Tuesday 10th August, 2021

Master Thesis

submitted to

Delft University of Technology

in partial fulfilment of the requirements for the degree of

Master of Science

in Management Of Technology

Faculty of Technology, Policy and Management

Supervisory Committee:

1st Supervisor:	Prof. Dr. Ir. Marijn Janssen
2nd Supervisor:	Dr. Ir. Zenlin Roosenboom-Kwee
Committe Chair:	Prof. Dr. Ir. Marijn Janssen

What drives cryptocurrency market dynamics?

Analysing external variable influence on cryptocurrency prices

---

Maximilian Sachs 4236262

To be defended in public on Monday 30th August, 2021

contact@arkmon.eu

Summary

Since the introduction of Blockchain technologies in Bitcoin (Nakamoto & Bitcoin, 2008), there has been over a decade of innovation and growth. The cryptocurrency markets have formed their own ecosystem and followings. On Wednesday 12th May, 2021, cryptocurrencies reached their highest total market capitalisation yet, at $ 2.5 trillion (coinmarketcap.com, 2021). This rapid growth in valuation has attracted many new entrepreneurs and investors into the industry. However, the cryptocurrency world comes with its own set of challenges. This research is the first to create a multivariate model of the effects of external variables on cryptocurrency prices. For the first time, the results enable to compare the magnitude and direction of effects from several external variables. The model proposed in this thesis can help managers, developers and investors alike in understanding external risks to cryptocurrency projects.

Existing literature, from the theoretical base of cryptocurrency market literature, was reviewed to identify which variables should be included in the analysis. The reviewed literature was found to be focused on either the technical implications of blockchain as a technology or on the narrow effect of a particular variable on the cryptocurrency price. Comparatively little research investigates the influence of external variables. Their focus lies on the narrow effect of a singular variable only. However, there might be important observations to be made when considering multiple variables together and within the context of each other. Based on findings of existing literature and supplemented by the author's experience, the following variables were selected for analysis: *price volatility, macro-economic trends, Github statistics, Google trends score, Twitter sentiment, community size and Internet activity levels*. Price volatility describes the magnitude of variance of cryptocurrency price. The macro-economic trends were represented by the close price of the SP500, Oil and Gold. Github statistics use the metrics of the cryptocurrencies publicly accessible Github repositories. The Github metrics are: number of open issues, closed issues, pull requests, forks and stars. Google trends score represent the relative popularity with which a particular cryptocurrency was searched. Twitter sentiment represents the general populations opinion about the different cryptocurrencies. A distinction to Influencers was made. Community size is the amount of followers a cryptocurrency has on Twitter and Reddit. Finally the Internet activity levels represent the daily cycles of global activity.

Over several months of real-time data collection efforts, datasets were compiled for both intra-day (variations of time-span smaller than 1 day) effects and inter-day (variations over multiple days) effects. The analysis was performed using an hierarchical Bayesian linear regression model (Gelman, 2006). This method allows for a distinction to be made between variables acting equally on all cryptocurrencies and variables having a specific effect on an individual cryptocurrencies. The stochastic nature of the method results in a probability distribution for the coefficient value of each variable. These probability distributions could be evaluated to not only judge the magnitude and direction, but also the significance of each variable. The most important results are summarised in the following paragraphs.

The cryptocurrency prices were estimated using external variables and the model achieved a mean R^2 of 0.69 for intra-day and a mean R^2 of 0.91 for inter-day effects. The most significant factor was identified to be the community size, however an opposite direction of effect was observed for the community size of Twitter vs. Reddit. An increase of the number of Reddit subscribes tends to have a positive effect on the cryptocurrency prices, while an increase in the number of Twitter followers tends to have a negative effect. The next most influential variable was found to be the long-term progress indicator of total issues on the cryptocurrency's Github repository. The Macro economic variables of SP500 Volume, Oil

Volume and Gold price had a significant influence as well. While cryptocurrencies have been described as hedges against conventional financial markets (Giudici & Abu-Hashish, 2019), it was found that SP500 and Gold price had a positive correlation on cryptocurrency prices. A reduction in price volatility tends to be followed by an increase in cryptocurrency price.

The external variables related to the social context of cryptocurrencies led to some of the most notable results. Twitter sentiment was identified as a powerful estimator. Both Group sentiment and Influencer sentiment are important. The effect of Influencer sentiment is significant, but may change depending on the cryptocurrency. Further research into key individuals in the social networks is needed to explain this variation. An increase in short-term Group sentiment largely tends to correspond with a decrease in cryptocurrency price. A novel type of social media user class was identified as an information Aggregator, which are Twitter users with a large number of friends. On Twitter the number of friends is defined to be the number of other users followed by the posting user. An increase in Aggregator sentiment correlates with a cryptocurrency price increase. The development progress of open source Github repositories were found to be a measure of the fundamental value of the cryptocurrency. Increasing the number of total issues on a cryptocurrency Github repository showed a significant and positive effect on the cryptocurrency price. When the global internet activity levels were low, a positive effect was observed on cryptocurrency prices.

In terms of the homogeneity of effects, it was observed that the Twitter sentiment and Github metrics had largely similar effects on all cryptocurrency prices, while macro-economic trends, Google trends and community sizes showed different effects depending on the cryptocurrency. It should be noted that the exact coefficient magnitudes varied amongst different cryptocurrencies and more research will be required to explain these differences.

While the current model has achieved a remarkably good performance in estimating the cryptocurrency price, there can still be more improvement. Incorporating theories about the dynamics of social networks in the context of the cryptocurrency innovation system may help to improve the explanatory power of the model. While working on this thesis many opportunities for future research have been identified. Especially interesting, and also challenging, would be the inclusion of the regulatory state and news stories. The difficulty of those variables arises from their discrete and event based nature, which cannot be analysed in a model that makes the assumption of variables to change smoothly over time. Key individuals in the social networks should be investigated further and best practices for managing cryptocurrency communities on social media platforms could be developed. This research furthermore indicates that there might be additional, as of yet unidentified, variables influencing cryptocurrency prices. Conventional narrow scientific research alone is insufficient to describe the complex dynamics of the cryptocurrency world, but approaching the problem from the broad perspective could have success.

This research has shown the hierarchical Bayesian model to be an effective model for multivariate effects in cryptocurrency market dynamics. As future research into additional variables is conducted, results can be linked back to the findings of this research to continually expand our understanding of cryptocurrency markets. Development progress on Github is an important metric for the fundamental value of the cryptocurrency. The price value is the result of a complex model involving a whole set of variables, with no singular variable being responsible for a disproportionately large part of the observed changes. Development efforts, Social Media sentiment and macro-economics need to be accounted for to estimate the cryptocurrency prices.

Contents

Glossary

bear market A market with a general down-wards trend.

blockchain Is a technical solution for a distributed and trust-able database.

bull market A market with a general up-wards trend.

cryptocurrency Is an application of blockchain technology to keep track of balance amounts in various wallets.

exchange A place hosting a number of and providing access to markets.

fiat currency Fiat currencies are the established centrally controlled currencies, e.g. Euro, Dollar.

granularity A high data granularity implies a small interval. A low granularity implies a large interval. Also referred to as the resolution in time.

initial public offering First sale of a companies stock to the general public on an stock exchange.

initial exchange offering An ICO executed by on the platform of an exchange.

initial coin offering Is analogous to an initial public offering (IPO).

inter-day effect Effects that only manifest over a time-span larger than a day.

interval The interval is the time between two data-points of a time-series, also referred to as the data spacing.

intra-day effect Effects that react to a cause in a time-span shorter than a day.

market A pairing of two currencies that can be traded for one another.

market capitalisation Total value represented by all units of a specific crypto currency.

open source The source code is publicly readable.

principal component analysis A statistical method to identify the most significant axis in a dataset.

retail investor A non-professional, part time trader/investor.

Acronyms

API application programming interface.

CNN Convolutional Neural-Network.

GRU Gated Recurrent Unit.

LSTM Long Short-Term Memory.

NYSE New York stock exchange.

PCA principal component analysis.

ROPE Region of Practical Equivalence.

SEC U.S. securities and exchange commission.

1 Introduction

Blockchain technology has found its first major application in cryptocurrencies. Since the introduction of Bitcoin (Nakamoto & Bitcoin, 2008) a multitude of clone projects have emerged improving on the original design and sometimes adding entirely new functionality. Most of these cryptocurrency projects have in common that they are open source and funded through the sale of their cryptocurrency. Developers and other expenses, like marketing, can only be paid for, as long as the cryptocurrency has a sufficiently high price[1]. Should the price drop too low, one could expect developers to leave the project and ultimately a lack of development could lead to a project failure. This may result in a vicious cycle as investors[2] sell their cryptocurrency, which is akin to their stake in the project, and thereby lower the price even further. Conversely as market capitalisation of a cryptocurrency increases, more opportunities might become available for the project, for example collaborations with other businesses or being able to afford more full time developers.

In the cryptocurrency world it is fair to claim that the project success is dependent on the sustained financial success in terms of price of the associated cryptocurrency. From a management and investment perspective, it is important to understand and be aware of influencing external variables. Understanding the magnitude and direction of the influence of external variables on the cryptocurrency prices is the first step in developing a strategy to mitigate or treat these additional risks in cryptocurrency projects. Due to the rapid pace of change of the number of market participants and the level of public interest, cryptocurrency market dynamics are not the same as even just a few years ago. The existing scientific literature could become outdated and opportunity for new research with larger and more up-to-date datasets arises.

A theoretical model relating influencing variables to the cryptocurrency market price could help practitioners involved in the management or development of cryptocurrency projects to benefit from a better planning and decision making framework. For investors in cryptocurrency projects the level of risk of an investment might be reduced by this framework as well.

In the following, the steps taken in this thesis to identify and analyse influencing external variables are outlined. In Section 1.1 the general types of actors and their involvement in cryptocurrency trading are described. Understanding their motivations and interests will further help to direct this research. The research questions are defined in Section 1.2. In Section 1.3 an initial set of potentially influencing variables is established.

The research question and the initial set of variables are used to conduct a literature research in Section 2. The literature research will help to both identify further influential variables and to identify research gaps. Based on the variables and literature research findings, Section 3.1 establishes hypothesis for each variable.

Section 3 describes the methodology steps to conduct the analysis. The core of the analysis is the application of an Hierarchical Bayesian linear regression model. In Section 4 the results of the analysis and the observed effects of each variable are described, and then discussed in Section 4.4. Recommendations for practitioners (Section 5.3) and future research recommendations are listed (Section 5.4). A final summary of the main findings is found in the conclusion in Section 5.1.

[1] This is making the assumption that the cryptocurrency itself cannot be directly used as payment, e.g. for the heating bill. Payment therefore requires conversion of the cryptocurrency into an acceptable means of payment.

[2] cryptocurrency investors are early buyers of a cryptocurrency at a low price, who anticipate the cryptocurrency utility to increase resulting in an increased price.

1.1 General Background Crypto Currency Markets

In this section some background on the terminology and the general working principles of the cryptocurrency markets are laid out.

A market is a digital place where a unit of a cryptocurrency can be exchanged into another cryptocurrency or fiat currency. Access to a market can be gained through exchanges, which commonly list a number of different markets on their platform. In order to trade, a user creates an account on an exchange and then deposits some capital, fiat or cryptocurrencies are usually accepted. With the cryptocurrency in their account, a user can place orders of types buy or sell. If an order of the opposite type can be matched at the selected price a trade occurs.

Many cryptocurrency traders can be considered as retail investors. They are non-professional traders with comparatively little capital. Retail investors often communicate their opinions and investments with each other using various social media platforms. This leads to a fast spread of information, which might help retail investors coordinate their efforts to exploit opportunities. An example of this could be seen in early 2021 with the highly unusual GameStop stock. The stock increased over 1000 % in value, caused by investors from the sub-reddit “WallstreetBets” (Boylston, Palacios, Tassev, & Bruckman, 2021).

At the time of writing there were 9756 cryptocurrencies listed on (coinmarketcap.com, 2021). There are likely even more cryptocurrencies, since Coinmarketcap might not be able to access all cryptocurrency exchanges in existence. Furthermore, there is a constant churn of new cryptocurrencies being created and new markets being launched. As new regulations are introduced markets and exchanges may be affected. The number of cryptocurrency, markets and exchanges is constantly changing. The numbers from data aggregators should therefore be considered as a lower bound for the true values.

There are technical differences between cryptocurrencies. While they are fundamentally based on Blockchain technologies, they leverage the benefits of the technology in different ways to solve a variety of problems, for example storing value, transacting value, authentication, executing arbitrary code and many more. The details of these differences are not directly relevant to this research, but it is important to avoid the misconception of all cryptocurrencies being the same.

Usually a cryptocurrency project creates some of its total coin supply at the beginning of the project. This supply of “coins” needs to be sold to network participants or investors. This can be achieved in a variety of ways. A common approach is an initial coin offering (Sachs, Balaji, Furtado, Sathyanarayanan, & Wong, 2021), where the project first offers a fixed amount of cryptocurrency for sale to the public on a project controlled website. Commonly other cryptocurrencies are accepted forms of payment, some projects also accept fiat currency through credit cards, Paypal or similar payment methods. A second common approach is an initial exchange offering, which is similar to the initial coin offering, however, instead of selling though their own website, the sale is conducted by and on the platform of an exchange.

The initial phase of a cryptocurrency market can have a very high level of price volatility associated with it. During this time the projects are commonly funded by the profits from the initial coin offering or initial exchange offering. After the cryptocurrency has been brought into circulation and has been listed by a number of exchanges some of the volatility may reduce and a stable phase of trading begins. Especially, when the cryptocurrency is purchased more for its utility rather than as a speculative investment. At this point the project may have used a large portion of its initial capital and will require to continually sell small additional amounts of its coin supply at market price to finance continued development. The project is interested in the cryptocurrency having the largest possible value at the time of sale, which means less coin supply needs to be used up for development. Conserving coin supply can increase the "runway" of the project or make the founders substantially more wealthy in the long term. Conversely if the value of the cryptocurrency drops too far the "runway" shortens and could become so short that it is no longer possible to complete the project, leading to project failure.

In "conventional" (non-cryptocurrency) startups the early phases of development are usually funded through investment rounds by angel or venture capital investors. If development progresses well, an initial public offering might follow. Afterwards, the company will be tradeable on a public stock exchange. Contrary to an initial coin offering of a cryptocurrency, there are strict and well-defined requirements and fees associated with an initial public offering of company stock. The process itself is regulated by a governing body of the country in which the exchange is located. Some of the most well known exchanges are the New York stock exchange (NYSE) and the NASDAQ, which are both under supervision of the U.S. securities and exchange commission (SEC). In most cases a company is listed on only a single exchange as a result of the regulatory hurdle of listing requirements. In contrast, many cryptocurrencies are listed on a multitude of cryptocurrency exchanges.

Figure 1 shows the total market capitalisation of cryptocurrencies in terms of USD. Before the first major valuation increase in early 2018 there was little capital bound in the cryptocurrencies. Simultaneously there was an increase in the number of papers published per year following the valuation spike. This will be further elaborated upon in Section 2.1.

The number of participants in the markets and the size of their stake have increased substantially since 2018, as can be seen especially well on the volume increase, shown in the grey area in the lower part of the 1. Wednesday 12th May, 2021, was one of the first days in history where cryptocurrency market volume over 24h exceeded that of the NASDAQ. The traded volume on the NASDAQ was \$274 744 402 783 (nasdaqtrader.com, 2021) and the total volume of cryptocurrencies was \$335 670 927 073 (coinmarketcap.com, 2021) during the same time.

Figure 1: Cryptocurrency total market capitalisation (coinmarketcap.com, 2021)

1.2 Research Questions

To enable the managers, developers and investors of cryptocurrency projects to better manage external risks, this research will attempt to identify and compare a number of variables with influence on the cryptocurrency prices. Understanding the drivers of dynamic changes in cryptocurrency markets and improving the predictability of the prices over time will enable better informed decision making and reduce the likelihood of cryptocurrency project failure. The following research questions are proposed:

RQ 1: Which variables have an influence on cryptocurrency price volatility?

RQ 2: What is the magnitude of effects of the variables influencing cryptocurrency price volatility?

RQ 3: Which variables have the biggest influence on price increase?

RQ 4: What theories can explain the observed effect of the variables?

The literature research, Section 2, will be used to identify the set of variables, answering RQ1. Then the relationships of variables to price will be analysed using a hierarchical bayesian linear regression model, Section 3, to answer RQ2. The obtained results, Section 4, are reviewed to identify the most important variables answering RQ3. And finally potential explanations of the observed results are given in the discussion, Section 4.4, to answer RQ4.

1.3 Influencing variables

Markets are social dynamic systems. They are being affected by the actions of many participants. The actions of these participants have direct and emerging effects on the markets and the social system surrounding them. For some of these effects metrics can be obtained, which will be referred to as variables, such as the price of a cryptocurrency.

An initial set of variables is needed as a starting point for the literature research, based on which the final set of variables for analysis will be selected in Section 2.4. The author's personal experience has been shaped by following and investing in various cryptocurrency projects since 2014. Reflecting on that time, allows a range of potentially significant variables to be identified.

Selecting an initial set of variables based on the author's experience with cryptocurrencies is of course highly subjective. However, the systematic application of the methodology employed in literature research and in the analysis should diminish the impact of this initially subjective choice of variables allowing this research to come to objective and scientific results in the end. The following variables are established as potentially useful to estimate the cryptocurrency prices:

- General economic trends
- Total capitalisation and volume of markets
- Regulatory restrictions
- General population and influencer sentiment
- Cryptocurrency development progress

In the following Section 2 the initial variables are used to identify relevant literature.

2 Literature research

The goal of this literature review is to identify existing research of the variables listed in Section 1.3 and to identify additional variables useful for further analysis in this research. The resulting set of variables, which are the answer to **RQ1**, are shown in Section 2.3.

2.1 Literature research methodology

The first step in finding literature to review is to make a selection of which repositories to search. An important criteria for repository selection is that it should be accessible and provide an up-to-date comprehensive list of peer reviewed papers (Jesson, Matheson, & Lacey, 2011, p. 45). The Scopus database is considered to sufficiently fulfil the selection criteria and will be used for the remainder of this literature research.

The theoretical base of this research is cryptocurrency market literature. Literature that is peer reviewed and was published in journals is the most desirable type. However, academic research on cryptocurrency markets is still in its infancy. Due to the novelty of the subject, selecting only peer reviewed papers limits the number of potential results, therefore conference proceedings will be included as well. The method of selecting literature will be based on a keyword search. The keywords should help find literature relevant to the initial set of influencing variables defined in Section 1.3. We are interested in research relevant to the dynamics of cryptocurrency prices in the broader context of financial markets, the economy and society. The initial variables were combined with the field and context relevant to this research resulting in the following sets of keywords are to be searched:

1. market dynamics
2. cryptocurrency economic trends
3. cryptocurrency regulation
4. cryptocurrency social sentiment
5. variables influencing market price volatility

Only literature in the English language will be included. For each keyword search results are sorted by relevance. Then for each result the relevance was checked by reading the paper abstract and confirming the paper to be related to the topic being investigated. If this was not found to be the case, then it was rejected immediately. There may be overlap among the results of different searches. Only the first encounter with a paper is included. A total of 473 results were obtained. Many of the results could quickly be excluded from further consideration as they were not found to be relevant to this research.

This approach may result in some selection bias, as the relevance to the topic at hand is subjectively judged by the author. However, within the scope of this master thesis the variance in search terms is considered sufficient to achieve the goal of identifying a set of relevant variables. The resulting set of variables in Section 2.3 appears to confirm this assumption, as a wide spectrum of variables is covered. Future research can of course always include additional variables, but for the scope of this research the obtained variables are considered sufficient.

Comparatively few papers investigate the link between the external environment of markets to the trends in prices. Some topics also saw varying degrees of popularity. As described in the introduction (Section 1.1), the fundamental cryptocurrency market dynamics have changed since the price bubble in 2017/2018. Literature published before this event has based their results on the dynamics before the bubble. This is assumed to be outdated. The focus in this research is placed on literature published in 2017 or more recently.

Fig. 2 shows the number of results for each keyword search per year. Especially after the market bubble in 2018, here visualised by the Bitcoin market capitalisation, the rate of publications and with that the academic interest in cryptocurrencies has increased. The publications related to cryptocurrency regulation (shown as green in Fig. 2) have been made since 2014 and appears to have reached its peak in 2019. The topics of sentiment and market dynamics have largely only been studied in the last 2 or 3 years.

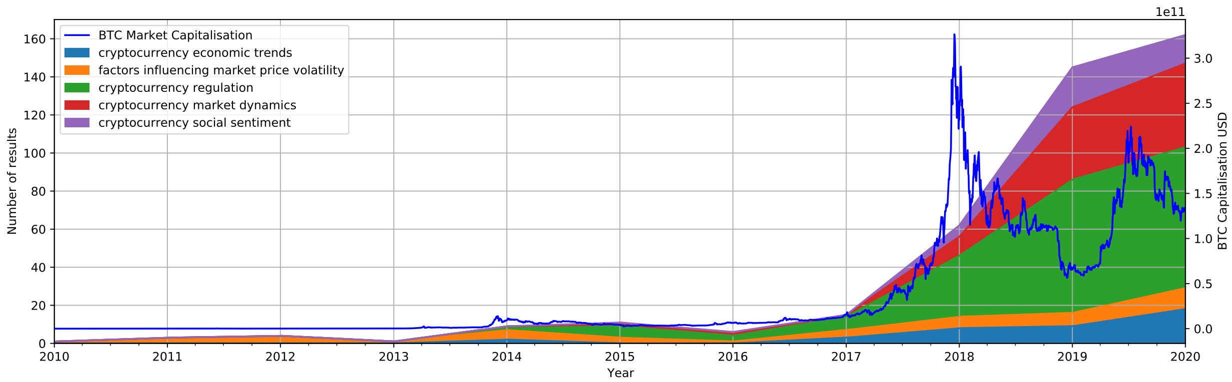

Figure 2: Number of results for the keyword searches over time

Scopus annotates each article with subject areas. Fig. 3 shows the number of times a subject area was encountered in the search results. The total number of subject areas might be larger than the total number of search results, as results might have one or more annotated subject areas. The 4 most common subject areas occur in 62.7 % of search results and are *Computer Science*, *Economics, Econometrics and Finance*, *Social Sciences* and *Business, Management and Accounting*. This is largely expected as the listed fields might have many promising applications for blockchain-technology. Additionally the cryptocurrency markets are becoming a social phenomenon.

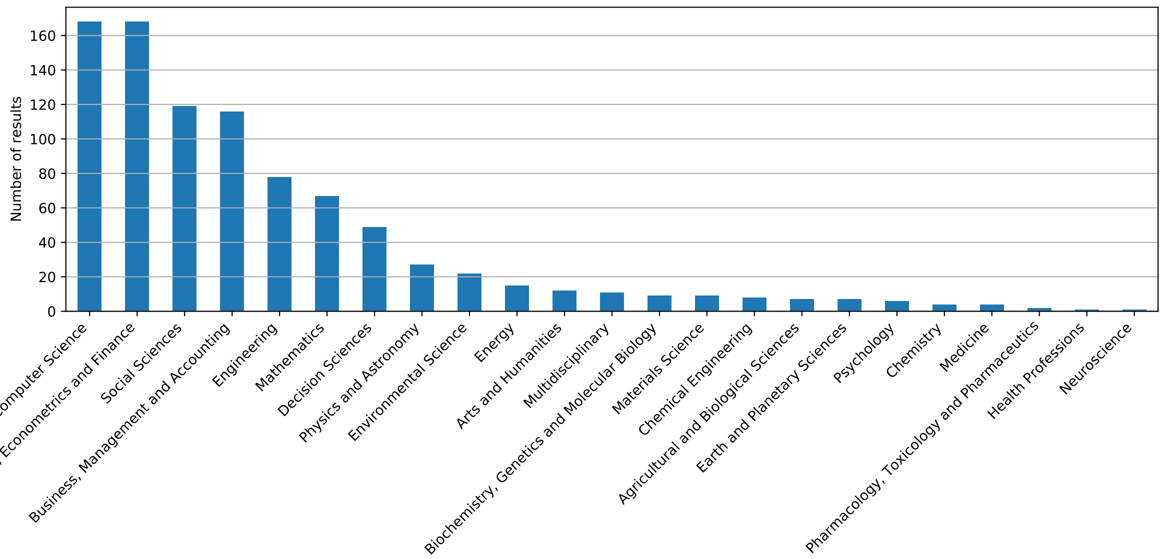

Figure 3: Number of results per subject area

Similarly Scopus provides a country annotation per result. Fig. 4 shows the distribution of results per country. An article may have involvement of multiple authors from different nations. The 4 most common countries cover 38.6 % of results are the *Russian Federation*, *United States*, *United Kingdom* and *China*. The interpretation of this result might be considered political and is therefore left to the reader. It appears some of the biggest global economies have an interest in supporting research into cryptocurrencies.

