            if (corr > 0):
                crypto[currency] = dollar[currency] / selection[index] * corr
                dollar[currency] = dollar[currency] - selection[index] * crypto[currency]
                crypto[currency] *= 0.998

        if ((selection[index] < selection.rolling(windowsSize).mean()[index]) and (crypto[currency] != 0)):
            dollar[currency] += selection[index] * crypto[currency] * 0.999
            crypto[currency] = 0
```

Listing A.6: Moving average day algorithm greedy with buying percentage correlation based parts approach 06