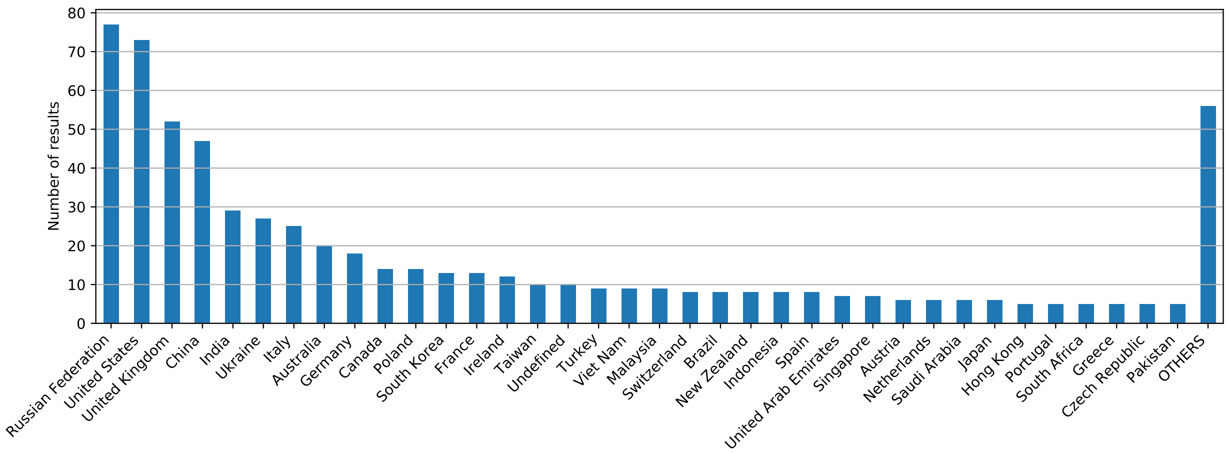

Figure 4: Number of results per country

In the following the selected results from each keyword search, as described at the beginning of this chapter, are discussed. The first keyword search ***cryptocurrency market dynamics*** obtained 113 results. The selected set of papers is shown in Table 1. Much of the literature is fairly recent and has few numbers of citations. A low number of citations is not an indicator for bad research, just as a high number of citations is not necessarily indicative of high quality research. For the purposes of this literature research the variance of number of citations is not considered relevant. To reiterate, the objective of the literature research is to identify a set of external variables which have been researched before. The selected papers were considered relevant to this goal based on their abstract. The individual papers will be discussed in Section 2.2.

Next the keyword search ***cryptocurrency economic trends*** obtained 42 results. The selected set of papers is shown in Table 2. Again a high degree of variance of number of citations is observed.

Next the keyword search ***cryptocurrency regulation*** obtained 197 results. The selected set of papers is shown in Table 3. The vast majority of the literature obtained from the search was focusing on the legal, ethical and social implications of cryptocurrency regulation. However, only few papers considered regulation as a measurable variable with direct impact on the cryptocurrency prices. This might be in parts due to the difficulty of transforming regulatory changes into a continuous time series that allows comparison to price. Changes to regulatory law are largely discontinuous events occurring at a single point in time. The problem of representing regulatory state over time is further discussed in Section 2.4.

Next the keyword search ***cryptocurrency social sentiment*** obtained 46 results. The selected set of papers is shown in Table 4. Comparatively few of the obtained results were in line with the objective of this literature research.

Finally the keyword search ***variables influencing market price volatility*** obtained 28 results. The selected set of papers is shown in Table 5. Of the five searches, this set of keywords collected the least number of results. This could indicate again the infancy of the field, as well as the difficulty of explaining price volatility.

Table 1: Selected papers for the keywords: cryptocurrency market dynamics

Title	**Paper**	**Source**	**Num. cited**
Correlation-based dynamics and systemic risk measures in the cryptocurrency market	(Liang et al., 2018)	*2018 IEEE International Conference on Intelligence and Security Informatics*	3
Evolutionary dynamics of the cryptocurrency market	(ElBahrawy et al., 2017)	*Royal Society Open Science*	58
Principal component analysis based construction and evaluation of cryptocurrency index	(Shah et al., 2021)	*Expert Systems with Applications*	0
Market dynamics, cyclical patterns and market states: Is there a difference between digital currencies markets?	(Bejaoui et al., 2019)	*Studies in Economics and Finance*	1
Towards an understanding of cryptocurrency: A comparative analysis of cryptocurrency, foreign exchange, and stock	(Liang et al., 2019)	*2019 IEEE International Conference on Intelligence and Security Informatics*	2
Herding and feedback trading in cryptocurrency markets	(King & Koutmos, 2021)	*Annals of Operations Research*	0
Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques	(Altan et al., 2019)	*Chaos, Solitons and Fractals*	100
The intraday dynamics of bitcoin	(Eross et al., 2019)	*Research in International Business and Finance*	24
What determines bitcoin exchange prices? A network VAR approach	(Giudici & Abu-Hashish, 2019)	*Finance Research Letters*	46

Table 2: Selected papers for the keywords: cryptocurrency economic trends

Title	**Paper**	**Source**	**Num. cited**
Buzz Factor or Innovation Potential: What explains cryptocurrencies' returns?	(Wang & Vergne, 2017)	*Public Library of Science*	31
The decade-long cryptocurrencies and the blockchain rollercoaster: Mapping the intellectual structure and charting future directions	(Klarin, 2020)	*Research in International Business and Finance*	13
Price clustering and sentiment in bitcoin	(Baig et al., 2019)	*Finance Research Letters*	13
Exploring evolution trends in cryptocurrency study: From underlying technology to economic applications	(Jiang et al., 2021)	*Finance Research Letters*	0
Data science in economics: Comprehensive review of advanced machine learning and deep learning methods	(Nosratabadi et al., 2020)	*Mathematics (MDPI)*	6
Deep Learning Approach to Determine the Impact of Socio Economic variables on Bitcoin Price Prediction	(Aggarwal et al., 2019)	*2019 12th International Conference on Contemporary Computing*	5
An econometric model to estimate the value of a cryptocurrency network. The bitcoin case	(Abbatemarco et al., 2018)	*European Conference on Information System*	1

Table 3: Selected papers for the keywords: cryptocurrency regulation

Title	**Paper**	**Source**	**Num. cited**
International models of legal regulation and ethics of crypto currency use: Country review	(Panova et al., 2019)	*Journal of Legal, Ethical and Regulatory Issues*	0
Taming the blockchain beast? Regulatory implications for the cryptocurrency Market	(Shanaev et al., 2020)	*Research in International Business and Finance*	12

Table 4: Selected papers for the keywords: cryptocurrency social sentiment

Title	**Paper**	**Source**	**Num. cited**
Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model	(Li et al., 2019)	*Frontiers in Physics*	9
Sentiment analysis using R: An approach to correlate cryptocurrency price fluctuations with change in user sentiment using machine learning	(Rahman et al., 2019)	*International Conference on Informatics*	2
Advanced social media sentiment analysis for short-term cryptocurrency price prediction	(Wołk, 2020)	*Expert Systems*	2
Bitcoin volatility, stock market and investor sentiment. Are they connected?	(López-Cabarcos et al., 2021)	*Finance Research Letters*	5

Table 5: Selected papers for the keywords: actors influencing market price volatility

Title	**Paper**	**Source**	**Num. cited**
Fundamental and behavioural determinants of stock return volatility in ASEAN-5 countries	(Thampanya et al., 2020)	Journal of International Financial Markets, Institutions and Money	4
Riding the Wave of Crypto-Exuberance: The Potential Misusage of Corporate Blockchain Announcements	(Akyildirim et al., 2020)	*Technological Forecasting and Social Change*	3
Financial variables affecting oil price change and oil-stock interactions: a review and future perspectives	(Liu et al., 2019)	*Natural Hazards*	5

2.2 Paper reviews

Two of the papers selected were systematic literature reviews. These will be reviewed and discussed first, to obtain an overview of the current state of research. Afterwards, the remaining papers relevant to the specific variables are discussed.

Klarin (2020) conducts a systematic and algorithmic analysis of scientific and practitioner oriented literature related to cryptocurrency development and trends of the last decade. The author creates a network graph, placing similar topics in scientific literature close to each other and identifying four main topics of research: Bitcoin and cryptocurrencies, Blockchain adoption, Blockchain environment and Business model innovations. The variable of trust is highlighted as one of the key offerings of blockchain technologies and is identified as instrumental in both the blockchain eco-system development and the adoption processes. By comparing the scholarly and practitioner literature Klarin (2020) identifies research gaps, as practitioner literature can give insights on the wider topics and trends with practical relevance. The main observed difference is that while scientific literature focuses on the underlying technology of Blockchain, media outlets focus more on the financial opportunities of investing and the political and economic implications of cryptocurrencies. Klarin (2020) also identifies the "global issues concerning growth of less developed countries, combating poverty and inequality" (Klarin, 2020, p. 9) as under-researched in academic literature. Political statements and actions could be giving credibility to cryptocurrencies, increasing trust in the technology and thereby increase adoption. Actions from other non-governmental reputable actions, like large Banks and Corporations, could have a similar effect.

Jiang et al. (2021) apply bibliometric analysis to cryptocurrency related papers published between 2009 and 2019. The authors distinguish two main fields of cryptocurrency literature: Underlying technology and economic applications. As part of their analysis they create an "institution cooperation network, which reflects the regional characteristics of institution cooperation" (Jiang et al., 2021, p. 4). Interestingly, institutions in Asia and Europe collaborate within their region and with each other, while American institutions "have not yet formed a cooperation network" (Jiang et al., 2021, p. 4) on the topic of cryptocurrencies. European institutions primarily focus on price prediction and volatility of cryptocurrencies, while chinese institutions focus on the underlying technology fundamentals of blockchain, like privacy and security. Klarin (2020) and Jiang et al. (2021) show research into Bitcoin alone is singularly larger than into the remaining cryptocurrencies as a whole field. Their main finding is "research priorities of cryptocurrency shift from underlying technology to economic applications" (Jiang et al., 2021, p. 7). The main proposed future research fields are in-depth research into underlying theories and price predictions and developing specific application schemes in various fields.

The works of Klarin (2020) and Jiang et al. (2021) indicate the need of further scientific research into the variables effecting cryptocurrency prices. In the following literature, the variables from Section 1.3 are investigated in more detail.

The risk of cryptocurrencies is one of the most discussed topics in many online communities. Liang et al. (2018) provide an analysis of the dynamics and systemic risks of the cryptocurrency markets. The authors collect publicly available market data from January 2015 until April 2018. Their data ends during one of the biggest retractions of cryptocurrency markets. As also discussed in Section 2.1, the authors have collected data during a time of increased uncertainty in the markets and results obtained from this data might not hold today. In their paper one of the results from their quantitative analysis was to label the cryptocurrency markets as "fragile and unstable" in early 2018, which in retrospect turned out to be correct.

Their methodology to turn daily price series correlation data to a connected graph of related currencies is very well described and can easily be followed. This method is likely a good starting point for repeated research with more recent data. In their correlation analysis they find Bitcoin forms the central node in the graph for half of their analysed days. This might be explained due to the fact that most other currencies used to be exchanged to Bitcoin before they could be exchanged into fiat currency. This pattern is still common today, although a repeated research using this methodology might find that Ethereum and a number of smaller cryptocurrencies can now also be used for this purpose.

In a different approach to understanding the market dynamics, especially of how new cryptocurrencies are being created and old projects are discontinued, ElBahrawy et al. (2017) apply a "neutral model of evolution" to cryptocurrency market data and are able to reproduce a number of key properties. "The number of active cryptocurrencies, market share distribution and the turnover of cryptocurrencies" are among the predicatable properties of their model, according to the authors. Bitcoin holds the top spot in terms of capitalisation over the analysed timespan. There are 33 cryptocurrencies fighting for position in the ranks 2-6, with an average time in the same rank of roughly 12 weeks. It is an interesting approach to describe the dynamics and competitiveness between cryptocurrencies in terms of their ranking. If the research was to be repeated today one would expect Ethereum to converge to the second rank, even though it did from time to time take the first place of largest capitalisation in the past.

The evolutionary model applied shows remarkable accuracy in modelling the capitalisation of the currencies. The dataset being analysed in the paper ends in May 2017. Shortly before the first major cryptocurrency price bubble. It could be interesting to repeat their analysis methodology on the tumultuous market data beginning in 2017 until today to confirm if the evolutionary model still holds. Overall the paper suggests that in the long term the importance of Bitcoin as the biggest cryptocurrency will reduce.

Shah et al. (2021) apply principal component analysis (PCA) to cryptocurrency pricing data with the goal of creating a robust cryptocurrency market index. An index helps investors avoid risks of selecting individual projects and alleviates the amount of effort investors have to spend in evaluating possible investments. The index is meant to track the movement of the market as a whole. According to the authors, most existing indexes do not clearly document their methodology and are based on subjective rules. The variables "consisting of market, size and momentum" (Shah et al., 2021, p. 2) are included in the mathematical model of PCA. Based on the resulting weights an objective index was constructed.

The resulting index demonstrates some interesting properties. While it does not always track the total market capitalisation optimally, it can be used to dynamically identify the most important markets at any given time. The proposed method allows for arbitrary specification of the number of components to be included, which if expanded could allow for a more complete, but also more complex, representation of the market. The paper proposes an interesting method and documents its application well.

While reviewing this paper, an additional application beyond the method of Shah et al. (2021) was identified. The scalar weights per market represent the relative importance of a particular market. Observing the weights of markets per cryptocurrency might allow for a comparison of the importance of a cryptocurrency at a given time. Project insiders might use this technique to compare their project to others. This should allow identification of the main project competitors and to track their relative success over time. Further investigation into the proposed approach could be beneficial.

Bejaoui et al. (2019) analyse the cryptocurrencies Bitcoin, Ripple, Litecoin and Ethereum market prices for cyclical behaviour and non-linear structures through the use of a Markov switching auto-regressive moving average analysis (MS-ARMA). The proposed method is complicated and difficult to follow, which makes interpretation of the results challenging. While the authors found evidence for the presence of statistically different states in the daily closing price data, it is difficult to imagine a theoretical reasoning able to explain all the observed findings. One finding is the presence of unusual spikes and trend reversals, which is proposed to be caused by short-term investors seeking overly excessive returns. This corresponds to my personal experience in markets.

Fundamentally the presence of non-linearities is demonstrated in cryptocurrency markets. The idea of multiple regimes is interesting to explore further. The main states found are similar to the commonly understood bull market and bear market conditions. While this is convincing, without a theory about when and why the markets switch states, the information of the presence of multiple market conditions is not of a high degree of usefulness to investors or developers. But it could form the basis of future research developing a theory about when markets change dynamic states.

Jiang et al. (2021) compare cryptocurrency as a financial asset to foreign exchange markets and stocks. The comparison investigates the "five properties: volatility, centrality, clustering structure, robustness, and risk" (Jiang et al., 2021, p. 1). One of the main findings is that the volatility of cryptocurrencies is considered quite high, but strongly anti-correlated to foreign exchange markets. The centrality analysis identifies Bitcoin as an originally highly important node, as it was necessary as an intermediate step to exchange between fiat currencies to cryptocurrencies. However, over time, other currencies have been taking this intermediate function as well and weakening the importance of Bitcoin for the market as a whole. The remainder of the analysis classify cryptocurrency markets as a risky investment with a continually changing composition of competing markets.

The paper confirms the general perception of cryptocurrency markets as risky. However, this was based on data from 2015 to 2018. The method used is a simple correlation analysis and might be repeated on the more recent data to check if the findings still hold 3 years later. While the analysis identifies the rapidly changing nature of the markets, it fails to propose a theory explaining the observed changes.

King and Koutmos (2021) approach their investigation from the novel angle of identifying herding behaviour caused by past price changes. This allows for insight into the psychology of human traders who are subject to natural human biases and behaviours. To find herding the authors analyse price data to identify if an "appreciation in the prior trading day, [is followed by] subsequent buying (i.e. trend chasing) or subsequent selling (i.e. contrarian trading)" (King & Koutmos, 2021, p. 2). One of their major findings is "that increases in volatility are associated with rises in prices and, on average, volatility is rewarded in the cryptocurrency market" (King & Koutmos, 2021, p. 3).

The authors provide a detailed explanation for their statistical evaluation of risk (represented with the Sharpe Ratio) and behavioural trends in the various currencies. The analysis shows mixed results. Some cryptocurrencies show herding behaviour, but most appear to be segmented still. The authors indicate, as the markets mature, more integration might be observed. However, this needs to be tested again with more up to date data. The authors point out that some market mechanisms, like the lack of margin trading might have helped in keeping the markets separated. Since the paper was published, margin trading has been introduced on various exchanges and different results are expected today if the analysis were repeated.

Altan et al. (2019) demonstrate a new hybrid model for cryptocurrency price prediction. The method first decomposes a price signal using Empirical Wavelet Transform, then applies a Long Short-Term Memory recurrent neural network for the prediction and finally uses cuckoo search optimisation for the Meta-heuristic optimisation of the combined model. The input data used is daily open, high, low and close ticker data for the 4 biggest cryptocurrencies at the time which were Bitcoin, Ripple, Dash and Litecoin. The proposed method is well documented and likely reproducible. The results obtained by the authors demonstrate a high degree of accuracy in the price prediction of the next days data based on the previous days data.

Their best model is capable of predicting the price value with an mean absolute percentage error of between 1.47 % to 3.55 %. This is impressive performance compared to other similar models. The value of such a system could lie in helping investors evaluate trends in the markets and select projects with a good expected future return. Interestingly the only variable used as input is the price. A basic economic principle is that supply and demand regulate the price and assuming a sufficient level of information available to both sides, the price is expected to converge at its correct value. In turn this implies information used by market participants is effectively "baked" into the price value. A sufficiently complex model might be able to make use of that contained information, which is demonstrated by the authors. Such a relationship indicates the price history is a highly useful variable to be considered in the estimation of future trends and volatility.

Eross et al. (2019) investigate the intraday behaviour of a number of cryptocurrencies. A major variable influencing the price of stocks is the opening or closing of traditional exchanges. When for example the NYSE opens there is often a bump in the trading volume. Based on the presented results, the authors conclude the relationships of returns, volume, volatility and liquidity of cryptocurrencies are inconclusive. There is a small observable effect, generally indicating the volume of the cryptocurrency is mostly affected by US and European traders. Generally this paper did not find its expected relationships, but has observed the important characteristic that the intraday behaviour of cryptocurrency markets changes with time, which needs to be taken into account when creating other models.

Giudici and Abu-Hashish (2019) analyse the correlations of the Bitcoin to USD market among 7 different exchanges and with a number of non-cryptocurrency markets, which are gold, oil, SP500, USD-Euro and USD-Yuan. A main finding is the volatility of Bitcoin is significantly larger than even the next most volatile market, the Bitcoin volatility is 20 times larger than the SP500 volatility. A second interesting observation is the importance of the price on one exchange being affected by that of another. While the prices in the short-term might be different between exchanges, this would create arbitrage opportunities and we would expect a correction of the price to the common value among the set of exchanges. The authors states it is important to consider multiple exchanges to get a true understanding of the value of a cryptocurrency at a given time. Based on the idea of the efficiently integrated market, prices between exchanges can be predicted based on prices from other exchanges and other exogenous markets. The authors suggest to investigate the reasons for deviations of an exchange to the expected efficient price value further.

Wang and Vergne (2017) state that much of the existing literature bases their analysis on the expectation that cryptocurrencies behave like traditional currencies, but their analysis shows the variable of "buzz" (hype) is a critical influence on cryptocurrency markets. Since the supply of a cryptocurrency is fixed algorithmically by its protocol design, the price is determined by the demand. Hype variables, like media coverage, could lead to short term surges in demand. However, in the long term the true innovation potential becomes the primary demand driver. By analysing news, market data and development statistics, they find that true innovation represented by Github metrics are a major driver in the long term returns. The public interest level they find to be negatively correlated with returns, which indicates their model does not support the idea of buzz leading to sustained price increases. This paper documents the method of analysing influencing variables well and especially their findings on the Github metrics justifies further research into this variable.

Baig et al. (2019) analyse the effect of Google Trends sentiment on Bitcoin price clustering. Clustering describes the effect of prices spending unusually much time unusually close to round number prices. The authors describe in their data to see trades executed at whole dollar prices more than 18 % of the time, while whole dollar increments account for only 1 % of possible values. The theory behind this behaviour is that it is costly (effort) for traders to specify more fine grained price values and during phases of high sentiment the trader is less concerned with such small differences in the short term compared to their long term expectations. The authors further remark that the presence of such price clustering might imply frictions in the market which prevent the price to converge to its equilibrium value. The results show a significant correlation of sentiment with price clustering.

Nosratabadi et al. (2020) conduct a systematic literature review of advanced machine learning and deep learning methods with practical economic applications. The authors find in their analysis stock price prediction using data science to be the objective of a majority of the analysed papers. Other studies applied sentiment analysis. Almost all papers use the time series of the prices as input with some considering the underlying companies reports as additional input source. Applying a wavelet transform before using an Long Short-Term Memory (LSTM) for prediction, significantly improved performance compared to using an LSTM alone. Analysing the wavelet transform of the price time series could be an interesting variable to analyse in this research as well.

Aggarwal et al. (2019) study the predictive quality of various social variables, like tweet sentiment score, number of followers of the tweeting user, number of retweets and likes, on the daily Bitcoin price. Aside from Twitter sentiment the authors use the gold price as an additional input for the neural network used to make the predictions. They the proceed to test a Convolutional Neural-Network (CNN), a LSTM and a Gated Recurrent Unit (GRU) model. While their analysis shows the sentiment to be a much more significant predictor than gold price, their dataset only covers daily data over a timespan of roughly 5 months. Based on the Author's experience with neural network based methods, this can be considered little training data. Nevertheless the idea of using gold as an indicator is interesting. And might inspire additional parameters, like other stocks, indices or resources to be included, e.g. SP500 and oil.

Abbatemarco et al. (2018) attempt to develop a measure for the price of Bitcoin based on its fundamental economics. Fundamentally distributed ledgers like blockchain are maintained by a network of computers, miners, providing computational power to process, verify and store transactions. In exchange for allocating resources the miners are paid with fees and block rewards. The cumulative processing power is measured as the hashrate. Abbatemarco et al. (2018) indicate the hashrate can be used as an estimator for true cryptocurrency value, as one would expect the processing power to deacrease or increase to the equilibrium point where the value of the obtained cryptocurrency matches the cost of mining for it. The authors find a correlation in their analysis, however, further research is required to determine if this relation is predictive. They conclude that it "is very likely that elements such as the greater acceptance of cryptocurrencies at a legal level, greater network diffusion, the birth and development of increasingly user-friendly cryptocurrencies wallets, are in the long-term Bitcoin price driving forces" (Abbatemarco et al., 2018, p. 11).

Panova et al. (2019) review the legal status of cryptocurrencies. They find in the early years, cryptocurrencies were seen strongly negative by governments around the world as it threatened one of the key sources of power for governments, which is to issue currency (Panova et al., 2019, p. 3). However, since then many governments have shifted their stance, as it became clear the nature of cryptocurrencies makes it nearly impossible to prohibit them. Instead there is a general trend to classify them similar to currencies of other countries and to levy income, corporate and capital gains tax against them to profit of the recent boom. There are still large grey-areas in regulations and vast differences between governments. A major requirement introduced by the EU is to oblige European cryptocurrency exchanges to identify their users (Know your customer), which might also be related to efforts in preventing money laundering (AML).

Panova et al. (2019) state in the coming years a a competition between national legal systems in terms of attractiveness for cryptocurrency related businesses will develop. Most governments appear to be awaiting further developments before taking more definitive actions. At the time of writing only Japan has recognised cryptocurrency as money. Most regulatory efforts appear aimed at preventing illegal activities. For the purposes of this thesis, the variable of regulation appears to not stand in the way of new business models and technical innovation. Measuring the state of national regulations over time is technically very difficult and will not be considered as a measurable variable in this research. This will be further discussed in Section 2.4.

Shanaev et al. (2020) perform a systematic analysis of the effect of news media reporting on regulatory events. They find anti-money laundering was among the most discussed topics followed by exchange regulation. However, mentioning regulation on the news media resulted on average in a slight reduction of cryptocurrency price. Similarly the mentioning of exchange regulations and issuance regulations resulted in a drop in market values as well. Overall their findings show the “market perceives regulatory events as value-destroying” (Shanaev et al., 2020, p. 9). They state, “government commitments not to overregulate cryptocurrencies and to let the industry develop in a freer environment can contribute to lower market volatility and more stable coin and token prices” (Shanaev et al., 2020, p. 9). However, due to the manual data collection approach used, only 120 events were analysed over a period of about a year.

Li et al. (2019) analyse Twitter sentiment based on hourly samples over a timespan of 3.5 weeks, a comparatively short timespan to analyse. Any results from this dataset has to be considered from that perspective. Using sentiment evaluation they assign a rating to each tweet which may have a value of positive, neutral or negative. As part of their analysis the authors observed a large number of automatically generated tweets in their dataset, however, most of them were given a neutral sentiment rating and were excluded. The approach of Li et al. (2019) might leave out some of the nuances that could be obtainable when using a more graduated set of values to use for the sentiment. Overall they find some predictive quality in the sentiment of the Twitter data for the prediction of cryptocurrency prices.

Rahman et al. (2019) aim to identify which types of emotion are the most significant in predicting bitcoin price. They base their analysis on the sentiment evaluation of tweets containing the word cryptocurrency. Additionally they use an initial worldcloud based on those tweets to come up with related and frequently used terms, they do not state the final set of keywords used, increasing the difficulty of replicating their research. They analyse both the total sentiment of each tweet and the sentiment per emotion of “anger, anticipation, disgust, fear, joy, negative, positive, surprise, sadness, surprise, trust.” (Rahman et al., 2019, p. 3). The authors use a binary representation for the price change, instead of a ratio value. The authors acknowledge their 500 tweets per day collected over only a number of weeks is insufficient for extracting significant conclusions. Their primary findings show emotions of anger and negativity are the most predictive from the set of emotions analysed. Their prediction only determines the direction of the trend but not the magnitude of the change.

Wołk (2020) uses tweet sentiment and Google Trends data to forecast short term cryptocurrency prices. The author used a hybrid model based on the mean prediction of 6 underlying methods, with support vector machine and simple neural network among them. The input data is based on 10 min intervals between datapoints. The overall prediction quality is promising, with the error being less than the usual daily volatility of the prices. An interesting observation is a price drop is followed by an increase in tweet frequency, indicating more people discussing the falling prices. Finally (Wołk, 2020) verifies the predictions by testing the method on a real portfolio with $100 on the BitBay exchange. Resulting in a gain of 14 % after the month. This indeed supports the idea of sentiment and trends data being useful for price predictions, however, the testing period is only a month long and due to the high volatility of cryptocurrency markets in 2018, the success of the method might be highly dependent on the specific month selected for backtesting. Overall the author finds a combination of Google Trends data and general negative sentiment to be the most powerful predictor.

López-Cabarcos et al. (2021) use the Standform CoreNLP software to evaluate sentiment of posts from the `StockTwits.com` platform. They find sentiment to be correlated and usable for price predictions. However, the SP500 was identified to be a more influential variable. The authors differentiate among different market conditions. When the stock market is volatile they find Bitcoin to be usable as a safe haven. When the stock markets are stable then Bitcoin can be used as a speculative investment. This is an interesting observation and we would expect to find evidence for this relationship to be found in the following research as well.

Thampanya et al. (2020) examine the effects of fundamental and behavioural variables on stock prices (not cryptocurrencies). According to the authors, investigating and comparing these variables simultaneously has only received little research. The authors compare the differences of the effect over five selected countries of Malaysia, Thailand, Singapore, Indonesia and the Philippines. Fundamental variables are designed to use risk-return considerations to judge potential investments. Behavioural variables on the other hand include influences of sentiment and psychological prejudices in their considerations. These variables are helpful in explaining some of the "irrational" trading behaviours observed in markets, e.g. herding (price bubbles) and loss aversion. The fundamental variables analysed are: Macro-economic indicators of "GDP, money supply, interest rate, inflation rate and exchange rate" (Thampanya et al., 2020, p. 3). Additionally coporate variables of balance sheets, cash flow information and other reporting about finanical status of a specific company are considered. As behavioural variables the authors analyse index composition for the percentage of firms less than 10 years old, with the argument a large number of young firms indicates herding behaviour.

The effect of each variable varies significantly among the stock markets of the 5 countries analysed. While in specific cases a strong correlation exists with the potential to have predictive uses, the variance among countries indicates the detected relationship might only be coincidental. The authors find "interest rates, exchange rates, and inflation rates are the key macroeconomic variables driving stock market volatility" (Thampanya et al., 2020, p. 23). Assuming the index composition to be usable as a proxy for behavioural variables may be questionable.

Akyildirim et al. (2020) investigate the effect of an announcement about potential interest in cryptocurrency projects by exchange listed companies (not cryptocurrencies) on their stock price. Due to the high levels of hype associated with the field, companies could expect a temporary stock price boost simply from such an announcement. The authors indeed find such a relationship. After correcting for other trends in the markets, companies that made such announcements outperformed their peers within several weeks. However, the difference disappeared again after about three months, indicating the "buzz" nature of the price change. This observation raises the interesting point about the moral issue of abusing the expectation of investors, as companies could obtain a "quick" profit for their shareholders by making cryptocurrency related announcements. The variable of publicly traded companies attaching their reputation to cryptocurrencies likely also affects the price of the connected cryptocurrency. Therefore company announcements could be considered as an additional variable in the analysis of the following research.

Liu et al. (2019) investigate variables influencing the Oil price. While this study is not directly related to cryptocurrencies, it has a similar research objective as this thesis, however, applied to a different asset. Their main finding is the oil price is in the long-term driven by the fundamentals, in the short-term this does not hold. The presence of an oil futures market changes the dynamics of the price and can lead to coupling. They conclude, a model to explain the short-term dynamics of the oil price needs to take into account structural changes, macroeconomic effects on the transmission of information and behavioural finance. Similar to oil, Bitcoin also has a futures market associated with it. This could be an additional variable to consider in this research. Although it is not clear if the futures market is affected by the current price, or if the current price is affected by the futures market.

2.3 Literature research findings

Existing literature analyses a range of different variables. The set of variables, as encountered in the literature reviewed in the previous Section 2.2, is summarised in Table 6. Literature uses a range of different methodologies to analyse the variables. The methodologies are summarised in an overview in Table 7. In the following Section 2.4 the variables to be analysed in this research are selected.

The literature applies a wide range of methods to analyse various factors. The majority of existing research focuses on the narrow effect of a single selected variable. This might result in model error, as the multivariate effects are unaccounted for. The price itself is one of the most commonly researched variables. This is likely the result of broad price data availability. However, models build only on the price variable are often highly mathematical in nature and at times act as "black boxes". This makes interpretation and discussion of the observed relationships between variables difficult.

Another observation from the literature research is the commonly used differentiation between intra-day effects (variation of timescale smaller than 1 day) and inter-day effects (variation of timescale larger than 1 day). This may provide additional insight into the nature of the variables.

According to the reviewed literature the first research question is answered by the set of variables in Table 6. These variables have been identified to have some influence on cryptocurrency prices. The direction and magnitude of the variable effects will be the focus of the remaining analysis of this thesis.

Table 6: Summary of variables used in literature.

Variable	**Literature**	**Theory**
Price	(Liang et al., 2018), (ElBahrawy et al., 2017), (Shah et al., 2021), (Bejaoui et al., 2019), (King & Koutmos, 2021), (Altan et al., 2019)	In efficient markets all information relevant to the market is contained within the price.
Volume	(ElBahrawy et al., 2017), (Shah et al., 2021), (Bejaoui et al., 2019)	More volume indicates more efficient markets and allows for comparison between markets.
Volatility	(Jiang et al., 2021)	Using a statistical measure to capture the amount of variance in price.
Risk	(Jiang et al., 2021)	Using the sharp ratio as statistical measure for the risk level.
Empirical wavelet transform	(Altan et al., 2019), (Nosratabadi et al., 2020)	Decomposing a signal, e.g. price, into its component frequencies.
Opening times of traditional exchanges	(Eross et al., 2019)	When exchanges open or close the set of traders changes in markets resulting in a change in market dynamics.
SP500, Gold, Oil	(Giudici & Abu-Hashish, 2019), (Thampanya et al., 2020)	SP500, Gold or Oil prices can be used as proxy indicators for macro-economic state.
Github insights	(Wang & Vergne, 2017)	Github metrics can be used as proxy indicators for fundamental value generation in open source software projects.
Hashrate	(Abbatemarco et al., 2018)	Hashrate is a measure for fundamental value in cryptocurrencies.
Google Trends	(Baig et al., 2019)	Sentiment score on a topic over time
Price Number Roundness	(Baig et al., 2019)	A statistically unusually large number of trades are executed at round values, indicator for bot vs. human trader presence.
Twitter insights	(Aggarwal et al., 2019), (Li et al., 2019), (Wołk, 2020)	Tweet sentiment score and meta-metrics can be used as indicator in hype-driven markets.
Regulatory state	(Panova et al., 2019)	Regulatory events likely affect prices.
News	(Shanaev et al., 2020), (Akyildirim et al., 2020)	Regulatory and other events mentioned in news-media likely affect prices.
Futures market price	(Liu et al., 2019)	Futures markets can affect the prices and there may be a reciprocal relationship.

Table 7: Methods used in literature and their input variables

Analysis method	**variables**	**Literature**	**Usecases**
Correlation analysis	price, volume, volatility, risk, Github insights	(Liang et al., 2018), (Jiang et al., 2021), (Wang & Vergne, 2017)	Discover relations between markets and other variables, creating a connected network graph.
Auto correlation	price	(King & Koutmos, 2021)	Measure correlation with the same signal but shifted in time, can indicate herding/trend chasing behaviour.
Neutral model of evolution	price, volume	(ElBahrawy et al., 2017)	Predict market capitalisation based on theory of evolution.
Principle component analysis	price, volume	(Shah et al., 2021)	Create objective market index.
Markov Model	price, volume	(Bejaoui et al., 2019)	Create multiple statistical regimes.
Empirical wavelet analysis	price, empirical wavelet transform	(Altan et al., 2019), (Nosratabadi et al., 2020)	Analysing the fundamental cycles in the price signal.
Intraday event analysis	price, exchange opening times	(Eross et al., 2019)	Identifying recurring changes in markets based on exchanges opening vs closed times.
Correlation with Macro-Economic trends	SP500, Gold, Oil	(Giudici & Abu-Hashish, 2019), (Thampanya et al., 2020)	Understanding how cryptocurrencies markets are affected by macro-economic trends.
Correlation with sentiment	Google trends, Twitter insights	(Baig et al., 2019), (Aggarwal et al., 2019), (Li et al., 2019), (Wołk, 2020), (López-Cabarcos et al., 2021)	Cryptocurrency markets are strongly driven by hype and buzz, which can be measured using sentiment analysis.
Sentiment correlation with round prices	Sentiment, prices	(Baig et al., 2019)	Measuring the effect of sentiment on roundness of prices. Are markets driven by rational investors or by buzz?

2.4 Variable Selection

In this thesis one of the objectives is to compare the magnitude and direction of effects of variables on cryptocurrency price volatility. So far in existing literature, single variables have been investigated individually. The narrow applicability of their results makes comparison between variables difficult. To overcome this problem, in the following sections an analysis shall be performed on a number of selected variables simultaneously. The distinction of intra- and inter-day effects will also be made, since the literature research findings (Section 2.3) indicate a difference of dynamics. This is further specified in Section 3. For now it suffices to specify, the analysis will be performed for a first dataset with datapoints being spaced 512 seconds apart and a second daily dataset with datapoints spaced 86 400 seconds apart.

The analysis will be quantitative in nature, applying the method of a Hierarchical Bayesian model. The reason for selecting this method is described in Section 3. Each variable shall be required to be measured in time series data of similar granularity (resolution in time). The method further requires a datapoint for each variable to be available at all time-ticks. This allows the following selection criteria to be defined for a variable to be included (inclusion requirements) in this analysis:

1. The variable needs to be continuous and smooth in time, discontinuities and gaps should be avoided.
2. The data for the variable needs to be available at the corresponding granularity (512/86400) being analysed for all markets.
3. The data needs to be obtainable in real-time or retroactively for the time range under analysis through an API.
4. Computed variables should have either an easy to implement algorithm or a freely available off-the-shelf software implementation.
5. The variable should describe an external property of the cryptocurrency market or its surrounding network.

If a variable fails to fulfil one or more of the above listed selection criteria it will be excluded from the analysis.

Before the variable selection is applied, there are potential gaps in the existing research. Based on the author's past experience of following cryptocurrency markets, there are a number of variables that could have a significant influence, but were not found to be included in the analysed literature. A set of additional potentially significant variables is shown in Table 8 and a reasoning for inclusion is given for each additional variable.

At this point a total of 18 variables could be identified. Next the selection criteria 1 to 5 will be applied to each variable. This will result in the final set of variables to be analysed. Table 9 contains variables that were excluded and the corresponding reasoning. Most exclusions are the result of a lack of sufficiently granular data or API availability.

After the filter process there are 8 remaining variables. The resulting set of variables will be used for the main analysis and is shown in Table 10.

Research question 1 had the objective of identifying variables relevant to market dynamics and price volatility. This has been answered by the combined set of variables listed in Table 6 and Table 8.

Table 8: Novel variables missing from the reviewed literature

Variable	Theory
Community Size	Community size could be a fundamental variable influencing the speed at which software issues are discovered. The total amount of capital available for investment as well as the ability to market the cryptocurrency through word of mouth is also a function of community size.
Key influencers	In the social networks there are key individuals that could influence the opinion of large followings.
General levels of internet activity	There are cycles of activity levels every day while different time zones are asleep. This likely affects market cycles as well. Some timezones might be more influential than others.

Table 9: Variables rejected from analysis, according to the inclusion requirements 1 to 5.

Variable	Reason for exclusion
Price	Endrogenous variable, we want to estimate this. This variable does not fulfil inclusion requirement 5.
Volume	Not an external variable of the market. This variable does not fulfil inclusion requirement 5.
Risk	Risk is fundamentally very difficult to define in cryptocurrency as not all affecting variables are understood yet. Mathematical definitions are similar to variable 3. Volatility. There is not much use in including risk in this research at the current state of knowledge about cryptocurrencies. This variable does not fulfil inclusion requirement 4.
Empirical Wavelet transform	Not an external variable of the market, is derived from price. This variable does not fulfil inclusion requirement 5.
Opening times of exchanges	This variable has large discontinuities and the opening times have no meaning on daily data level. This variable does not fulfil inclusion requirements 1 and 2.
Hashrate	Data not available at sufficient granularity for selected markets. This variable does not fulfil inclusion requirement 2.
Price number roundness	This is a derived variable based on price. This variable does not fulfil inclusion requirement 5.
Regulatory state	This is highly discontinuous both in time and geographically. There is no easily accessible time-series data source. Variable violates inclusion requirements 1. and 3.
News	While interesting theoretically, there is no easily accessible time-series data source. Variable violates inclusion requirement 3.
Futures market price	The corresponding data is not accessible through a free API at sufficiently high time granularity. Variable violates inclusion requirements 2. and 3.

Table 10: Final set of variables selected for analysis in this reserach

Volatility	Twitter insights
SP500, Gold, Oil	Community Size
Github insights	Key influencers
Google Trends	General levels of internet activity

3 Methodology

In Section 1.3 an initial set of variables was established. Based on literature research findings, additional variables missing from literature were added (Section 2.4). By applying appropriate theories, hypothesis for the relationship of the influencing variable on the price will be established in Section 3.1. Then the analysis method is outlined (Section 3.2) and finally the various data sources and corresponding data pre-processing steps are described (Section 3.4).

3.1 Hypothesis

The research questions (Section 1.2) and selected variables (Section 2.4) are combined to result in a set of hypothesis to be tested in the following analysis.

Cryptocurrency price is the endogenous variable which the model will try to estimate using the exogenous variables from Table 10. In the following the term price will be used as a short-hand for cryptocurrency price. In an efficient market any opportunity for arbitrage is immediately exploited and the price is assumed to converge to the true value. However, this true value may change over time, as for example the underlying cryptocurrency project is creating new useful features and thereby increase fundamental value to its users. The exogenous variables are assumed to be drivers to changes in the fundamental value of the market.

Volatility is a statistical measure for the amount of variance in the price value. It is not entirely clear how volatility relates to price and the relationship might differ between various cryptocurrencies, however, this research might shed further light on this relation. Inspired by the herding behaviour described by King and Koutmos (2021), we establish the volatility hypothesis as:

H1: When Volatility increases, the price increases.

SP500, Gold, Oil can be seen as indicators of the macro-economic state (Thampanya et al., 2020). The SP500 is an index tracking the 500 large US companies and is widely considered the best indicator for the overall economic state of the US market. In the past, cryptocurrencies have been marketed as an alternative to the existing financial system. Bitcoin has been referred to as an alternative to Gold. Oil is an ingredient for various products and one of the primary fuels used around the world for transport and energy generation. As mining cryptocurrencies is rather energy intensive (De Vries, 2018) there could be a relationship to the oil price. We establish the following macro-economic hypothesis:

H2.a: The cryptocurrency market prices decrease when SP500 increases.

H2.b: The cryptocurrency market prices decrease when Oil price increases.

H2.c: The cryptocurrency market prices increase when Gold price increases.

Github insights can help understand cryptocurrencies projects in a variety of ways. A majority of these projects are open source and make their code publicly available. This leverages the benefits of open source software development through freelance developers contributing code improvements and the public being able to report on bugs and other problems easily. From an analysis point of view, statistics about the code contributions of programmers can give insight into the pace and scale at which a project is improving its code base. Based on the author's experience from monitoring various open source projects in the past, the following relationships are proposed. While issues can be opened by either developers or users, more issues closed indicate active programming efforts. However, If the difference between issues opened and issues closed is too large, this may indicate the developers are slow to address those issues on the short-term. The metrics of Github stars is comparable to likes and indicates the popularity of the code overall. Similarly forks occur when other developers take the current state of the software as starting point for their own copy with their own modifications. Akin to the saying "Imitation is the sincerest form of flattery", the more people fork the project the more popular it could be considered. The following Github hypothesis are established:

H3.a: When the number of closed issues increases, the price increases.

H3.b: When the number of open issues decreases, the price increases.

H3.c: When the number of stars increases, the price increases.

H3.d: When the number of forks increases, the price increases.

Google Trends is a sentiment score provided by Google based on the frequency of searches related to a topic. Before someone might invest in a particular cryptocurrency they are likely to search for background information on it using the most popular search engine Google. Based on the findings of Baig et al. (2019):

H4: When the Google Trends rating increases, the price increases.

Community Size is a big driver from a technical development and from a marketing perspective. Many cryptocurrency projects are primarily marketed through word of mouth. From the author's observations: A larger community likely has more reach and can convince more investors to invest. The community hypothesis are:

H5: When the community size increases, the price increases as well.

General levels of internet activity follows a cyclical pattern. As the earth rotates different timezones are awake or asleep at any given time. This changes the subset of the population active on the internet. As a result also the set of traders of cryptocurrency changes. At times of peak internet activity we would also expect increased trading activity.

H6: An increase in global Internet activity corresponds to an increase in price.

Twitter insights are a way of obtaining sentiment about cryptocurrencies. By using natural language processing tools, the sentiment of a statement can be determined. Similar to the hypothesis by Aggarwal et al. (2019), Li et al. (2019) and Wołk (2020):

H7: When Twitter sentiment becomes more positive, the price increases.

Key influencers are especially powerful in spreading the word about a specific cryptocurrency project. These key nodes in the social networks are typically identified by a large following. When they change opinion their followers are likely to agree. Compared to the group as a whole, the sentiment of key influencers might be especially predictive.

H8: When key influencer sentiment becomes more positive, the prices increase.

3.2 Statistical Analysis

The goal of this research is to investigate the magnitude and direction of the effect of the selected variables on the price of cryptocurrencies. Section 3.1 established a set of hypothesis to be tested. The statistical analysis follows the method by Gelman (2006) and the applicability of the method is described in the following.

Cryptocurrencies are not all the same. There are varying magnitudes of market capitalisation and different communities associated with each cryptocurrency. Furthermore, different currencies solve different technical problems. For example Bitcoin is at the time of writing a currency and store of value, while Ethereum is a smart contract[3] system and a currency. The different markets might not behave all the same, since the underlying assets are different from each other. If the markets are dependent or in-dependent from each other is not clear at this stage.

Each selected variable could be classified into one of the following three groups. In the first group the variable effects all cryptocurrency markets equally. In the second group the variable effects each market differently. In the third group the variable has the same effect for a group of markets, but a differing effect between groups.

Generally three analysis approaches are possible with regards to accounting for the differences between effect groups. Each market could be analysed on its own in an unpooled model. The unpooled model assumes all markets to be independent of each other. Alternatively all the data could be analysed as if it came from a single market, being a pooled model. Which would assume no differences to be present between markets. The third approach is to employ a hierarchical model, which forms a hybrid of the other two approaches. For this research the third approach is the most sensible, it allows the greatest flexibility without having to make arbitrary assumptions about the nature of a variable effects and markets.

Gelman (2006) describes the hierarchical Bayesian regression as a generalisation of regression methods. In a hierarchical model the effect of a variable is considered to be composed of a component affecting all markets and a component specific to an individual market. For example if it is a variable affecting all markets equally then the individual market component would converge to zero. This allows the model to be applied on data where a mixture of effects is expected to be present, as is the case in this research. Bayesian models can be applied without having to have prior knowledge about the nature of the effects and avoids making unnecessary assumptions.

The Hierarchical Bayesian Linear Regression model is based on the Bayesian viewpoint. Each coefficient is represented by a probability density distribution, in contrast to the frequentist approach of an Ordinary Linear Regression where only a single value is obtained for each coefficient. The parameters describing the probability distribution are in turn defined to be part of a higher order prior distribution which is the same for all markets. The probability distributions for each variable per market and the hyper-prior of each variable are the result from the Bayesian model. For each market the variables' probability distributions can be compared based on their mean, their standard deviation or their significance. This can be visualised using forest plots to allow for clear and easily interpret-able presentation of results (see Section 4.3).

Price is the endogenous variable which the model is trying to estimate. Each variable is taken as an input and multiplied with a coefficient distribution, β_n. The price estimate is

[3] Smart contracts allow arbitrary code to be executed by the network.

then defined as the sum of all the products of the coefficients and the corresponding variable plus some constant, the intercept. The resulting estimate is compared to the known true value of the price. The difference between them is the estimation error, ϵ. Eq. (3.1) shows the corresponding equation for a single market.

$$\text{price} = \text{intercept} + \sum^{n} \beta_n \cdot \textit{variable}_n + \epsilon \tag{3.1}$$

The Eq. (3.1) is applied to the price and variables for each market. The model will contain a selection of 18 markets. The market selection procedure is described in Section Section 3.3. The Bayesian model requires a prior distribution to be defined for all of the coefficients. The prior distributions are taken to be normal distributions centred around some μ with some standard deviation σ. Each market has its own coefficient distributions, but sample their hyper-parameters μ and σ from a shared distribution per variable. The hyper-priors are defined to be a normal distribution with parameters $\mu = 0$ and $\sigma = 1$ for μ and a half normal distribution (only positive values) with the same parameters for σ. An additional control variable is included, the market volume.

Since the markets have differences in amount of trading volume this could effect the nature of the relationships the exogenous variables have on the price. To account for this, the analysis is first done with the control variable and then repeated without it. The difference between the models accuracies can then be compared to quantify the impact of the control variable. The effects of the control variable and of the exogenous variables are discussed in Section 4.3.

To do the numeric evaluation of the model, the Python module ***PyMC3***[4] will be used. It makes use of a Markov Chain Monte Carlo inference algorithm. The general algorithm can be described as follows. First a set of possible coefficient values is sampled from the prior distribution. Then the coefficients are evaluated and the weight of the coefficients in the probability distribution is adjusted according to the error ϵ to the observations. By repeating this process over and over, the probability distribution is gradually adjusted to converge to the posterior distribution that best describes the model with the given data.

3.3 Market Selection and Price data

The exchanges `Kraken.com` and `Poloniex.com` were selected as data source for the market prices, as a result of availability from the `Arkmon.eu` market monitoring system. `Arkmon.eu` was selected as data provider for the pricing data, since it is a project the Author has founded and access was available to the `Arkmon.eu` pricing data which is monitoring markets at a high resolution otherwise only available at high cost from professional data brokers. To limit the complexity of this analysis a selection of cryptocurrencies was made. First all markets were select which were available on both `Kraken.com` and `Poloniex.com`. Next a minimum volume requirement was applied. Markets with very small volume likely are subject to high levels of noise and might not lead to useful results. For each market it was checked if at least an equivalent volume of 20 Bitcoin was traded in the last 8192 s (which is the largest time-frame natively supported by the `Arkmon.eu` backend systems). Next markets were filtered for all markets where a cryptocurrency is denominated with respect to EUR, USD or a stable coin pegged to USD. The stable coins are DAI, USDC (USDCoin) and USDT (Tether). Finally the remaining markets were filtered again by requiring data to be available for all exogenous variables. The selection was executed on Tuesday 27th April, 2021 12:13:52 UTC, resulting in 18 markets selected for analysis:

[4] `http://docs.pymc.io/api.html`

["ADA_EUR", "ADA_USD", "BTC_DAI", "BTC_EUR", "BTC_USD", "BTC_USDC", "BTC_USDT", "ETH_EUR", "ETH_USD", "ETH_USDC", "ETH_USDT", "LTC_EUR", "LTC_USD", "LTC_USDT", "XLM_USD", "XRP_EUR", "XRP_USD", "XRP_USDT"]

Within the selected markets there is some overlap of cryptocurrencies. The label commonly used is called the ticker symbol. The corresponding full names are found in Table 11. Throughout the remainder of this report the ticker symbol or the full name are used interchangeably.

Table 11: Currency Symbols and Fullnames

Ticker symbol	Full name
ADA	Cardano
BTC	Bitcoin
ETH	Ethereum
LTC	Litecoin
XLM	Lumen
XRP	Ripple

The data obtained from the two exchanges is a live feed of every trade occurring on the tracked markets. Each trade contains information about the volume traded and the price at which the trade occurred. Converting a sequence of trades that have occurred into a representative value is a problem that the financial industry has solved in the past through various approaches. Commonly this is achieved using a ticker. Ticker data is provided by many exchanges at various intervals between datapoints. Within each interval information is returned about the open, high, low and close price. These correspond to the prices at which the first trade, the highest trade, the lowest trade and the last trade occurred within the interval. Computationally this is very cheap to do and has been used since the beginning of computerised trading, however, it contains only information from at most 4 trades in the interval. Information from any additional trades is simply thrown out. The `Arkmon.eu` system approaches this problem in a different way by taking the volume weighted average price of all trades which have occurred within the interval. This means that an informational contribution from each trade is contained within the final price value representing the interval. The volume weighted average price also combines the trades from the two exchanges being monitored in a way that accounts for variance in trading volume between them.

As part of data pre-processing any gaps that might occur in any of the market price data are filled through linear interpolation in time.

3.4 Data collection

This section describes for each variable how appropriate data is sourced and pre-processed. To perform the analysis we need time-series data for each variable. Each data series should ideally have the same number of datapoints and the datapoints should be equally spaced over the time range analysed. In the case of data-points being available at a resolution higher than needed, the variables with the higher granularity will be down-sampled to obtain the correct granularity.

Wołk (2020) and Altan et al. (2019) indicate that there are different market dynamics when considering inter-day (over multiple days) and intra-day (within the same day) effects. To explore this as well, two different time-series will be used at a different time interval between datapoints.

Intra-day effects can be analysed with data spaced at 512 seconds between datapoints. From past experience this was found to be a good compromise between noise in the time series data and having sufficient granularity in time. The current best performing `Arkmon.eu` AI trading system makes use of the 512 seconds data-spacing, as through trial and error it was found to have the best predictive quality. This is likely the result of the signal to noise ratio changing. At higher granularity the frequency and amplitude of noise in the price signal may become too large to extract reliably useful information and at lower granularity some information that could have been used might be lost due to averaging. The 512 seconds spacing should furthermore allow other time series to match the data spacing, as not all data sources are available at very high granularity in time.

Inter-day effects will be analysed with data spaced at 86 400 seconds between datapoints. Where daily data is not directly available it can be obtained through appropriate weighted averaging of the underlying higher granularity data. The higher granularity data at 512 seconds between datapoints will for brevity be referred to as the 512 data and the lower granularity daily data will be referred to as the 86 400 data.

For all variables open data sources were used. Some of the data-streams, like trade- or tweet-feeds are available in real-time only. They require continuous monitoring over the time-range under analysis. But in principle anyone with an always running computer can have access to this data. Trades are broadcast without any identifier of the parties involved in the trade. Tweets do contain the posting accounts user-name and id. However, this data is broadcast by the posting user themselves and every reasonable Twitter user can be assumed to be aware that they are not acting anonymously when posting Tweets. Any of the activity on the Twitter platform is in the public domain and is accessible to anyone else without restriction. For the purposes of this analysis the sentiment data of many users is aggregated and averaged into a singular datapoint. Only the aggregated sentiment was used for the analysis, following the Data Minimisation principle. The analysis methodology complies with Twitters terms-of-service. No ethical issues were encountered in the data collection phase.

Volatility

The volatility is defined as the standard deviation of the logarithmic rate of return as described in Eq. (3.2). This variable is calculated from the price data and helps test hypothesis 1. The value of a standard deviation increases with the number of datapoints used in its calculation. As we would like all data series to have a data spacing of 512 seconds per datapoint, we need to use a time-series with denser datapoints to determine the underlying volatility within a 512 second interval. The `Arkmon.eu` Market Tracking system also records the volume weighted average price at 32 second spacing between datapoints. The 32 time-series contains 16 data points for each sample from the 512 time-series. This will allow a standard deviation to be calculated for each datapoint in the 512 time-series.

$$\text{Logarithmic rate of return:} \qquad r_{log} = \frac{\ln\left(\frac{x_2}{x_1}\right)}{t} \tag{3.2}$$

SP500, Gold, Oil

To obtain the macro economic data Yahoo finance was chosen. It is the only data provider offering data at sufficient granularity free of charge. Other data providers, like NASDAQ, exist but would incur a large fee for the required data. The hypothesis 2 will be tested using this data. Yahoo finance provides data at 2 minute spacing for the last 60 days. A python software was written to update the newest datapoints every 4 hours and combine all the obtained data into a continuous *.csv* file. The oldest datapoint dates back to Friday 19th March, 2021 04:00:00 Eastern Time Zone. The provided data is ticker information, so open, high, low, close and volume data. The volume column will be used to re-sample the price columns to a volume weighted average price. To interface with the Yahoo API the python library ***yfinance***[5] will be used.

The selected variables (Section 2.4) require the data from the Gold, SP500 and Oil markets in terms of USD corresponding to the tickers of *GC=F*, *SPY*, *CL=F* respectively. Complementary to the intra-day data, data spaced at daily intervals was obtained for the same time-range. Fig. 5 shows an example of the macro economic data re-sampled to the 512s interval. The gaps in the graphs are the times when exchanges are closed and no trading occurs. This is one of the main difficulties of using conventional financial markets in relation with cryptocurrency markets. Exchanges, like NASDAQ or NSYE, have specific opening times during which trading occurs and are closed on bank holidays. cryptocurrencies on the other hand can be traded 24-7 during the entire year. Gold and Oil are traded on multiple exchanges around the world, so a price value can be obtained for example from NASDAQ Dubai while NASDAQ New York might be closed. However, SP500 is based on companies exclusively traded in US based exchanges. Fig. 5 shows that in a given week there is more time without any SP500 trading than with. This problem is addressed by using linear interpolation between on all gaps in data.

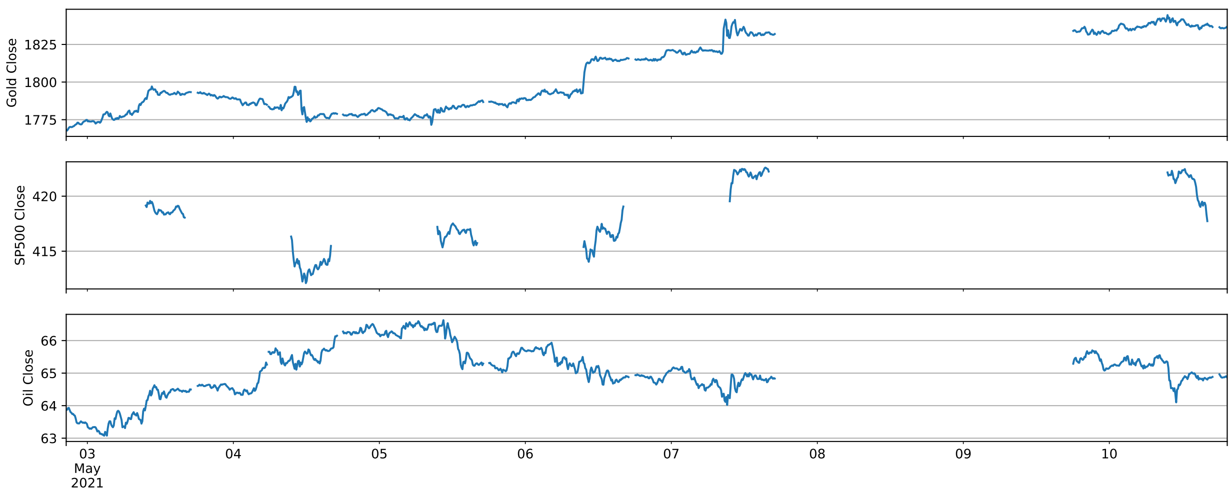

Figure 5: A Week of market data for Gold, SP500 (SPY) and Oil from Monday 3rd May, 2021.

Github insights and Community Size

Hypothesis 3 requires data for the development progress and technical maturity of a cryptocurrency project. Hypothesis 5 requires data about the community size of the "fan club" for a particular cryptocurrency. While manual data aggregation would have been possible by directly interfacing with the various Github projects and social media APIs, there already exists a data aggregation platform that collects the variables needed to test hypothesis 3 and 5. For the development statistics and community size the data from `Coingecko.com` is used.

[5] `https://pypi.org/project/yfinance/`

It aggregates the number of users and activity statistics from `Reddit.com` and `Twitter.com`. This data is obtainable at a "daily" resolution. To reach higher resolution real-time tracking would have to be employed, which is left as an improvement opportunity for future research. To interface with the `Coingecko.com` API the Python module ***pycoingecko***[6] was used. And a script was created to update the data once per day. A range of datapoints is obtained from which the following are selected for analysis:

1. *Social media statistics*:

   twitter_followers, reddit_subscribers

2. *Development (Github) statistics*:

   forks, stars, total_issues, closed_issues, pull_requests_merged, commit_count_4_weeks

The cryptocurrency projects analysed in this research have made their source code openly available to the public. By tracking changes to this code repository, the technical maturity of the project can be determined and rates of progress can be compared between multiple projects. Based on the author's personal programming experience, the above listed fields were selected as being likely significant indicators for the technical maturity of the cryptocurrency. The metric of number of open_issues was not directly obtained from `Coingecko.com`, but could be calculated by subtracting the number of closed_issues from the number of total_issues.

Google Trends

Google trends is a data source provided by Google, to measure the relative interest people show in a particular search term based on the search queries made through the Google search engine. Changes in the amount of searches made for a particular topic can indicate a change in public interest (Choi & Varian, 2012). This data source is used to test hypothesis 4.

Google trends data was accessed using the Python module ***pytrends***[7]. The main complication of collecting Google trends data was the strict rate limit. Google automatically selects a data spacing depending on the requested time-range. When requesting 4 hours of data, the data would be returned at 1 min resolution, however, at any time-range larger than 4 hours up to 7 days the data was returned at hourly and beyond that at only daily resolution. To obtain a time series at sufficient granularity many requests had to be executed while shifting the time-range of data being requested.

This introduced the next problem: Google automatically normalises the returned data by dividing by the maximum value in the interval and multiplying by 100. This means the data obtained from different requests cannot be compared. To solve this problem the requested data ranges were shifted in such a way that at least 1 datapoint would overlap between the data ranges obtained. The overlapping datapoint is part of both time ranges for which data is requested and should have the same value in either. This allows the newly obtained data to be scaled to the already recorded data.

Shifting the window over which data was to be obtained by 4 hours every-time with at least 1 datapoint of overlap allows for a continuous Google trends data series to be obtained. A downside of this solution is that the data needs to be uninterrupted, and once a 0 value is reached no older data can be obtained anymore since it cannot be scaled to the overlapping value, as a divide by 0 would be encountered. Another downside of this approach is the many requests needing to be made. Google limits their API to only return 1 result every

[6] `https://pypi.org/project/pycoingecko/`
[7] `https://pypi.org/project/pytrends/`

90 seconds. Exceeding this limit would result in a temporary block of further requests. The data collection for this process took approximately 2 months.

The data going back to the Monday 1st February, 2021 is obtained for the selected currencies (see Section 3.3) as far as possible. As a last step the data is divided by the maximum and multiplied by 100 to scale the resulting time series.

General levels of internet activity

Hypothesis 6 requires data for a variable representing global Internet activity levels. Surprisingly it turned out that obtaining a metric for the general levels of internet activity at any given time is a rather complex endeavour. The sheer almost unimaginable scale of Internet activity levels defeats even the most sophisticated monitoring. There simply is no measure which captures the entirety of the internet. The closest to it might be traffic statistics from the companies forming the backbone of the Internet, however, those are very secretive with their data and do not make it publicly accessible. So a new idea was needed.

From my observations of providing computer support to friends and family, I have noticed a common pattern of people accessing the Internet to be through the use of Google. However, often times people access the search engine Google not with the direct URL, but by simply typing Google into their search bar, which usually is configured to use Google and so as a result there are many searches on Google for the search term Google.

To confirm this anecdotal theory, Fig. 6 shows the Google trends rating for the search term Google trends of the USA. A daily cycle is clearly visible with a low on the weekend. The daily uptick corresponds to the time when the working day usually starts. The assumption is made that this metric may be used as a rough indicator for the levels of internet activity.

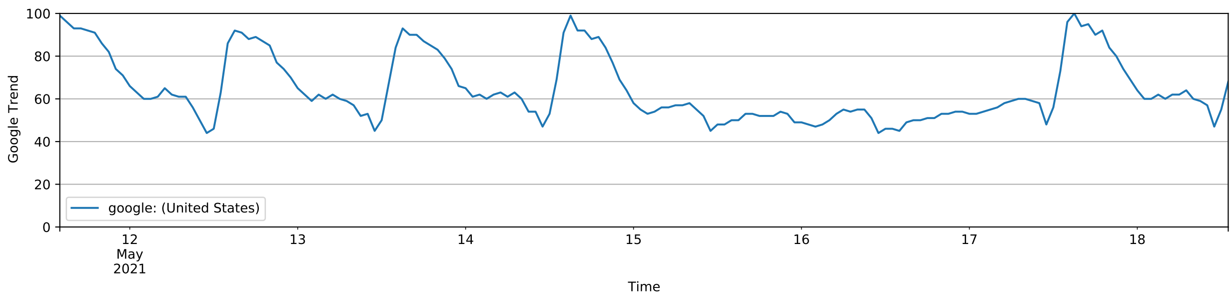

Figure 6: Google trends score for search term Google in the USA

The same data aggregation implementation was used as was previously implemented for the selected cryptocurrencies. Now the search term "google" has been added to be monitored.

Twitter insights and Key influencers

A second source for the public's sentiment is social media. Twitter is one of the most popular social media sites and changes in sentiment have been shown to be predictive for cryptocurrency prices (Aggarwal et al., 2019), (Li et al., 2019), (Wołk, 2020). The data for this variable will be used to test hypothesis 7 and 8.

On Twitter there is a constant stream of posts. Each tweet comes with many types of meta data and the text posted. The freely available Twitter API limits returned tweets to at most two weeks old. So to obtain data for more than two weeks real-time logging of tweets is required. To connect with the API the ***tweepy***[8] Python module was used. It connects to

[8] `https://www.tweepy.org/`

the web socket data streamer from Twitter which allows filters to be specified and then each tweet which matches a filter will be broadcasted to the listener. Any content on Twitter, according to their terms-of-service[9], is considered to be in the Public Domain and made freely accessible to anyone. All active Twitter users have by definition agreed with the terms-of-service and therefore no individual consent is required to record the users' Tweets. The data collection of Tweets and the following analysis is implemented in accordance with the Twitter terms-of-service.

The key-word filter obtains any tweet containing any of the words: "blockchain", "crypto", "cryptocurrency", "bitcoin", "ethereum", "cardano". Additionally the user search API was used to obtain some of the most popular Twitter users, and the 1000 users with the most followers (influencers) are being tracked as well, regardless of their connection to cryptocurrencies. Appendix A contains the complete list of user ids being tracked. Any tweet from those 1000 users was stored. Before analysing the influencer tweets, they were filtered. Only tweets containing any of the keywords from the previous filter were included in the influencer sentiment analysis. This ensures relevancy to cryptocurrencies, since most influencer tweets were concerned with different topics. A total of 288 207 996 tweets have been recorded to database in addition to 2659 cryptocurrency related tweets from the selected influencers. The metrics of *follower count* and the *friends count* are especially useful. The follower count is defined as the number other users following the tracked user. The friends count is defined as the number of users being followed by the tracked user. The main content of the tweet is the text.

Each tweet contains a large range of keys containing information. A small subset of available tweet meta-data relevant to this research is shown below:

```
{'created_at': 'Mon May 17 18:38:10 +0000 2021',
 'id': 1394361618994192385,
 'retweeted': False,
 'text': 'RT @[redacted] Watch this! And think about #Xpos &amp; #BOB phone '
         '&amp; #functionX blockchain and understand how privileged  you are '
         'to know about...',
 'timestamp_ms': 1621276690611,
 'user': {'created_at': 'Sun Mar 21 18:56:05 +0000 2021',
          'followers_count': 19',
          'friends_count': 54}}
```

However, getting the tweets is only part of the story. The raw tweet text needs to be preprocessed into a sentiment value. Various off-the-shelf modules exist for this purpose. Some experimentation showed the off-the-shelf neural network based FLAIR sentiment evaluation module to be most effective.

FLAIR was developed by researchers of Zalando (Akbik et al., 2019). It is especially able to deal with previously unencountered words as are common in jargon or slang. This method is available as a python module[10] and easily applied to the tweet text. It requires no text pre-processing and is capable of handling emojis and urls in the text. Running the algorithm over each tweet results in a sentiment coefficient per tweet. It can have value 1 for very positive or value -1 for very negative or any value inbetween.

For the analysis datasets spaced at 512 and 86 400 seconds are needed. The sentiment coefficients for all tweets occurring within an interval have to be combined into a single value per interval. This can be achieved in multiple ways. The first approach is to simply take the equally weighted mean of the sentiments. The second approach weights the sentiment

[9]https://twitter.com/en/tos
[10]https://github.com/flairNLP/flair

coefficients by the number of followers of the poster and the third approach weights the mean by the number of friends of the poster. This results in three different time series.

The same process is repeated for the 1000 selected influencer users, however, due to their lower number their frequency of tweets is too low for a datapoint every 512 seconds, so the influencer mean sentiment is only available in the 86 400 dataset.

The raw data of tweets contains user identifiable data, however, for the purposes of this analysis this data is not used, according to the Data Minimisation principle. Only the various weighted mean sentiment values are used in the following analysis. By analysing the group of users as a whole, no personally identifiable conclusions can be made about any individual in the dataset.

Final processing

The data was obtained for all variables and combined into one data frame for 512 data and one for 86 400. In the next step each column in the dataframes was min-max normalised, such that the largest value in the column would equal to 1, and the smallest value equal to 0. The time range was limited to start and end points such that data for all variables were available in that time range. The time range is further described in Section 3.5.

As a last step outliers were removed from the data. For each sample per variable the z-score was determined. This indicates the number of standard deviations that this sample is away from the mean of the variable. Any sample with a standard deviation larger than 3 was removed and replaced by a linear interpolation between its predecessor and successor values, to maintain continuity of the data.

3.5 Dataset description

Following the steps described in the Methodology Section 3 data was collected and combined in a dataframe. The general format is similar to a spreadsheet, with each row containing the data corresponding to a particular point in time and each column containing the data of a single factor. Two dataframes were created. The first is used for intra-day analysis at data spacing of 512 seconds and the second at data spacing of 86 400 seconds is used for analysing inter-day effects.

The 512 data covers the time from Sunday 31st January, 2021 23:57:52 until Friday 2nd July, 2021 13:00:48, which corresponds to a total of 4 months, 29 days, 13 hours, 2 minutes and 56 seconds. The dataframe has 25574 rows and 1065 columns. An example of some data, the last 7 datapoints, from this dataset are shown in Table 12 on p. 32.

Table 12: Example 512 data

	BTC_EUR	ETH_EUR
2021-07-02 11:44:00+00:00	27857.747413	1721.538805
2021-07-02 11:52:32+00:00	27997.053802	1728.492447
2021-07-02 12:01:04+00:00	27920.675394	1722.926749
2021-07-02 12:09:36+00:00	27978.709049	1726.715299
2021-07-02 12:18:08+00:00	27920.544072	1726.029414
2021-07-02 12:26:40+00:00	27882.367126	1723.138894
2021-07-02 12:35:12+00:00	27884.021475	1723.128673

The 86 400 data covers the time from Monday 1st February, 2021 until Saturday 3rd July, 2021, which corresponds to a total of 4 months and 30 days. The dataframe has 153 rows

Table 13: Example 86 400 data

	BTC_EUR	ETH_EUR
2021-06-26 00:00:00+00:00	27798.319314	1564.363114
2021-06-27 00:00:00+00:00	26296.452631	1489.488892
2021-06-28 00:00:00+00:00	27774.741150	1570.615974
2021-06-29 00:00:00+00:00	28935.586609	1725.960808
2021-06-30 00:00:00+00:00	29970.263157	1829.683578
2021-07-01 00:00:00+00:00	29273.064514	1821.506457
2021-07-02 00:00:00+00:00	28366.735589	1807.918481

and 1356 columns. An example of some data, the last 7 datapoints, from this dataset are shown in Table 13 on p. 33.

The shown data is only an example for the general structure of the data of two selected colums for the market prices of Bitcoin and Ethereum. Displaying the whole dataframe unfortunately is not possible in printed form. The dataset has been made available for public download through 4TU.ResearchData at the following URL: `http://dx.doi.org/10.4121/14904813`

The data aggregation process took over 2.5 months in total. One of the biggest challenges was ensuring temporal consistency between different variables. Especially for the intra-day data it was critical to ensure the variables were all sampled at the exact same time. Just because variables might be obtained at a spacing of 512 between datapoints does not mean they may be compared. Particular care was taken to ensure the correct alignment of all data series with each other. Having a variable shifted in phase compared to others would reduce the quality of the results as the shifted variable would effectively contain information “from the future”. As part of the time conversion also time-zones had to be properly accounted for. The time-range under analysis contained a switch from daylight savings time to summer time for most of the data sources. Again special care was taken to ensure no shifting of any variable data occurred.

The intra-day datapoints were defined such that a datapoint was placed whenever the seconds of the unix-timestamp since its epoch of Thursday 1st January, 1970 were evenly divisible by 512 or in other words when the remainder modulo of 512 was equal to 0. The inter-day datapoints were sampled at exactly midnight for each day.

All the combining of data was done using custom Python scripts leveraging the Python module Pandas. Spreadsheets are commonly used in academia and industry, however, due to the large number of variables, markets and samples this was not a feasible option. Pandas allows to do spreadsheet-like operations within Python, which was crucial for the operations required as part of the analysis. The intra-day dataset is stored in csv format and has a filesize of 523.8 MB. The inter-day dataset is stored in the same format and has a filesize of 3.5 MB

4 Results

In Section 3.2 the reason to use a Hierarchical Bayesian linear regression was described. In this section it is applied to the dataset of Section 3.5. Each β coefficient in Eq. (3.1) determines the magnitude and sign of the contribution of a specific variable to the price estimate. Variables that are observed to have a coefficient distribution with a larger magnitude relative to the other variables can be seen as having a larger influence. The significance of a particular variable is implicitly obtained by making use of the Bayesian inference approach. With a conventional ordinary linear regression one would obtain a single value coefficient per β, however, the Bayesian inference results in a posterior probability distribution per variable. The magnitude, variance and significance can be compared between variables.

Variables with a high degree of variance and low magnitude are less significant and could be independent. Conversely a variable with low variance and high magnitude is likely significant and the price could be dependent. For a completely linearly dependent variable a scatter plot is expected to show a line. For a completely in-dependent variable a scatter plot is expected to show a uniformly random distribution over the entire plot. Appendix B.2 contains scatter plots for the variables of the ETH_USD market illustrating most variables are neither clearly dependent nor in-dependent. This re-affirms the choice of an hierarchical analysis method, however, it is difficult to draw any conclusions from scatter plots alone. So we continue with the main analysis method here.

As part of the results the significance of variables will be discussed. In frequentist methodologies the p-value is commonly used to represent the probability of the null-hypothesis. In this Bayesian analysis the p-value is not defined (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019). An alternative metric for significance is needed. Makowski et al. (2019) propose to use the p-direction, the probability of direction. It is defined as the ratio of samples with the same sign as the median value. If it is high (close to 1) this signifies a high certainty of the null-hypothesis to be false. A low value (close to 0) signifies a high un-certainty of the null-hypothesis, but does not necessarily poof the null hypothesis to be false, contrary to the p-value. A second useful metric identified by Makowski et al. (2019) is the Region of Practical Equivalence (ROPE). Kruschke (2014) suggest to only consider coefficients with values outside of the -0.1 to 0.1 range as significant when using linear regression models (like the one used in this analysis). The ROPE is defined as the ratio of samples within the -0.1 to 0.1 range to the total number of samples. A small ROPE value indicates significance, however, a large one does not necessarily guarantee insignificance. It could be the case that a coefficient has a small influence but with high certainty. This would manifest in a large p-direction and a large ROPE value. No single metric captures all properties of significant coefficients, so for each variable the selected metrics have to be considered in their context.

The main output from the Bayesian model is a sequence of traces. Each trace contains a set of sampled values for all variables and parameters in the linear regression model, see Section 3 for more detail on this process. The last 1000 traces are selected as the result of the analysis. The selected traces should all have been sampled after the model has converged to a steady solution. This is checked by creating auto-correlation plots of the traces. When the auto-correlation of the traces is high, this indicates the solution has not converged yet. Appendix B.1 shows the auto-correlation plots. The solution appears to have a low auto-correlation which is sufficient to assume convergence and we can continue the analysis.

In Section 4.3 research question 2 “What is the magnitude of effects of the variables influencing cryptocurrency price volatility?” is answered by reviewing the individual coefficient distributions. Research question 3 "Which variables have the biggest influence on price increase?" and research question 4 "What theories can explain the observed effect of the variables?" will be addressed in Section 4.4 by comparing the coefficients effects to each other and deriving theories to explain the observed effects.

4.1 Model Accuracy

Before we investigate the effect of variables we should get a better understanding of the power of the obtained model. Since the model uses multiple exogenous variables we would like to penalise the R^2 when too many variables are included in the model, which might lead to over-fitting. The adjusted R^2 values are reduced in magnitude compared to the standard[11] R^2 values and are considered to be a more realistic representation of the models performance (Miles, 2005). In Eq. (4.1) k is the number of exogenous variables and n is the number of samples. For intra-day $k = 14$ and $n = 25574$. For inter-day $k = 25$ and $n = 153$. Table 14 shows the obtained adjusted R^2 and the standard R^2 values for each market. Generally the accuracy appears to be higher for the inter-day data compared to the intra-day data. The mean adjusted R^2 value for intra-day data is 0.69 and the mean adjusted R^2 value for inter-day data is 0.90. The model shows an ability to estimate the current price fairly accurately. This is considered sufficiently accurate to allow for analysis of the effects of each variable in the model and to achieve the research objective of gaining a deeper understanding of the dynamics of cryptocurrency markets.

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \qquad \text{(Miles, 2005)} \tag{4.1}$$

Table 14: R^2 values of the bayesian hierarchical model estimate

	R^2 512 data		R^2 86400 data	
	standard	adjusted	standard	adjusted
ADA_EUR	0.687787	0.687616	0.826547	0.792403
ADA_USD	0.702227	0.702064	0.866141	0.839791
BTC_DAI	0.839789	0.839701	0.937655	0.925382
BTC_EUR	0.860061	0.859984	0.950259	0.940467
BTC_USD	0.848056	0.847973	0.952661	0.943342
BTC_USDC	0.839457	0.839369	0.942918	0.931681
BTC_USDT	0.853741	0.853661	0.945905	0.935256
ETH_EUR	0.611348	0.611135	0.918509	0.902467
ETH_USD	0.618681	0.618472	0.913469	0.896435
ETH_USDC	0.605739	0.605523	0.910797	0.893238
ETH_USDT	0.614628	0.614416	0.916854	0.900487
LTC_EUR	0.636351	0.636151	0.943396	0.932253
LTC_USD	0.629898	0.629695	0.951092	0.941465
LTC_USDT	0.632229	0.632028	0.951288	0.941699
XLM_USD	0.592198	0.591975	0.830205	0.796780
XRP_EUR	0.620392	0.620185	0.916288	0.899810
XRP_USD	0.623134	0.622927	0.909318	0.891467
XRP_USDT	0.625338	0.625133	0.907757	0.889599
mu	0.691170	0.691000	0.916170	0.899668
sigma	0.100709	0.100764	0.037884	0.045342

[11] un-adjusted

Fig. 7 and Fig. 8 on page 36 are shown here as examples of the model fit. In each plot the mean estimation is shown by the orange line. The light orange area around it visualises the confidence interval containing one standard deviation (68 %) of values.

In Fig. 7 the intra-day price estimate is shown. The high degree of variance between datapoints is noteworthy. This might indicate the variables used in the model are noisy. The general trend of the estimate appears to track the market price (blue line) indicating that the model can be used to estimate the price. Fig. 8, in contrast, shows less variance in its estimate for inter-day price, as can be seen by the light orange one standard deviation confidence interval being smaller. The orange estimate closely tracks the blue price curve.

The inter-day model makes use of additional variables that were not available at the required resolution in time for the intra-day data. The unavailable variables were influencer sentiment, community size and Github statistics. This difference in variables could explain the larger variance in Fig. 7. Future research should therefore expand the number of variables accounted for in the intra-day dataset by improving the real-time data collection. In general the result shows that the price of a cryptocurrency market can be estimated using data from external variables. The adjusted mean (mu) R^2 values of Table 14 show that a larger portion of inter-day changes can be explained with the analysed variables than for intra-day. The standard deviation (sigma) between markets is larger for intra-day than for inter-day. The Bitcoin markets appear to reach the highest R^2 values for both intra- and inter-day data. Each market has its own set of coefficient distributions and hyper-prior distributions. In the following sections the individual effect of the variables will be discussed. First the hyper-priors are described in Section 4.2, then the forest plots for each variable are shown and analysed in Section 4.3.

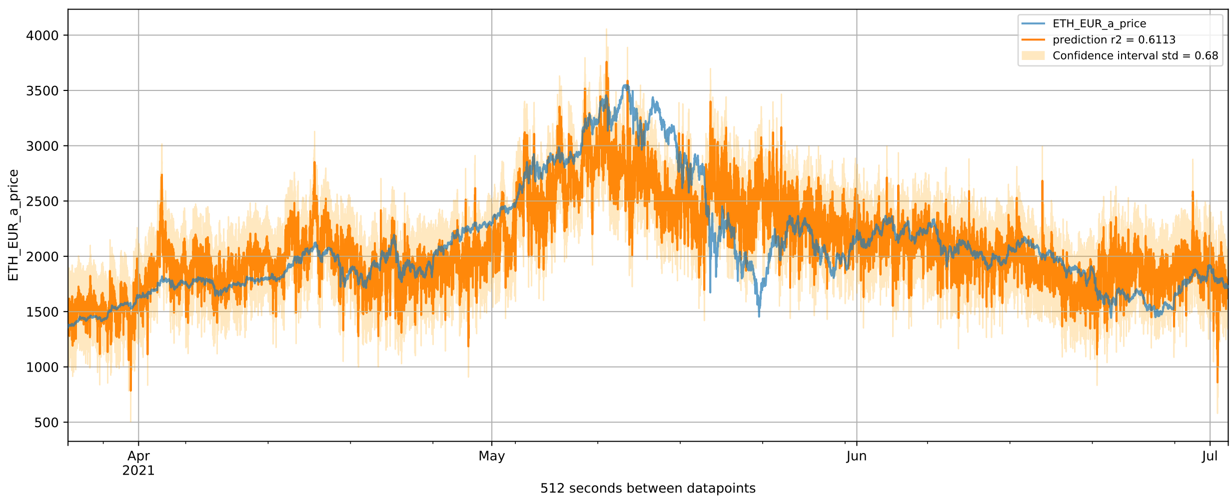

Figure 7: Time series ETH_EUR, 512 data

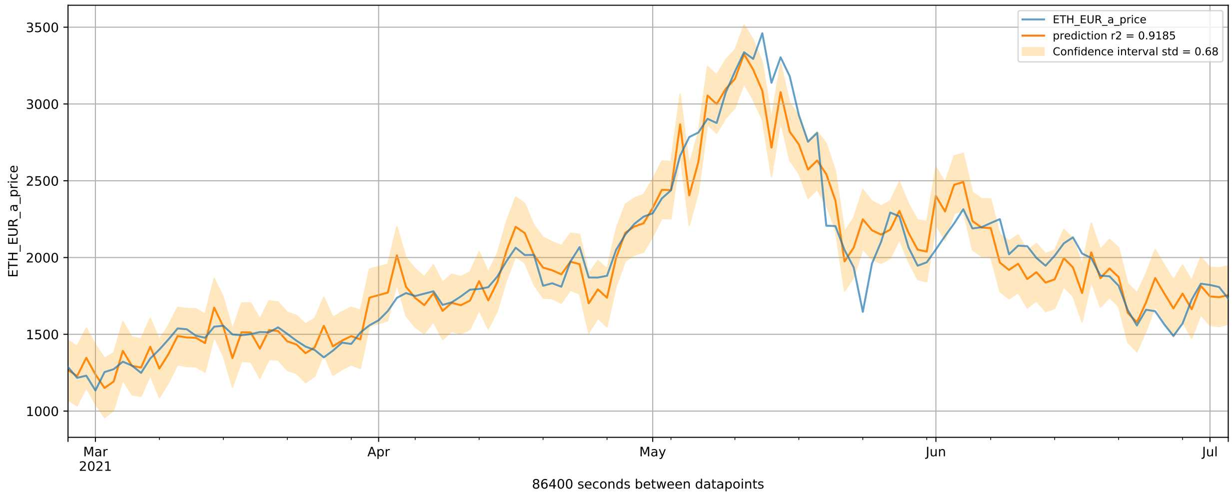

Figure 8: Time series ETH_EUR, 86400 data

4.2 Hyper-Prior Distributions

Next we plot the hyper-priors. Fig. 9 on p. 39 shows the posterior distribution corresponding to the 512 data. Fig. 10 on p. 40 shows the posterior distribution for the 86 400 data. Table 15 contains the significance statistics for the intra-day model. The theory behind p-direction and ROPE are explained in mode detail in the introduction to the results Section 4. Some of the variables have a p-direction of 1. This is the result of a numeric inference using a finite number of samples. It is possible for all samples taken to be entirely positive or negative, even when the true underlying normal curve distribution would in theory extend to both positive and negative infinity. The most significant variable hyper-prior means are found to be oil close, gold volume, Twitter sentiment polarity, Google trends, sp500 close, general internet activity levels, price volatility and the Twitter sentiment weighted by number of friends. Potentially insignificant variables are the Twitter sentiment weighted by number of followers, gold close price, sp500 volume, oil volume and volume log return volatility. The control variable of market volume had the least significance of all.

Table 15: Cluster A hyper-prior variable distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
intercept mu	0.621	0.106	0.170	1.000	0.000
mu oil close	-0.563	0.102	0.181	1.000	0.000
mu gold volume	0.396	0.033	0.083	1.000	0.000
mu polarity flair list	-0.385	0.033	0.087	1.000	0.000
mu quote googletrends	0.366	0.112	0.306	0.997	0.012
mu sp500 close	0.334	0.068	0.203	1.000	0.001
mu google googletrends	-0.247	0.060	0.245	0.999	0.005
mu a price log return volatility	-0.199	0.058	0.293	0.998	0.042
mu polarity flair user friends count list	0.159	0.025	0.159	1.000	0.010
error	0.127	0.001	0.005	1.000	0.000
mu polarity flair user followers count list	0.041	0.009	0.216	1.000	1.000
mu gold close	-0.054	0.074	1.360	0.557	0.726
mu sp500 volume	-0.029	0.006	0.214	1.000	1.000
mu oil volume	0.008	0.012	1.555	0.521	1.000
mu a volume log return volatility	0.005	0.023	4.473	0.193	1.000
mu a volume	0.001	0.018	13.280	0.059	1.000

For some of the variables there might still be variance between the different markets. For example gold close might still be important for some of the markets, as it is showing a high variance. A high variance may indicate the variable is highly dependent on the specific market and not a global effect. This can also visually be confirmed by reviewing the gold close distribution in Fig. 9. The magnitudes on a per market bases will be discussed later in Section 4.3.

In Table 16 we find the variables with low significance for inter-day to be volume log return volatility, the number of forks and stars and the general Twitter sentiment polarity. The number of forks and stars have a large variance, indicating these variables to be highly market specific. The most significant variables appear to be number of Twitter followers, total issues, the influencer sentiment polarity and the number of Reddit subscribers. The Twitter followers and Reddit subscriber variables have a large variance, indicating these variables to have varying effects across markets. This is confirmed by visually inspecting the distributions in Fig. 10.

Table 16: Cluster B hyper-prior variable distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
mu quote twitter followers	-1.633	0.505	0.309	0.998	0.001
mu quote total issues	1.062	0.309	0.291	0.996	0.002
mu influencer polarity flair list	-0.690	0.120	0.175	1.000	0.000
mu quote reddit subscribers	0.717	0.619	0.864	0.758	0.071
mu influencer polarity flair user followers cou...	0.410	0.083	0.203	1.000	0.000
mu influencer polarity flair user friends count...	0.351	0.073	0.208	1.000	0.000
mu sp500 close	0.274	0.064	0.234	1.000	0.007
mu quote open issues	-0.254	0.055	0.216	0.999	0.012
mu a price log return volatility	-0.175	0.040	0.231	1.000	0.028
mu sp500 volume	0.165	0.016	0.100	1.000	0.000
mu quote pull requests merged	0.269	0.353	1.309	0.562	0.174
mu gold volume	-0.149	0.013	0.089	1.000	0.001
mu gold close	0.150	0.058	0.389	0.986	0.185
intercept mu	0.162	0.099	0.614	0.898	0.252
mu quote googletrends	0.115	0.074	0.641	0.879	0.409
mu google googletrends	-0.099	0.025	0.248	1.000	0.509
mu oil volume	-0.087	0.017	0.191	1.000	0.788
mu polarity flair user followers count list	0.082	0.024	0.295	0.999	0.782
mu polarity flair user friends count list	-0.085	0.043	0.503	0.950	0.636
error	0.078	0.001	0.016	1.000	1.000
mu quote closed issues	-0.146	0.275	1.886	0.450	0.226
mu oil close	0.052	0.027	0.513	0.949	0.963
mu quote stars	-0.126	0.284	2.256	0.346	0.254
mu a volume	0.045	0.042	0.928	0.724	0.909
mu polarity flair list	-0.035	0.037	1.055	0.664	0.958
mu a volume log return volatility	0.019	0.036	1.860	0.427	0.986
mu quote forks	-0.038	0.219	5.776	0.185	0.354

Fig. 9 and Fig. 10 allow for an initial high level understanding of the variable effects, but much of the detailed nuance the hierarchical Bayesian linear regression analysis method can offer is lost when only considering the hyper-priors. This detail can be found in the forest plots for each variable. The forest plots will be discussed in detail in Section 4.3.

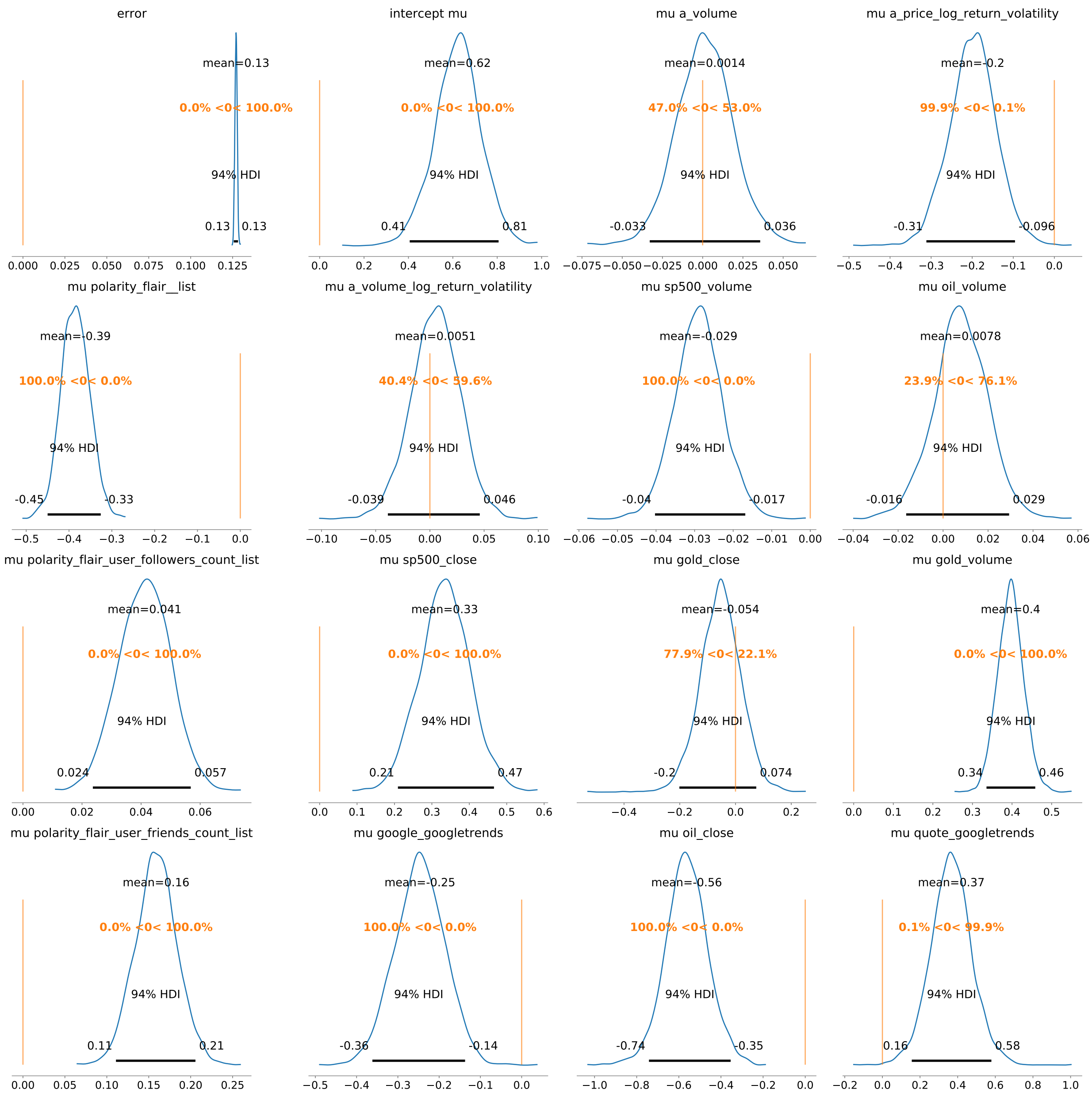

Figure 9: Hyper-priors Posterior plot 512 data

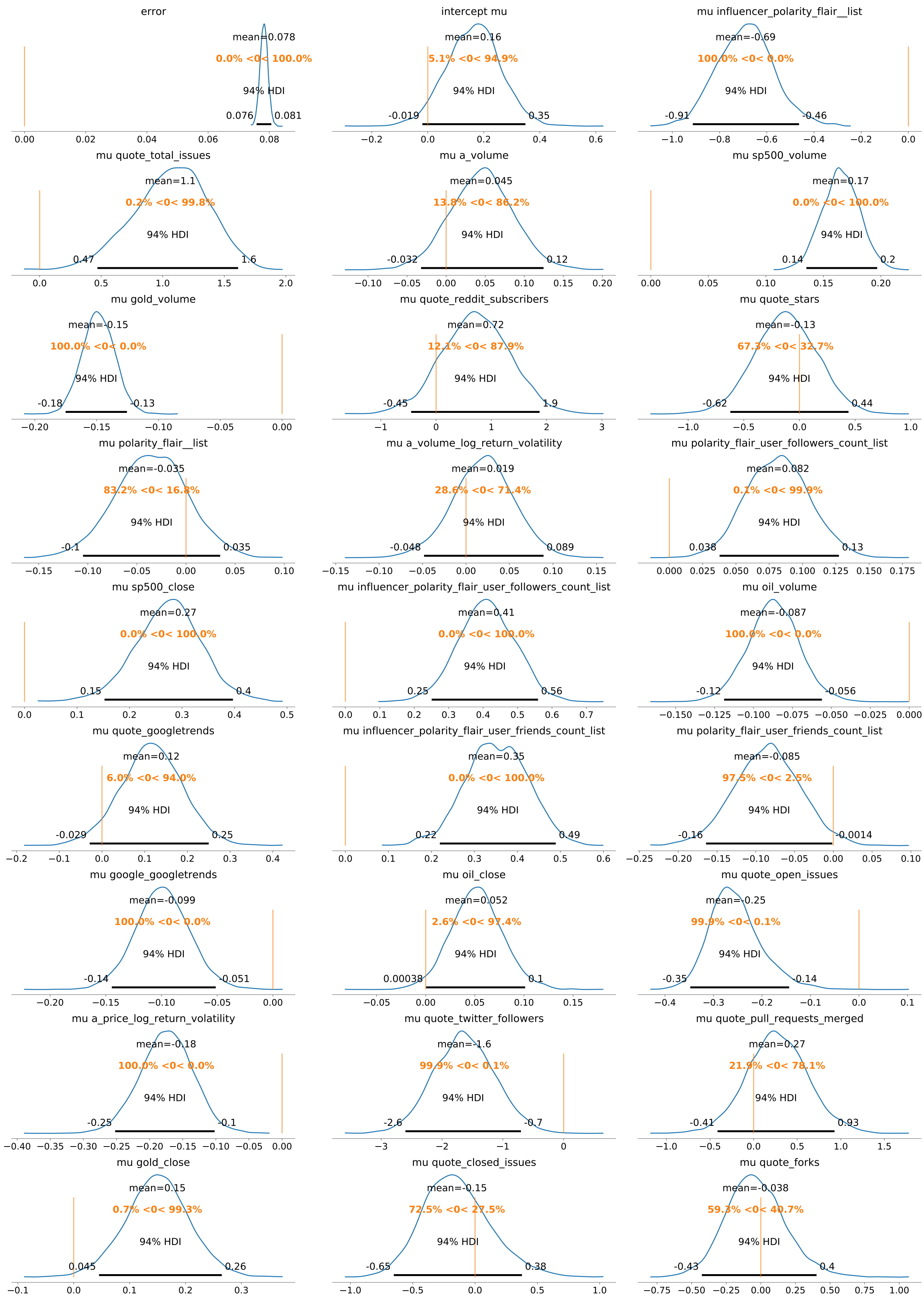

Figure 10: Hyper-priors Posterior plot 86 400 data

4.3 Variable Analysis Results

In the following the results obtained for each variable are shown and used to argue about the established hypothesis. Later in Section 4.4 potential reasons for the differences and unexpected observations will be discussed. These observations answer **research question 2** by visualising the direction and magnitude of effects of the variables on the cryptocurrency prices. To keep the text shown here concise, only the most important beta-coefficient means and variance are mentioned in the text. The full tables for each variable per market can be found in Appendix C.

Control Variable Volume

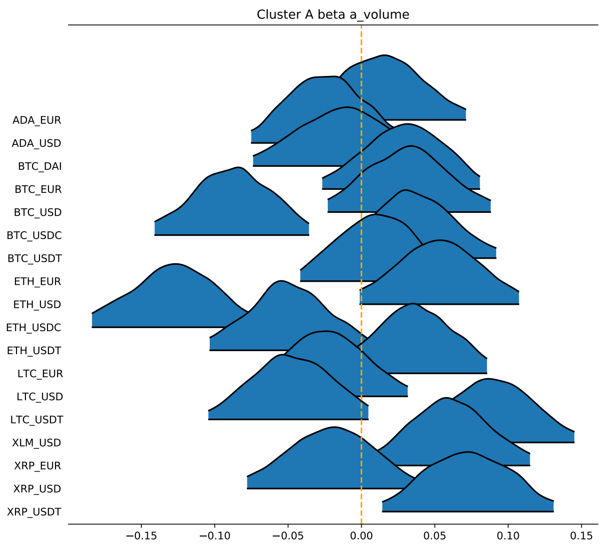

Figure 11: Forest plot control variable volume 512 data

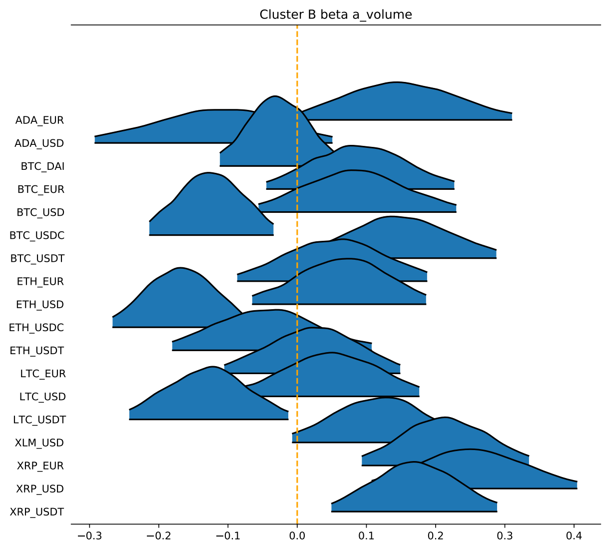

Figure 12: Forest plot control variable volume 86400 data

The hyper-prior for the volume beta-coefficient of the intra-day model had a mean of 0.001 and a variance of 0.018 with p-direction of 0.059. Together these values indicate a high degree of uncertainty and variance of the influence of volume in the model. This is confirmed by the per-market distributions shown in Fig. 11. There is no homogeneous effect observable and the magnitude of the effect is mostly within the ROPE range. The coefficients in Table 19 on p. 75 in Appendix C show ETH_USDC to be the market most significantly influenced with a p-direction of 1 and ROPE coefficient of 0.172, however, the market to be least significantly affected is ETH_EUR with a p-direction of 0.29 and ROPE of 0.999. We would expect the same cryptocurrency, in this case Ethereum, to be effected similarly strongly regardless of which fiat currency it is being traded against. In the following, for example in the community size variable, such per-currency clustering can be seen.

Similarly reviewing the distributions in Fig. 12 and statistics in Table 34 on p. 83 in Appendix C for the inter-day data leads to the conclusion that the effect of the control variable volume is considered insignificant and the model appears to hold independently of the volume of the markets analysed. An exception might be Ripple, which is the cryptocurrency most significantly affected by the volume control variable on inter-day data. Ripples inter-day mean coefficient value range from 0.17 for XRP_USDT to 0.25 XRP_USD.

The analysis was repeated without the inclusion of the control variable. This resulted in the R^2 values shown in Table 17 on p. 74 in Appendix B.3. Compared to the R^2 values in Table 14 in Section 4.1, we find the inclusion of the volume control variable results in a percentage mean increase of R^2 value of 0.33 %. This indicates the volume has negligibly little impact on the model performance.

Volatility

The Figures below show the posterior distribution for possible values of the volatility variable β-coefficients for each market.

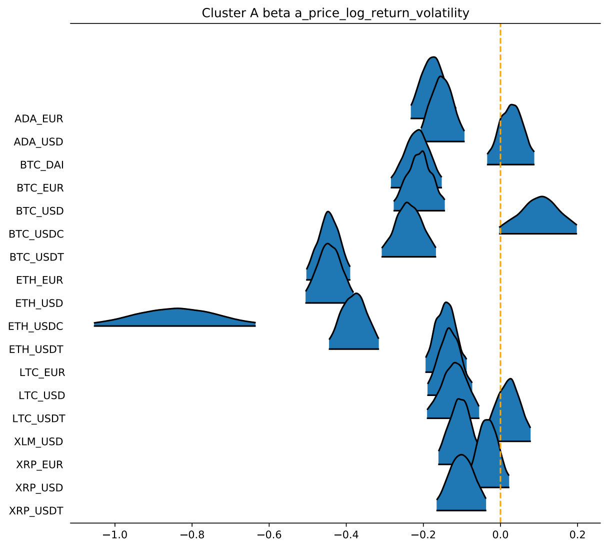

Figure 13: Forest plot Price log return Volatility 512 data

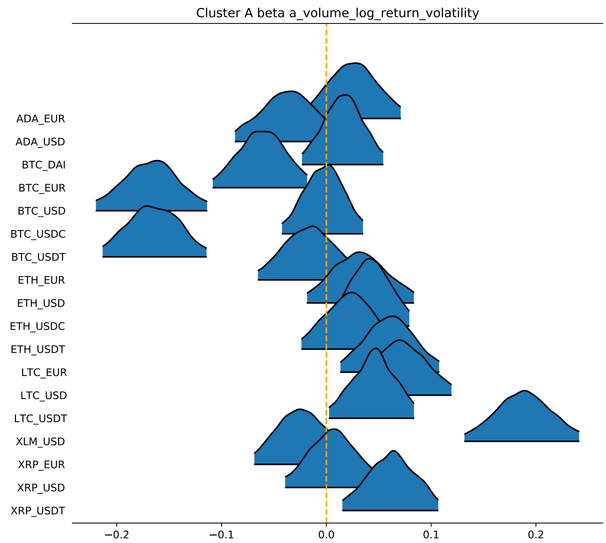

Figure 14: Forest plot Volume log return Volatility 512 data

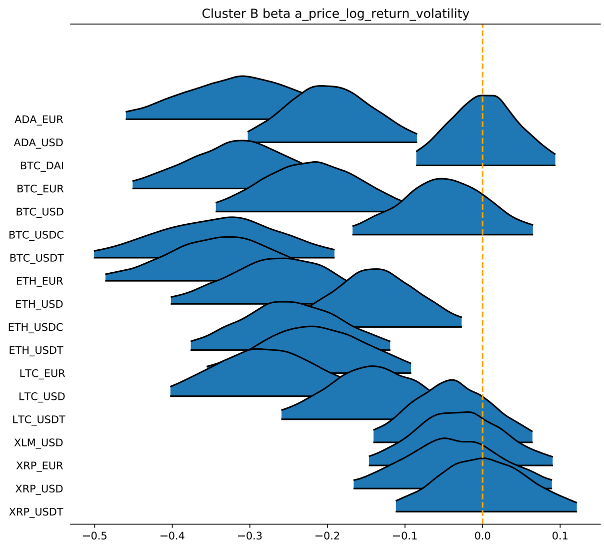

Figure 15: Forest plot Price log return Volatility 86400 data

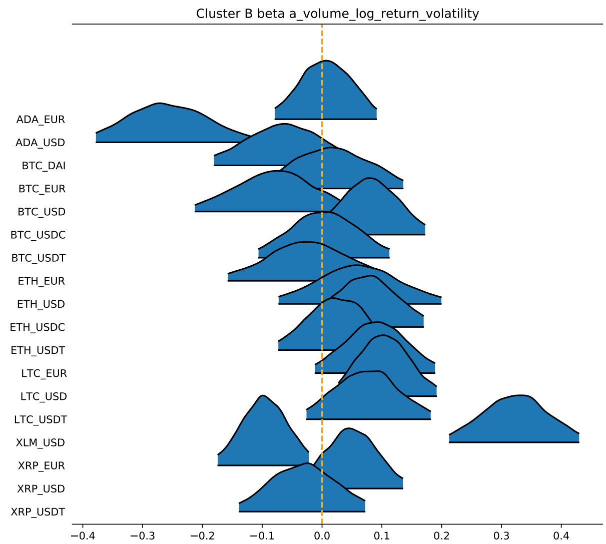

Figure 16: Forest plot Volume log return Volatility 86400 data

The following observations are made for the coefficients on the intra-day data based on the forest plots in Figures 13 and 14:

For most currencies price volatility appears to have a generally small effect. Bitcoin Cardano and Litecoin have mean coefficients from -0.12 to -0.24. Ripple, Lumen and the BTC_DAI outlier market have coefficients to small to assert significance. This is supported by ROPE coefficients ranging from 0.43 to 0.99 Ethereum is the most influenced of the analysed cryptocurrencies. The Ethereum to EUR, USD and USDT markets have a mean coefficient value of -0.44 to -0.38. A potential outlier might be the ETH_USDC market, which shows the largest coefficient magnitude of all markets at -0.84. Due to the variance between markets there does not appear to be a homogeneous relationship. The full set of statistics for price volatility on intra-day can be found in Table 18 on p. 75 in Appendix C.

Volume volatility has a small coefficient value as well. All markets, with the exception of XLM_USD, BTC_USD and BTC_USDT, have a high ROPE coefficient values ranging from 0.87 for LTC_USD to 1 for BTC_USDC. This indicates a high degree of uncertainty with relatively low levels of influence. The full set of statistics for volume volatility on intra-day can be found in Table 20 on p. 76 in Appendix C.

The following observations are made for the coefficients on the inter-day data based on the forest plots in Figures 15 and 16:

Price volatility coefficients have a similar effect on inter-day data as on intra-day. The markets BTC_USD, XLM_USD, XRP_USD, XRP_EUR had a p-direction value between 0.58 to 0.39. The p-direction values for BTC_DAI and XRP_USDT are 0.03 and 0.001 respectively. The remaining markets showed a p-direction close to 1 with a corresponding ROPE close to 0. This indicates the price volatility has a heterogeneous effect on the markets. Table 33 on p. 82 in Appendix C shows the coefficients for all markets.

Volume volatility appears to be largely centred around the 0 coefficient and no clear relation is visible. This is supported by the mu volume volatility statistics of $\mu = 0.019$ and $\sigma = 0.036$ Table 16. The variance is considerably large compared to the mean value. While an effect might be present for some markets it must be concluded to be too uncertain in this analysis. With the exception of XLM_USD and ADA_USD, all markets have a ROPE coefficient from 0.47 to 0.97 indicating a high degree of uncertainty as well. The full set of statistics for volume volatility on inter-day can be found in Table 33 on p. 82 in Appendix C.

The volatility hypothesis was defined as: ***H1: When Volatility increases, the price increases***. This hypothesis is not true according to the observed coefficients. There is no uniform relationship, since some currencies seem unaffected while others exhibit a behaviour inverse to the hypothesis. The magnitude of the effect from price volatility appears larger than from volume volatiltiy.

Macro Economic variables

Similar to before, we select the distributions of the variables of interest. The close price and volume coefficients for Oil, Gold and SP500 are shown below.

The following observations are made based on the coefficients of the intra-day data as shown in Figures 17 to 22:

Oil close price shows various coefficients depending on currency. Bitcoin has the largest negative coefficient of -1.09, followed by Lumen (-0.89) and Ripple (-0.80). Ethereum and Litecoin show a lower but still negative coefficient of -0.28 and -0.34 respectively. Finally Cardano has a positive coefficient of 0.11. However, due to Cardano's high ROPE value of 0.29 for ADA_EUR and 0.52 for ADA_USD the significance of oil close price on Cardano is likely low. For the remaining markets a high degree of significance is observed. The full set of statistics be found in Table 24 on p. 78 in Appendix C.

Gold close price shows both negative and positive coefficients. Bitcoin shows the most negative coefficients from -0.29 to -0.42. Ripple, Lumen and Litecoin are still negative but with a much smaller magnitude. Ethereum is showing a positive coefficient and Cardano is also positive with the largest coefficient magnitude of 0.52. The ROPE statistic indicates Gold close might not be significant for Litcoin and Lumen. The full set of statistics be found in Table 21 on p. 76 in Appendix C.

Most currencies are positively correlated with SP500 close price, with the exception of Cardano which is negative. Ripple has the largest magnitude, followed by Lumen, Bitcoin,

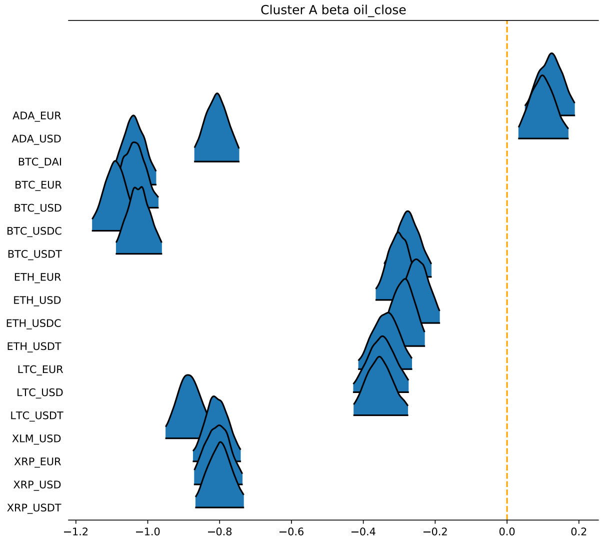

Figure 17: Forest plot Oil Close 512 data

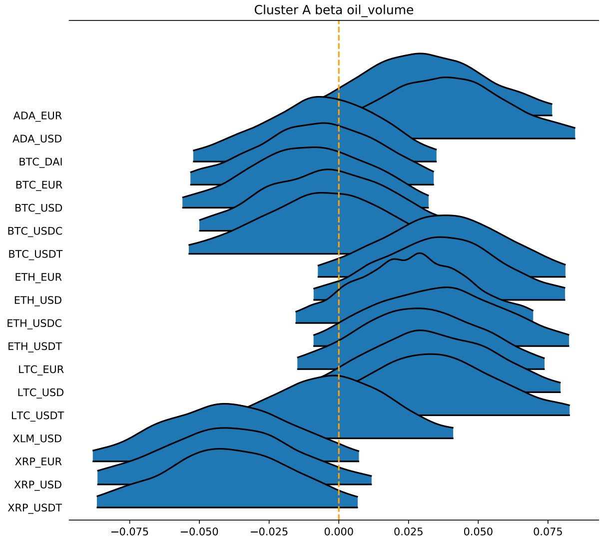

Figure 18: Forest plot Oil Volume 512 data

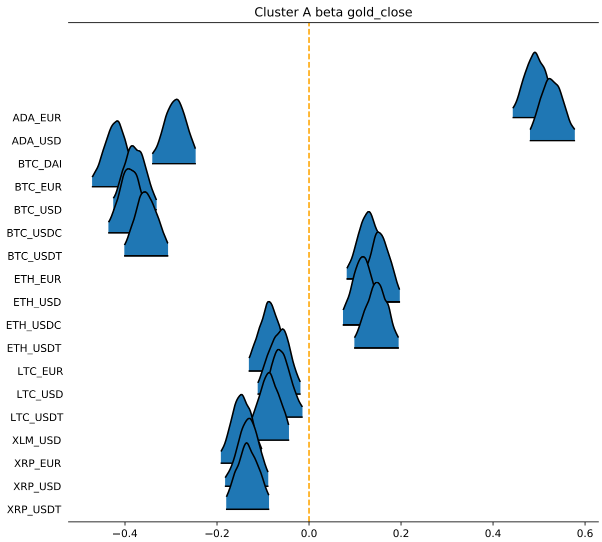

Figure 19: Forest plot Gold Close 512 data

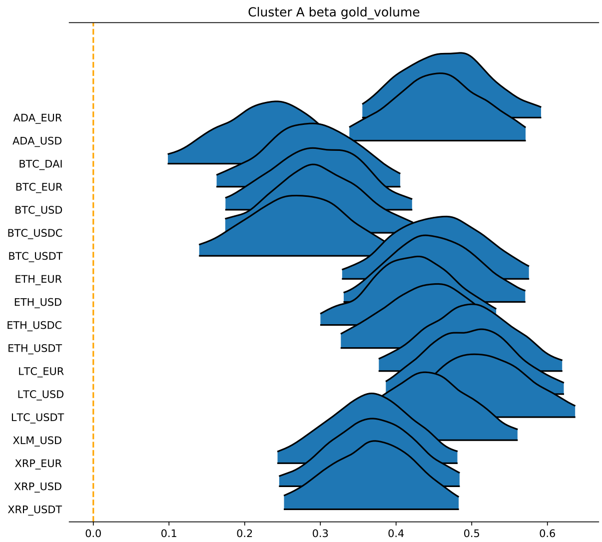

Figure 20: Forest plot Gold Volume 512 data

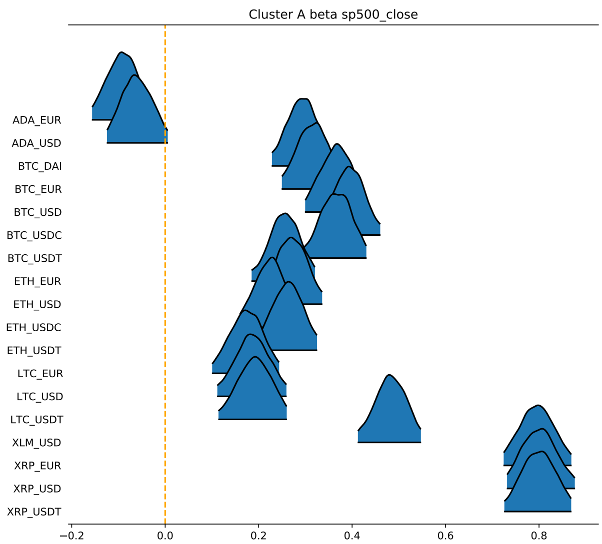

Figure 21: Forest plot SP500 Close 512 data

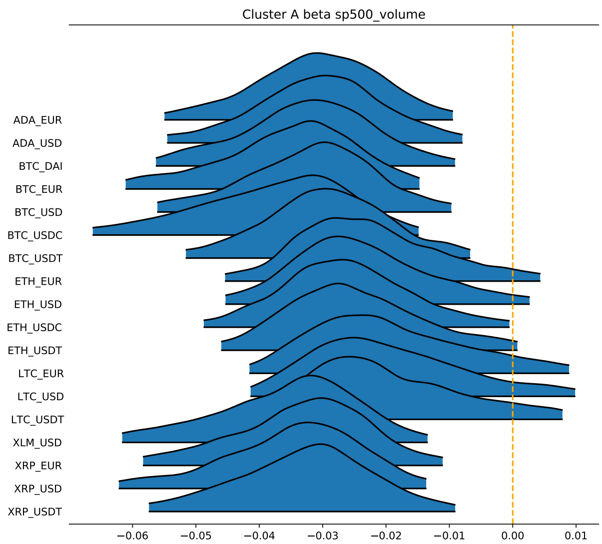

Figure 22: Forest plot SP500 Volume 512 data

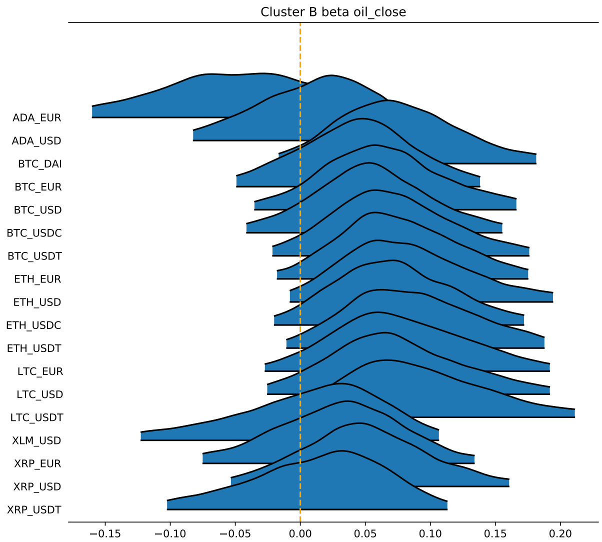

Figure 23: Forest plot Oil Close 86400 data

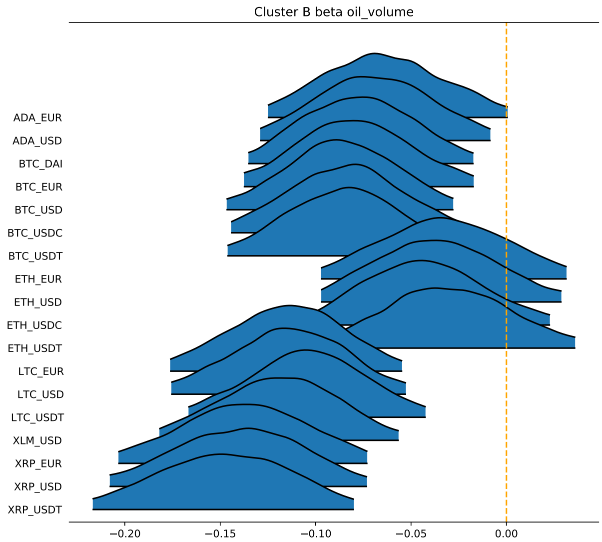

Figure 24: Forest plot Oil Volume 86400 data

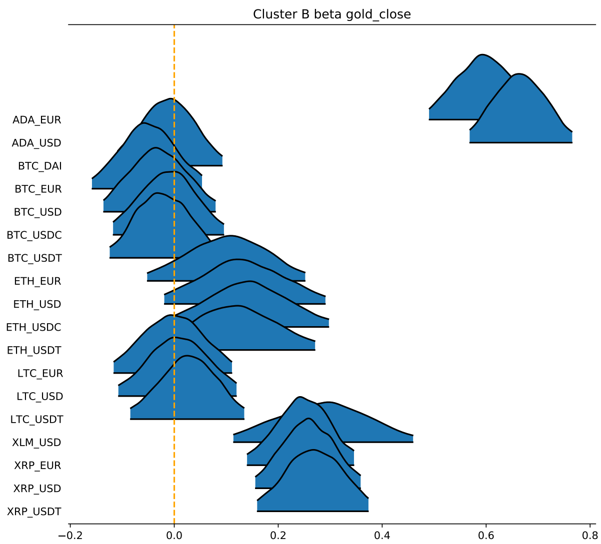

Figure 25: Forest plot Gold Close 86400 data

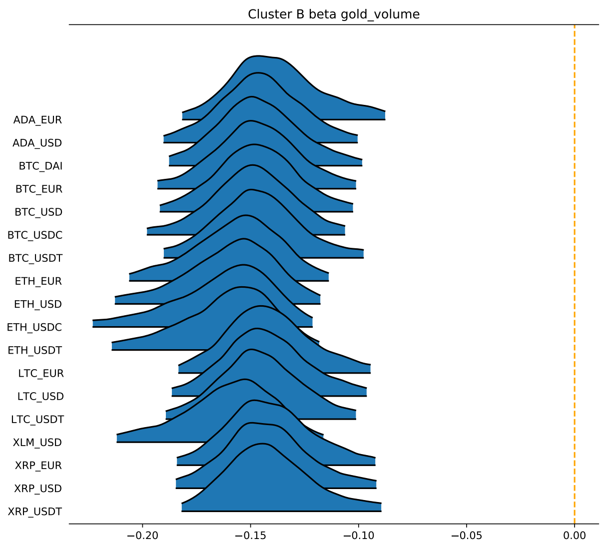

Figure 26: Forest plot Gold Volume 86400 data

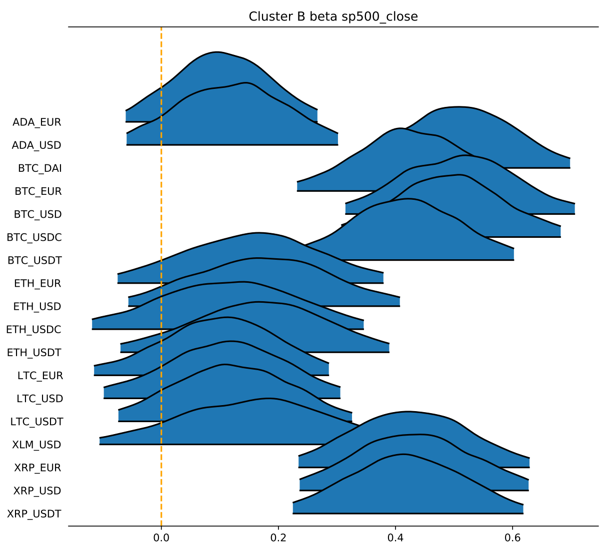

Figure 27: Forest plot SP500 Close 86400 data

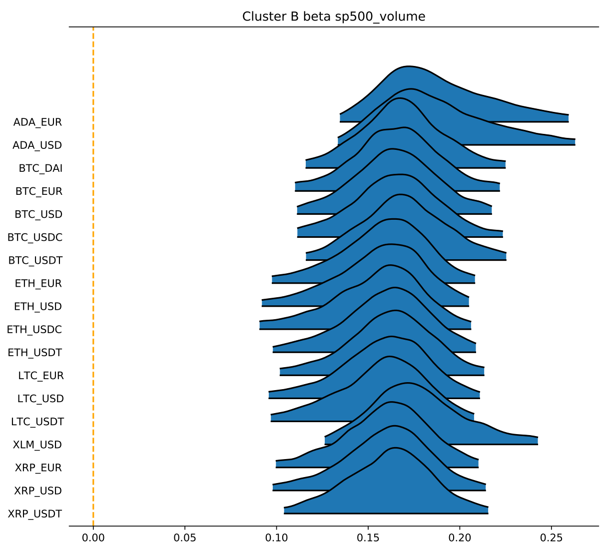

Figure 28: Forest plot SP500 Volume 86400 data

Ethereum and Litecoin in descending order. The full set of statistics be found in Table 30 on p. 81 in Appendix C.

The variables of SP500 volume and Oil volume have coefficients close to 0, indicating no dependence to exist. This is also confirmed by the hyper-prior statistics for those coefficients in Table 15. However, reviewing SP500 volume more closely, it shows a fairly high p-direction statistic for all markets in Table 31 on p. 81 in Appendix C. SP500 volume could be significant with a very low magnitude of effect. The statistics for oil volume can be found in Table 25 on p. 78 in Appendix C.

Gold volume hyper-prior has a positive mean coefficient value of 0.40. All p-direction and ROPE statistic in Table 22 on p. 77 in Appendix C indicate significance for all markets.

The variance between markets appears to be reduced in the inter-day data compared to intra-day. The following observations are made based on the coefficients of the inter-day data based on Figures 23 to 28:

Oil close coefficient values in Table 42 on p. 87 in Appendix C are close to zero for Ripple, Lumen and Cardano, as indicated by the large ROPE. Bitcoin, Ethereum and Litecoin show a slightly larger positive coefficient. It appears that oil close price is fairly insignificant.

The volumes of SP500 and gold have a fairly homogeneous effect amongst markets. Gold volume shows a small negative coefficient for all markets with a mean value of 0.15 and SP500 volume shows a small positive coefficient of 0.17. The statistics for sp500 volume are shown in Table 57 on p. 94 in Appendix C. The statistics for gold volume can be found in Table 37 on p. 84 in Appendix C.

Oil Volume coefficient shows less magnitude than gold and appears to have the largest magnitude for Ripple, Lumen and Litecoin. Cardano and Bitcoin follow with a similar magnitude. Ethereum finally shows a coefficient value close to 0. The statistics can be found in Table 43 on p. 87 in Appendix C

Gold Close shows insignificant coefficients close to zero for Litecoin, and Bitcoin. Ethereum has a small positive coefficient together with Ripple. Cardano shows a large positive coefficient. The statistics for gold close can be found in Table 36 on p. 84 in Appendix C. SP500 Close shows positive coefficients for all cryptocurrencies, but Cardano, Ethereum, Litecoin and Lumen have coefficients close to 0 and might not be considered significant. Ripple and Bitcoin show a fairly large positive coefficient. The statistics for gold close can be found in Table 56 on p. 94 in Appendix C.

Overall we find a range of patterns for coefficients of macro economic variables. The strength of the relation appears stronger but more varied on intra-day data and weaker but more homogeneous on inter-day data. The following hypothesis were established for the macro-economic variables:

H2.a: The cryptocurrency market prices decrease when SP500 increases. The opposite is observed. The coefficients for almost all markets are positive, although with varying magnitudes depending on currency. However, a positive coefficient with SP500 Volume can be found for all currencies.

H2.b: The cryptocurrency market prices decrease when Oil increases. The coefficients found for intra-day are largely negative indicating a potential relation as proposed. However, on the inter-day data the relation reverses and becomes less significant. The hypothesis cannot be accepted as true.

H2.c: The cryptocurrency market prices increase when Gold increases. This appears to be false. While a small relation as stated can be seen for a number of currencies on the intra-day data, on the inter-day data most currencies show a small coefficient only. The observed variance is too large to come to a broad conclusion. A notable exception is Cardano which shows a strongly positive coefficient with Gold price on both intra- and inter-day data.

Github Insights

Github stats are only available for the 86 400 data.

The following observations are made based on the coefficients of the inter-day data based on Figures 29 to 34:

Number of closed issues has a fairly homogeneous effect on all markets, however, a large part of the markets have a low probability of direction. Only Lumen, Ripple and Litecoin have a p-direction above 0.72. This indicates a high degree of uncertainty associated with this variable. Using more data in future research could potentially increase the certainty to allow a significant relation to be found. The statistics for closed issues can be found in Table 47 on p. 89 in Appendix C.

The number of pull requests has a wide range of effects and appears dependent on the specific market. The variance between coefficient values is large. Cardano shows the largest variance and has a p-direction value of only 0.15. Bitcoin has the largest coefficient value with a mean of 2.25 with a p-direction of 1 and a ROPE of 0. The variable pull requests is likely significant for Bitcoin, Litecoin and Ripple, but insignificant or uncertain for the others. The full set of statistics for pull requests can be found in Table 51 on p. 91 in Appendix C.

An increase of total issues appears to affect all markets homogeneously and with a mean coefficient value of 1.06. In Table 16 it was shown that this variable is the second most significant for the inter-day model. All markets have a high p-direction (>0.95) and low ROPE coefficient (<0.018), confirming a high likelihood of significance. The full set of statistics for total issues can be found in Table 54 on p. 93 in Appendix C.

The number of open issues is also fairly homogeneously correlated with a mean value of -0.25. The statistics for open issues can be found in Table 50 on p. 91 in Appendix C.

The number of stars and forks have varying effects depending on market and generally shows a high degree of uncertainty. This repeats the same observation as seen in their hyper-prior distributions of Table 16. The statistics for number of stars can be found in Table 53 on p. 92 in Appendix C. The statistics for number of forks can be found in Table 48 on p. 90 in Appendix C.

H3.a: When the number of closed issues increases, the price increases as well. This hypothesis cannot be considered true. The probability distribution of coefficients covers a fairly large range of values. However, the number of total issues has a fairly homogeneous positive coefficient with large magnitude.

H3.b: When the number of open issues decreases, the price increases. This hypothesis appears to be true. The analysed markets show a homogeneous distribution around negative coefficient values for all markets. The coefficient magnitude is not particularly large. The standard deviations of the distributions is fairly low compared to its magnitude. This variable can be considered significant.

H3.c: When the number of stars increases, the price increases. There is no clear relationship either way. It appears very currency specific and has a high degree of uncertainty. The magnitudes of coefficients is surprisingly large, however, given the level of spread. The hypothesis appears false.

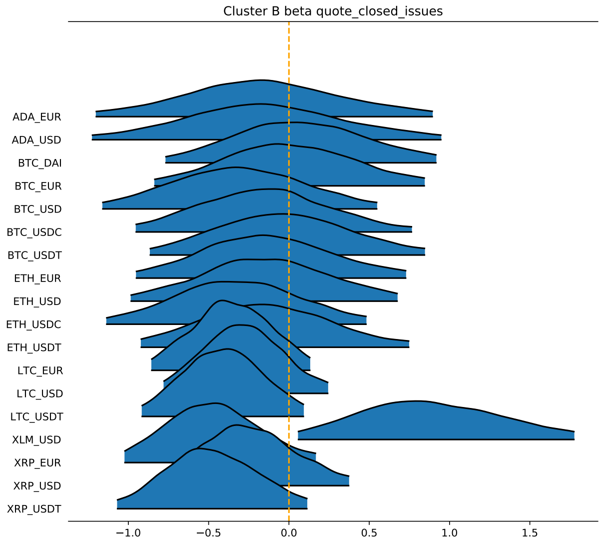

Figure 29: Forest plot currency project number of closed issues 86400 data

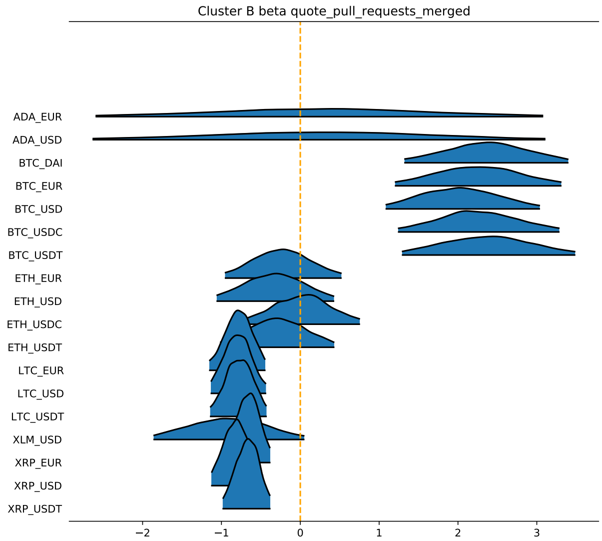

Figure 30: Forest plot currency project number of pull requests merged 86400 data

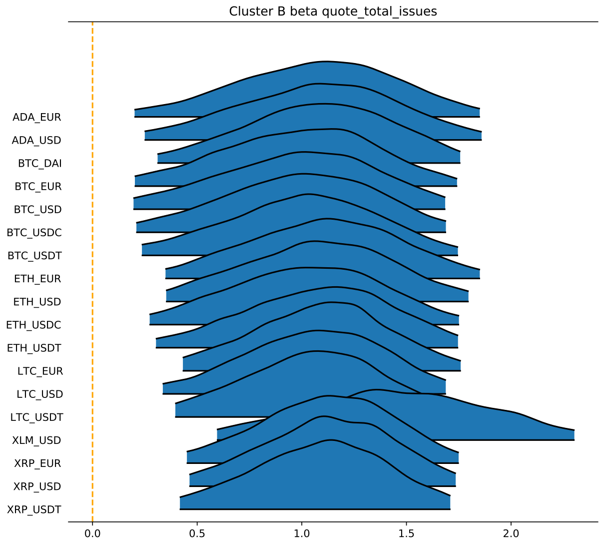

Figure 31: Forest plot currency project number of total issues 86400 data

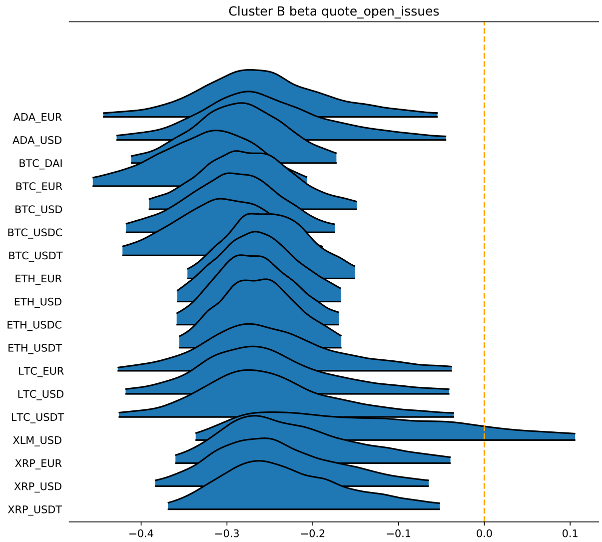

Figure 32: Forest plot currency project number of open issues 86400 data

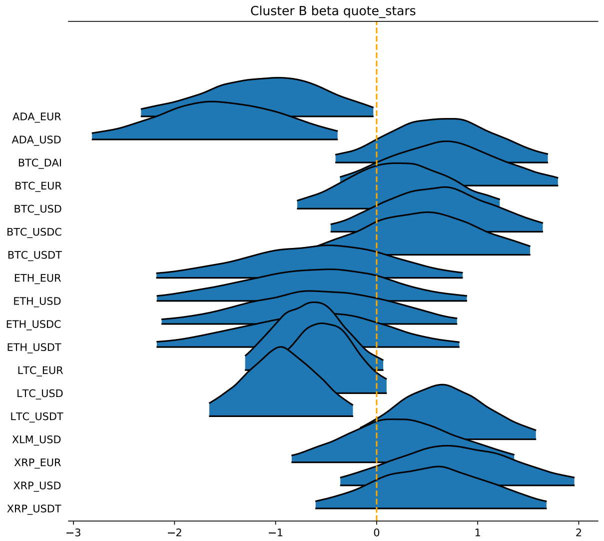

Figure 33: Forest plot currency project number of stars 86400 data

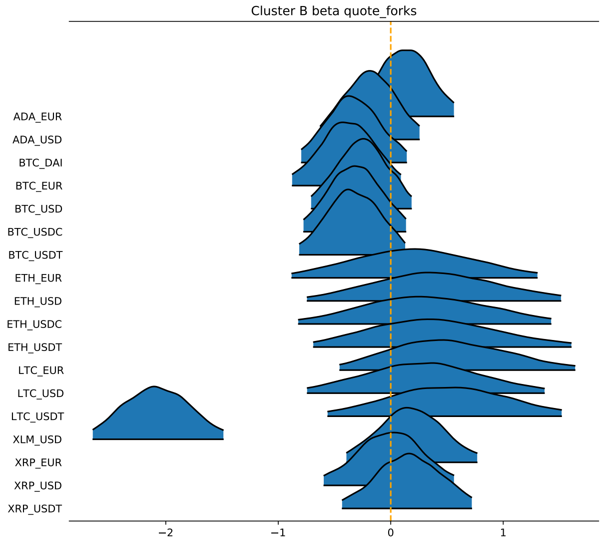

Figure 34: Forest plot currency project number of forks 86400 data

H3.d: When the number of forks increases, the price increases. Most currencies show a high degree of uncertainty around the 0 coefficient value. Some currencies might have a tendency to a negative value. However, this is not sufficiently significant or homogeneous. The hypothesis appears false.

Google Trends

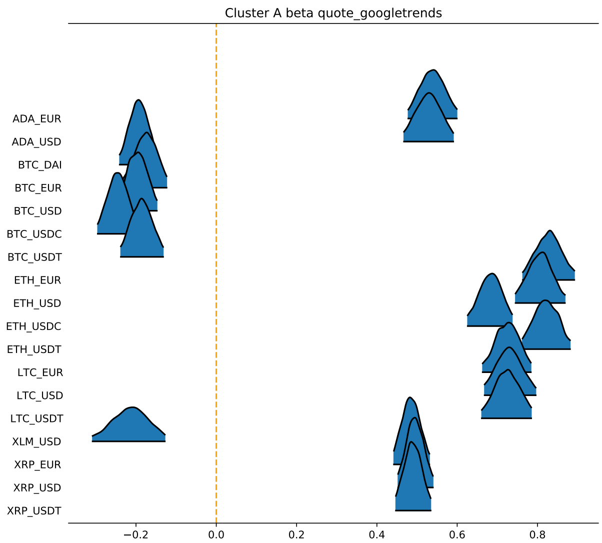

Figure 35: Forest plot currency Google trends 512 data

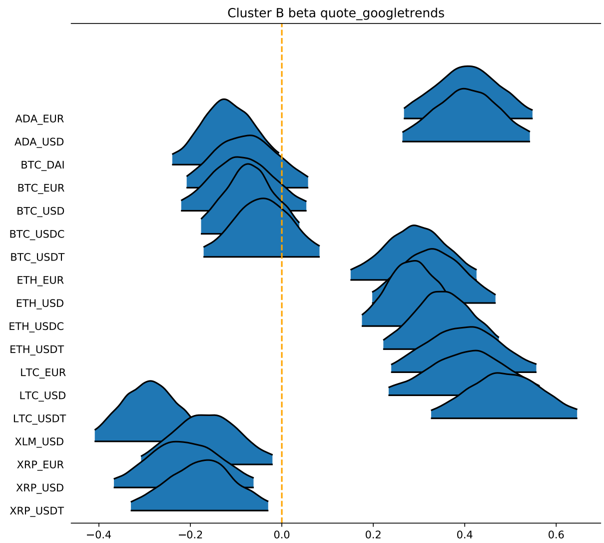

Figure 36: Forest plot currency Google trends 86400 data

For intra-day Fig. 35: Google trends for Bitcoin and Lumen appear to have a negative effect with mean coefficient values of -0.19 and -0.21 respectively. Google trends for Ethereum is a strongly positive effect with a mean value of 0.82. The p-direction and ROPE values indicate a high degree of signficance for all markets. The statistics for Google Trends on intra-day model can be found in Table 29 on p. 80 in Appendix C.

For inter-day Fig. 36: Bitcoin shows a less strong but still negative effect with a mean coefficient of -0.09. The p-direction values for Bitcoin have are in the range from 0.94 to 0.46 indicating uncertainty and potentially low significance. Ethereum is also still positive but less strong at 0.31. The coefficients for the remaining markets have reduced in magnitude as well compared to the intra-day. The statistics for Google Trends on inter-day model can be found in Table 49 on p. 90 in Appendix C.

Ripple is the only cryptocurrency which changed direction of coefficient, being positive for intra-day (0.48) and negative for inter-day (-0.18).

The hypothesis ***H4: When the Google Trends rating increases, the price increases.*** cannot be considered true in general. But for some markets, especially Ethereum, Cardano and Litecoin, it appears to hold. The observed differences between markets could potentially be explained by an unknown moderator variable changing the effect of the Google trends variable.

Community Size

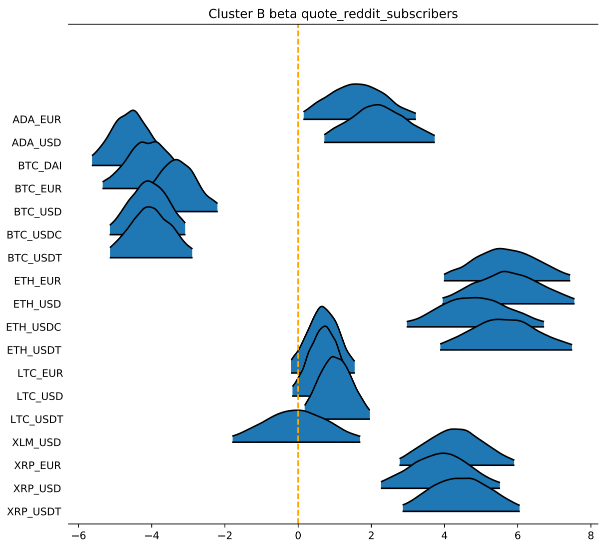

Figure 37: Forest plot number of Reddit subscribers 86400 data

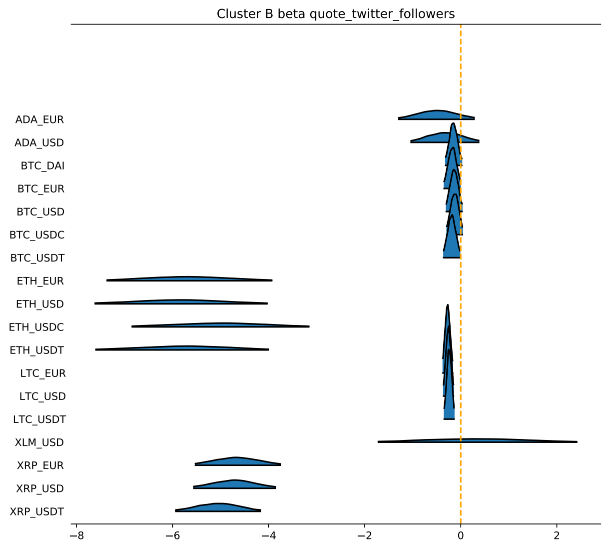

Figure 38: Forest plot number of Twitter followers 86400 data

Community size data for Reddit subscribers and Twitter followers is only available for 86 400 data.

The following observations are made based on Figures 37 and 38:

Lumen has a small coefficient of -0.06 for number of Reddit subscribers. The corresponding p-direction value is 0.04 indicating a high degree of uncertainty. The p-direction for Lumen on the number of Twitter followers is also low at 0.2. Cardano and Litecoin show a higher degree of significance but still have small coefficient values compared to the other markets.

Bitcoin has a negative coefficient of -4.1 for the number of Reddit subscribers, while the coefficient to the Twitter followers is comparatively smaller at -0.15. Ethereum has the largest positive coefficient with Reddit subscribers (5.6) and the largest negative coefficient with Twitter followers (-5.7). The Twitter followers coefficient probability distribution is rather spread out, indicating some uncertainty. Ripple follows a similar pattern as Ethereum, but with a reduced magnitude. The p-direction and ROPE coefficient values indicate a high degree of significance for Ethereum, Bitcoin, Litecoin and Ripple. This corresponds to the analyis of the hyper-priors Fig. 36 which identified Twitter followers as the overall most significant variable and Reddit subscribers as the fourth most significant variable.

The statistics for number of Reddit subscribers on inter-day model can be found in Table 52 on p. 92 in Appendix C. The statistics for number of Twitter followers on inter-day model can be found in Table 55 on p. 93 in Appendix C.

The hypothesis for this variable was ***H5: When the community size increases, the price increases as well.*** The hypothesis does not appear to be true. Not only are there differences in behaviour between currencies, but it appears that a positive coefficient with number of Reddit subscribers (the community size on Reddit) corresponds with a negative coefficient for the number of Twitter followers (the community size on Twitter). This might be interpreted as a hint that the delta of community size growth between the Reddit and Twitter community is a variable of significance. This could be a starting point for future research.

General Internet activity level

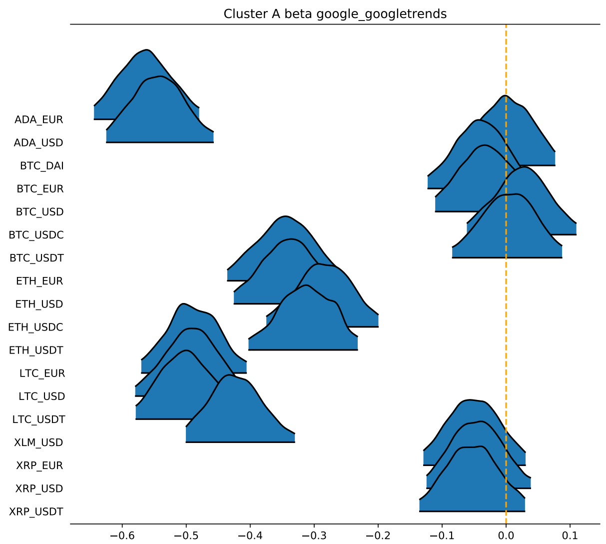

Figure 39: Forest plot Internet activity levels 512 data

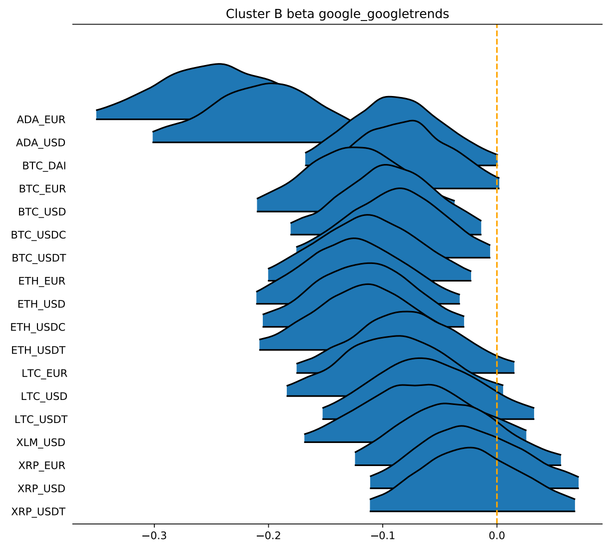

Figure 40: Forest plot Internet activity levels 86400 data

Fig. 39 shows the coefficient distributions for intra-day and Fig. 40 shows the coefficient distributions for inter-day. The observations made are:

Internet activity levels have a negative coefficient for almost all markets both on intra- and inter-day data. Ripple and Bitcoin show coefficient distributions close to zero. Their p-direction and ROPE coefficient indicate uncertainty and low significance. Cardano has the coefficient with largest magnitude, -0.56 on intra-day and -0.24 on inter-day. Litecoin, Lumen and Ethereum have coefficients ranging from -0.5 to -0.28 on the intra-day. Aside from Cardano, all markets appear to have low significance on the inter-day data, which is supported by a high ROPE coefficient value. The coefficient magnitudes are reduced on the inter-day compared to the intra-day model. This might be explained by the difference between days mostly being the result of weekends, while on the intra-day the daily activity cycles might have an effect.

The statistics for Internet activity levels on intra-day model can be found in Table 23 on p. 77 in Appendix C. The statistics for Internet activity levels on inter-day model can be found in Table 55 on p. 85 in Appendix C.

The hypothesis ***H6: An increase in global Internet activity corresponds to an increase in price*** appears to be false, rather the opposite seems to be the case. The effect is stronger for intra-day than for inter-day.

Twitter insights

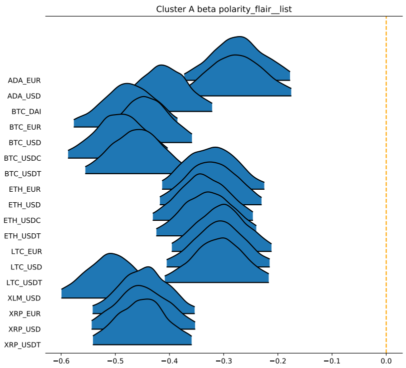

Figure 41: Forest plot mean Twitter sentiment 512 data

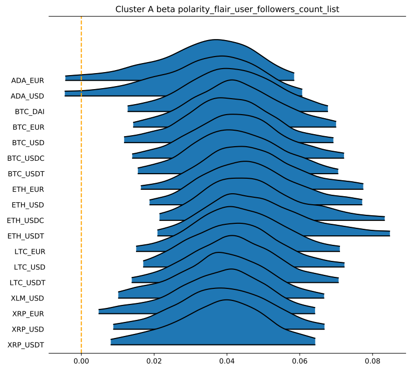

Figure 42: Forest plot mean Twitter sentiment, followers weighted, 512 data

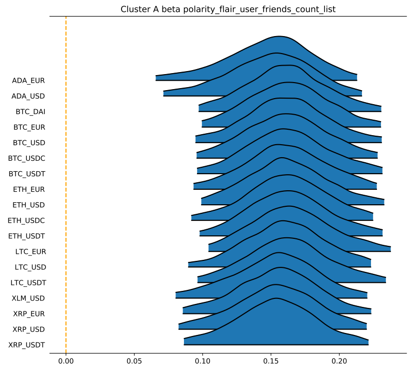

Figure 43: Forest plot mean Twitter sentiment, friends weighted, 512 data

The following observations are made based on the coefficients of the intra-day data as shown in Figures 41 to 43:

Lumen (-0.51), Bitcoin (-0.47) and Ripple (-0.45) have the largest coefficients and appear to be more influenced by sentiment than Ethereum (-0.32), Litecoin (-0.31) and Cardano (-0.27). The statistics for sentiment polarity on intra-day model can be found in Table 26 on p. 79 in Appendix C. It is noteworthy that for all markets the coefficients are significant and have a negative direction.

When weighting the mean sentiment by the number of users followed by the posting account (the number of friends), the direction changes to be positive. A fairly homogeneous distribution is observed for all markets with a mean value of 0.16. The statistics for friends weighted sentiment polarity on intra-day model can be found in Table 28 on p. 80 in Appendix C.

When weighting by the number of followers the coefficient distributions are fairly homogeneous as well. While the ROPE value is high, potentially indicating significance, the probability of direction is high as well. This indicates the variable to be of high significance but having a small magnitude of effect. The mean value is 0.04. The statistics for followers weighted sentiment polarity on intra-day model can be found in Table 27 on p. 79 in Appendix C.

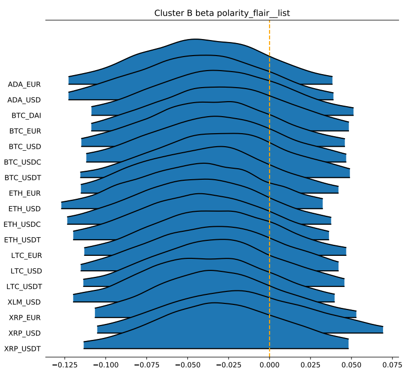

Figure 44: Forest plot mean Twitter sentiment 86400 data

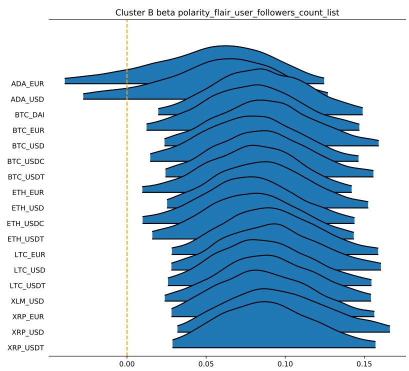

Figure 45: Forest plot mean Twitter sentiment, followers weighted, 86400 data

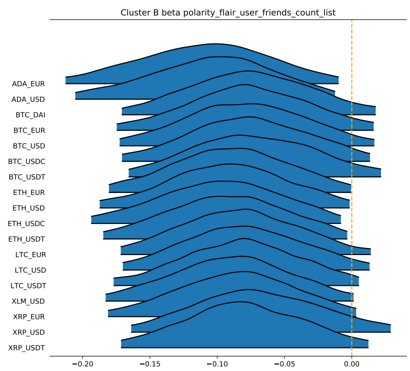

Figure 46: Forest plot mean Twitter sentiment, friends weighted, 86400 data

The following observations are made based on the coefficients of the inter-day data as shown in Figures 44 to 46:

Simple mean sentiment is observed to be more homogeneous than for intra-day. However, a high degree of uncertainty is indicated by a low p-direction value and a high ROPE value. This may indicate the effect to be uncertain for intra-day and likely is insignificant. The statistics for sentiment polarity on inter-day model can be found in Table 44 on p. 88 in Appendix C.

Weighting by followers has a mean magnitude of 0.08, which is an increase over intra-day. The p-direction values indicate signficance for all markets. The statistics for followers weighted sentiment polarity on inter-day model can be found in Table 45 on p. 88 in Appendix C.

Weighting by friends results in a negative coefficient with a mean value of -0.08. The p-direction is reduced compared to the followers weighted mean sentiment, which indicates a higher degree of uncertainty. However, the variable is still considered significant. The statistics for friends weighted sentiment polarity on inter-day model can be found in Table 46 on p. 89 in Appendix C.

Twitter sentiment appears to have a largely homogeneous effects for all markets.

The hypothesis ***H7: When Twitter sentiment becomes more positive, the price increases.*** might be considered true for intra-day effects, but false for inter-day effects. It furthermore depends on how the sentiment value is calculated. Weighting the mean sentiment of tweets by the number of the friends of the author seems to result in the most significant variable for intra-day, while on inter-day data weighting by number of followers is most significant.

Twitter Influencers

Since Influencers are lower in number than the general population of users, the number of their tweets is lower as well. The mean sentiment of Influencers could only be obtained for the inter-day data.

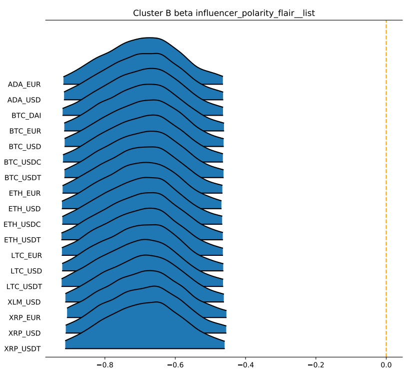

Figure 47: Forest plot mean Influencer Twitter sentiment 86400 data

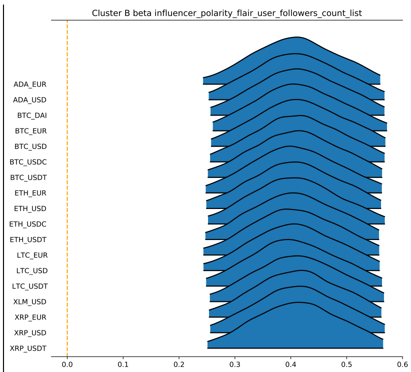

Figure 48: Forest plot mean Influencer Twitter sentiment, followers weighted, 86400 data

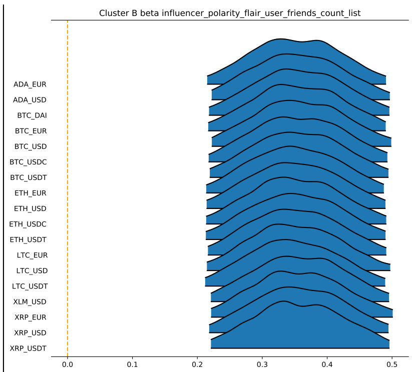

Figure 49: Forest plot mean Influencer Twitter sentiment, friends weighted, 86400 data

In Figures 47 to 49 we find:

Influencers show a homogeneous coefficient distribution of mean value -0.69. The influencers appear to have a larger effect than the general Twitter sentiment on inter-day data. The simple weighted average sentiment has a negative coefficient direction. The p-direction and ROPE both indicate a high degree of significance and certainty. The statistics for influencer sentiment polarity on inter-day model can be found in Table 39 on p. 85 in Appendix C.

Weighting by followers or friends results in similar coefficients for both, in terms of magnitude and spread. Followers weighting results in a mean coefficient value of 0.41 and weighting by friends results in a coefficient value of 0.35. Both coefficients show a high degree of significance and certainty. The statistics for followers weighted influencer sentiment polarity on inter-day model can be found in Table 40 on p. 86 in Appendix C. The statistics for friends weighted influencer sentiment polarity on inter-day model can be found in Table 41 on p. 86 in Appendix C.

The hypothesis ***H8: When key influencer sentiment becomes more positive, the prices increase*** appears inconclusive. When weighting by the amount of followers of the Influencer, it seems apparent that their sentiment has a positive effect. However, this effect appears to depend on the specific influencers. The group of influencers when weighted equally seems to have a negative effect. The hypothesis can therefore not be accepted as true.

4.4 Key Findings

Research question 3 aimed to identify which variables have the biggest influence on cryptocurrency price increase. While identifying a single killer variable would be nice, the reality appears to be more complex.

For the intra-day data the factors with the largest mean coefficient magnitude were found to be oil close (-0.56), gold volume (0.4), sentiment polarity (-0.39), google trends (0.36) and SP500 close (0.34). These coefficient values are similar in magnitude, which indicates no singular factor can be identified to explain the majority of changes. A model therefore will have to account for a multitude of variables which are all fairly important to consider.

On the inter-day data the picture changes slightly. This might in part be the result of more variables being available at this resolution. The largest hyper-prior mean values were observed for the variables of the number of Twitter followers (-1.7), the number of total issues (1.1), number of Reddit subscribers (0.7) and influencer sentiment polarity (-0.66). Similarly to the intra-day analysis this shows that a model needs to account for a multitude of factors. However, it also illustrates the importance of social media presence, influencer opinion and fundamental development progress for the cryptocurrency price. Unfortunately influencer sentiment, community size and Github statistics were also the three variable types unavailable at the 512 second intra-day data resolution. Since they showed some of the largest magnitudes for inter-day, they could also be important for intra-day. In future research these variables could be included by building customised real-time data monitoring systems. Building such a software was beyond the scope of this research. Section 3.4 elaborates on the challenges encountered during data collection.

In the following each hypothesis is discussed with the goal to answer **Research question 4**: What theories can explain the observed effect of the variables?

H1: When Volatility increases, the price increases. The opposite to the hypothesis was observed. Generally it makes intuitive sense that decreasing volatility could result in a price increase. Stability of cryptocurrency has long been discussed in social and news media alike. Increased stability of an asset can improve its ability to store value and thereby make it more attractive to investors. The hypothesis was based on the findings by King and Koutmos (2021), however, here an effect opposite to King and Koutmos (2021) is observed.

H2.a: The cryptocurrency market prices decrease when SP500 increases. The opposite to the hypothesis was observed. Especially for Bitcoin, Giudici and Abu-Hashish

(2019) has already established a significant correlation to conventional financial markets (represented here by SP500). The more established cryptocurrencies become, the more we might expect the same traders to be interacting with both conventional stocks and the new markets of cryptocurrencies. Research on herding behaviour of traders King and Koutmos (2021) indicates groups of traders often behave similarly.

H2.b: The cryptocurrency market prices decrease when Oil increases. There does not appear to be a clear pattern in either direction. The cryptocurrency prices seem largely unaffected by oil prices. This might be interpreted in multiple ways. First the oil price itself could potentially be considered as a somewhat "manipulated" market. Cartel-like organisations like OPEC and others come to price agreements and many other political interests are involved in price finding for oil. So in a way the lack of a relationship to oil indicates that the same forces do not act on cryptocurrencies. An exception is Bitcoin with a negative coefficient of largest magnitude (-1.1) on intra-day. This might be explained through energy consumption. Abbatemarco et al. (2018) have shown a relationship between the networks hashrate and its fundamental value. The hashrate is a result of many computers contributing processing power and consuming energy. Some part of the energy used by the Bitcoin network likely comes from oil or other correlated energy sources. A change to the oil price could effect the electricity cost of large mining operations resulting in changes to their profitability and with that the price point at which they would be willing to operate.

H2.c: The cryptocurrency market prices increase when Gold increases. While Bitcoin is often touted as "the digital gold", from the markets perspective this does not seem to be the case. Rather the opposite is true. Bitcoin might be considered as a hedge to gold. New investors interested in new forms of long term value storage might choose Bitcoin over Gold, resulting in a drop of Gold demand while increasing Bitcoin demand. However, this explanation might be more on the side of speculation. Further research will be required to obtain a strong theory on the relation between Gold and cryptocurrencies.

H3.a: When the number of closed issues increases, the price increases as well. The lack of a clear significant coefficient value here is surprising. Wang and Vergne (2017) have shown Github statistics to be an indicator for fundamental value growth of cryptocurrencies. The number of closed issues represents the total amount of concluded work by the project. The analysis results from Section 4.3 showed the coefficients for number of total issues to be much more positive and significant. The total number of issues is the sum of closed issues and open issues. It could be that the open issues signify future technical improvements which are already priced in. This might indicate the total number of issues is a better indicator than number of closed issues for the total long-term development progress or fundamental value of the cryptocurrency.

H3.b: When the number of open issues decreases, the price increases. The hypothesis appeared to be true. This might confirm the theory that a reduction of currently open issues indicates short-term development progress and thereby makes the cryptocurrency an attractive investment leading to a price increase.

H3.c: When the number of stars increases, the price increases and ***H3.d: When the number of forks increases, the price increases*** both do not hold. The relationships are too varied and not significant enough to propose possible explanations. The variance might be a result of the differences with which the different projects manage and organise their development efforts. There are different possible approaches in how version control systems like Github are utilised. The projects approach to organising development might be a moderating variable that should be investigated in future research.

H4: When the Google Trends rating increases, the price increases. Bitcoins negative intra-day coefficient might potentially be because it is the best known cryptocurrency, and media reporting might more commonly report on significant drops in value. From the author's personal observation, it appears that negative news has a wider reach in mainstream media, while positive news is often limited to niche and social media. The increased awareness in the mainstream through negative reporting could result in people searching for Bitcoin to search for more details on what has happened after the price reducing event already has occurred. Ethereum on the other hand is less well known and receives nearly no mainstream media attention. Its positive coefficient might indicate more serious/researching investors informing themselves about its properties, or following largely positive news mentions. Ripple has a positive coefficient for intra-day but a negative coefficient on inter-day, which indicates they might be driven by short-term hype, but over the long term investors could become disillusioned.

The observed effect needs to be investigated further to analyse potential interactions with news and other types of events. The reason someone is searching for a specific cryptocurrency might turn out to be more important than the number of searches recorded. This reason for searching could be considered a moderating variable for the effect the search will have. News events, as indicated by Baig et al. (2019), could be catalysts for searches being biased in some way.

H5: When the community size increases, the price increases as well. Bitcoin shows little effect from the number of Twitter followers, but a strong negative effect from an increasing number of Reddit subscribers. This could potentially be the result of its status as a cryptocurrency gateway. It is commonly the most well known cryptocurrency and is an entry point for research by many investors and users. During such research investors will likely learn about various other cryptocurrencies. These might be attractive in their own right, or used to hedge against risk, but in effect lead the investor to convert some of their Bitcoin holdings into other currencies and thereby using Bitcoin only as a gateway. The magnitude of the observed effect is larger than we would expect to see as a result of the explanation above. So there might also be some other variable at play. The most plausible explanation is that the dataset has simply captured a time of continual decline of Bitcoin price. Ethereum shows to benefit strongly from more Reddit subscribers, but appears to be negatively affected by the number of Twitter followers.

This is an unexpected finding overall. The expectation that different internet communities behave similarly appears to be false. Existing literature by Aggarwal et al. (2019) and Li et al. (2019) focuses on the effects of Twitter. Reddit appears to be somewhat ignored by the research in cryptocurrencies. However, the findings shown here, indicate that Reddit is a significant force on par with Twitter. Future research on finding differences in the dominant user groups of Twitter and Reddit appears highly interesting.

H6: An increase in global Internet activity corresponds to an increase in price. The hypotheis is false. The highest levels of internet activity are generally when people are at work or otherwise occupied. So it could make sense that a drop in internet activity corresponds to people being at home or otherwise having more free time. Non-professional retail investors might then make use of this time by trading cryptocurrencies online. Ripple and Bitcoin appear to be the least affected by this variable. They are also the currencies most commonly associated with institutional investors, thus diminishing the effect of retail investors daily activity cycles. More research into the behaviour of retail investors is needed also beyond the cryptocurrency markets, as access to conventional financial markets becomes increasingly easier and retail investors coordinating their efforts could be a threat to the established market participants (Boylston et al., 2021).

H7: When Twitter sentiment becomes more positive, the price increases. Weighting the mean sentiment by number of friends appears to be the most useful metric. Conventionally a distinction is made between Influencers (accounts followed by many other accounts) and the general user (low number of followers). The results of this research indicate there might be a third significant category of information aggregators as key elements of social networks. Accounts who follow a lot of other accounts, might be considered information aggregators, and as they re-post findings, their sentiment could have a higher informational density.

H8: When key influencer sentiment becomes more positive, the prices increase. To some degree the expected effect was observed for influencers, but also depending on the weighting, the opposite effect could be seen. Overall this seems to indicate the effect of an influencer is specific to the individual influencer. Their potentially differing groups of followers might have varying degrees of interest and purchasing power and might react differently strong to the influencers personal sentiment. While further research might be interesting in this line of thought, it also seems difficult to arrive at a generally applicable model based on Influencers.

The above described theories are generated based on the results of the hierarchical Bayesian regression model. However, some theories, for example for hypothesis 4, required the inclusion of additional anecdotal evidence as part of their explanation since the analysis result alone could not yet explain the observed effects. These theories should be taken as starting points for future research to test their assumptions and to expand current understanding of cryptocurrency market dynamics.

Furthermore there are some noteworthy differences between the cryptocurrency markets and conventional financial markets. The Social Media sentiment and community size have shown themselves to be highly significant variables. Investors in conventional financial markets may benefit from including sentiment and community size in their models as well. However, depending on the domain, not every company will have a sufficiently strong user-engagement on social media to allow for analysis. Increased globalisation and importance of social media will likely require not just big players but also nieche and small companies to increase their social media presence to stay competitive (Holt, 2016). Future research could help identify best practices to follow to leverage the network effects enabled by social media.

Another interesting aspect is the approach to intellectual property as it is practised by cryptocurrencies projects. Most conventional organisations try to maintain a competitive advantage through intellectual property protection (Reitzig, 2004). However, making the source code publicly available in cryptocurrencies and other software projects can be a competitive advantage to leverage the communities feedback and code contributions and to provide a proof for trustworthiness (Maurer & Scotchmer, 2006). Combined with the ability to automate data collection and monitor variables around the clock, cryptocurrency markets have attractive properties without an equivalent alternative in conventional financial markets. It is the author's expectation that in the future there will be an increasing number of organisations offering cryptocurrency based and open source solutions to problems currently addressed by closed source established corporations. This could become an existential threat to those too slow or unable to innovate their business models and offerings.

5 Conclusion, Limitations and Recommendations

5.1 Conclusion

This research was driven by four research questions. The first question aimed to identify the set of variables influencing cryptocurrency prices. A range of variables were identified from literature in Section 2. However, not all could be included for analysis in this research. A selection of variables to be analysed further was made in Section 2.4. The second research question aimed to identify the magnitude and direction of each variables effect. This was achieved by analysing the coefficient distributions of each variable in a hierarchical Bayesian linear regression model. The results (Section 4) showed the variables had a multitude of effects on different markets. Especially noteworthy is the differing effect of community size on Reddit vs Twitter. Furthermore, cryptocurrency prices were found not to be independent of conventional financial markets. Influencers were found to be more important drivers of sentiment than the general population as a whole. A potentially novel group of information aggregators, who themselves follow many other users, were identified and found to be significant.

The third research question, identifying the variables with the biggest influence on price increase, can unfortunately not be answered with a single selected variable. The results showed that a range of variables interact with the price and no single variable is exclusively responsible for describing the changes over time. A model for estimating the cryptocurrency price will have to make use of a multitude of variables and with the current set of variables analysed there still remains significant variance among the variables effect on different markets.

This does not yet provide strong indication of causation of any singular variable. However, this research did take a step towards showing that the market dynamics are caused by a composite of variables that together form a social dynamic system influencing the markets. News media, social media and development efforts all affect markets at once and arguably also each other in a never ending feedback-loop. The price estimate model achieved a mean R^2 of 0.69 for intra-day and a mean R^2 of 0.91 for inter-day.

The theories in Section 4.4 proposed explanations for the observed variable effects and answered the forth and final research question. Some of the explanations also indicate the likely presence of still unidentified moderating variables for several effects. Without identifying more of the currently unknown variables, it is not possible to create a single definitive model which holds for all cryptocurrencies alike.

Most of the established hypothesis in Section 3.1 were concluded to be false, which on the one hand shows the difficulty of explaining market dynamics, but on the other hand points to a range of new research opportunities, which are summarised in Section 5.4. Especially promising might be the application of a hidden state markov chain model to this field, as it will allow unknown variables to be estimated without having to identify them beforehand.

Development metrics like the total number of issues and the number of open issues of cryptocurrency projects were found to be influential on price. The research has shown that focusing on developing the core technology and building the community around it are some of the most effective ways to boost a cryptocurrencies price.

This research has identified a number of relationships between variables and cryptocurrency price and proposed theories for them. Existing research is often focused on a very narrow aspect of a specific variable or a specific part of the social system surrounding cryptocurrencies. This may help increase the scientific validity of the research, but the draw back of this approach is that it might miss the big picture of relationships between variables. Especially when the rapid pace of technological innovation changes the way society interacts as a whole, it might be prudent to take a step back and observe the bigger picture. The main scientific contribution of this research is combining the cross-disciplinary variables from social, financial and managerial fields into one model. Much of the proposed future research from Section 5.4 includes identification of additional variables, which can then be integrated back into the foundation laid in this research to build a deep understanding of the dynamic systems across multiple scientific fields.

5.2 Limitations

There are plenty of other relationships to still be found in the data. Some of the observed relationships might not be causally linked, but could be caused or moderated by an unknown third variable. This could potentially be addressed in the future by applying a more advanced modelling technique using a hidden state markov chain model, instead of the current markov chain monte carlo Bayesian inference approach.

The analysis is based on data covering only a fairly short time-span of roughly 5 months. During this time cryptocurrency markets went through a phase of extra-ordinary change with a significant influx of investors and media interest. This coincided with a global pandemic, following an unprecedented decade of economic growth on a planet with mankind exponentially increasing in numbers for over a century and access to the Internet reaching an all time high. So it is an all around unprecedented time and just recording data at different times might cause significant sampling error leading to different results all-together.

Throughout the analysis markets are assumed to be efficient and that the true price value is only influenced by the exogenous variables. This assumption should be quantified in future research.

Some of the findings showed a large degree of homogeneity of coefficients for some of the variables. These are the most likely to still hold if the analysis were repeated using a different dataset. The rapid change of cryptocurrency markets also threatens this research to become outdated. This might be unavoidable when doing research on a highly dynamic system which is driven by changing social dynamics. Where the dynamics are both drivers and the result of the social evolution in this niche social system. However, the research is still important as a record of the dynamics at this time. The analysis methodology can, once implemented, be repeated continually to observe changes to the coefficients over time.

Expanding the model with further variables and markets could further increase the chance to come to long-term valid results. One of the main drawbacks of the current analysis is the difficulty to prove the causality of the variable effects. While some speculation was done, each of the explanatory theories requires its own full research and analysis to confirm the reasoning. This research is therefore not to be taken as the end-all model on cryptocurrency market dynamics, but could be a starting point for future research on the individual variable theories.

5.3 Recommendations for practitioners

Based on the findings of this research a number of recommendations may be made. One of the objectives of this research is to help managers and investors of cryptocurrency projects to reduce their risks. Cryptocurrency projects issuing their own token or cryptocurrency will be exposed to market forces which are considered an external risk to the project success, as was described in Section 1. The risk of a changing cryptocurrency price is of interest to both project insiders and to project investors. The following suggestions apply to both investors and managers of cryptocurrency projects (the practitioners). These may help to evaluate cryptocurrencies as potential investments and may allow the managers to compare their projects to other competitors.

- Inter-day changes appear more easily estimatable than intra-day.
- The General Twitter sentiment may be monitored to gain a broad feedback. However, the method by which sentiment is aggregated affects the interpretation significantly.
- Twitter influencers are still important, but the opinion of the crowd matters, especially that of aggregators, who are users who follow many other users.
- The effect of Google trends score is not well understood yet. While search engine optimisation is likely helpful, there are cases where increased numbers of searches coincide with decreases in price. The reason why people Google the project might be important. Is it to find confirmation bias on negative news, or to learn about the potential of new features?
- Macro-economic variables are big indicators for prices. It is important to keep an eye on economic developments overall, not just the cryptocurrency competition. Some downturns of the cryptocurrency prices might be explained by downturns of the broader economy. Cryptocurrency markets are not independent of the economy as a whole. Care needs to be taken if investments are to serve the purpose of hedging against risks of the conventional financial markets.

In addition, the following recommendations are made to cryptocurrency managers directly. They can also be used as a metric by investors, but appear especially powerful when considered in project internal decision making processes. Promoting the observed effects within the cryptocurrency project is expected to largely be under the control of the manager.

- Community size appears a very significant variable for inter-day dynamics, however, it also showed a wide range of effects. But it seems self apparent that a good community makes development of open source projects better. So the project should establish guidelines and frameworks to facilitate a constructive discorse. This could help both with development (bug-testing or ideation) and improving the general social sentiment.
- Increasing the number of total issues on Github indicates long term growth (showing progress long-term).
- Reducing the number of open issues shows short-term growth (Quick response to issues and short-term progress).
- A reduction in Volatility leads to an increased cryptocurrency price.

5.4 Future research

The following topics are suggested as potentially fruitful subjects to conduct future research on:

- The Neutral Model of Evolution by ElBahrawy et al. (2017) appears a highly interesting approach to model the market share of various cryptocurrencies. This research should be repeated on larger and more recent datasets.
- The PCA model from Shah et al. (2021) is another candidate for repeated research. It could help further compare influencing variables and potentially even find hidden unknown variables.
- The lack of systematic and automated tracking of national regulation makes accounting for their effects difficult. A new framework for analysing the regulatory impact on various cryptocurrency projects could help identify another important external variable.
- Aside from regulation, there are a number of variables whose reasons for exclusion were driven by the current state of data (API) availability. As more APIs might become available in the future, additional variables can be included in future repeats of this analysis. These include: Hashrates, opening times of exchanges, price number roundness, news and other discrete events.
- Some of the variables in this research were only available at daily resolution. This could be improved by building a real-time monitoring system, for example for the Github statistics or number of Twitter followers.
- In future research the effect of non-linear behaviour should be expanded upon. Knowing the amount of effect from non-linearities could help further reduce the uncertainty associated with some variables.
- The analysis focuses on fiat currency denominated markets to allow for comparison between markets. Further interesting relationships might be found in analysing crypto markets denominated in terms of other cryptocurrencies.
- The analysis of global Internet activity levels is using the Google trends score of the keyword “google” as a metric. However, there might be alternative ways to measure global Internet activity that could be better suited to cryptocurrency markets or even other research altogether.
- There could be other, as of yet, unidentified variables to help improve the explanatory power of the models and to explain the variance between cryptocurrencies. More advanced modelling techniques, like hidden state Markov chains, could help estimate those variables. Then further research could try to identify the unknown variable based on its estimated data.
- Social media communities were identified as highly important for cryptocurrency market dynamics. Future research should analyse social media behaviour in more depth to identify best practices for companies and organisations to follow in order to leverage the network effects they enable.

Finally it is important for the findings from the future research suggestions to be integrated back into the cross-disciplinary model of this research. Drawing from a multitude of scientific backgrounds appears to have the best chance of explaining cryptocurrency market dynamics.

6 Reflection

The findings from this research turned out quite different from what I had originally expected. My prior experience with cryptocurrencies informed the hypothesis generation and I turned out to be wrong on some of them, or at least not completely right. I was initially confident to create a model describing the price changes of cryptocurrency markets. However, it was discovered that there is much more to the story. There must still be more, currently unknown, variables with big effects on markets, as the model cannot explain the observed price volatility completely. The amount of variance observed between currencies was surprising, but is in itself a satisfactory result as well.

Compared to the planned timeline created at the beginning of the project, Appendix D, the process to arrive at the version of the thesis in-front of the reader today, was significantly more iterative than expected. The data aggregation took longer than anticipated and the methodology was adjusted several times throughout to increase the value of the findings from all that collected data. I would like to thank my thesis supervisors Prof. Dr. Ir. Marijn Janssen and Dr. Ir. Zenlin Roosenboom-Kwee for their advice and patience! Their guidance has contributed significantly to refine and focus this research and increase its scientific value.

My interest in cryptocurrencies and blockchain as a technology has been ongoing for several years already. I first got the opportunity to dive deeper into the field in an academic context in the course "Emerging and Breakthrough technologies". As part of the course a project was conducted in which the innovation diffusion of Ethereum was analysed (Sachs et al., 2021).

Approaching cryptocurrency from an academic perspective was a great experience. There is plenty of opportunity for new academic research and the amount of research published in this field has only started to increase in the last two or three years (Fig. 2). The blockchain technologies are not just a new technology but have formed new social dynamics around their development and might turn out to be one of the most important technologies of the 21st century. I am thankful and proud to contribute to the research into this phenomenon.

The MOT TPM courses of Data Analytics, Web-Sciences, Economic Foundations and Financial Management were all high influential to guide me through the process of this research and inform the decisions made throughout. Especially thinking across multiple scientific disciplines was required, as cryptocurrency markets do not exist as an isolated system, but are influenced by social effects from news and social media as well as financial markets and technical innovation systems. However, some problems were still encountered.

One of the most challenging aspects of this research was the data collection. Google trends for example imposes strict rate limits on data requests, and some data, like tweets, could only be recorded in real-time. This unfortunately resulted in the dataset only being able to cover roughly 5 months of data. Aside from data availability, big data is difficult to handle but very rewarding in the end. New software had to be written to combine and aggregate data correctly into a dataset before any analysis could even begin. I would have liked to include more variables in the analysis, but the complexity of the process scales exponentially with the number of variables included, becoming increasingly harder to manage data and analysis. Overall 4349 lines of code were written for this project, not including proprietary code used from Arkmon. Knowledge about variables from this research will be useful to guide the expansion and improvement of the `Arkmon.eu` trading system in the future.

In the variable data collection I was also relying on third party data collectors. This is not sufficient as they all record data at own rates and have different policies on public data accessibility. If I were to do this research again, I would build a complete data tracking system in-house that can monitor each variable directly without requiring any third parties. Throughout the data collection process I put an emphasis on using open data sources. Open source intelligence is a fascinating concept to me and the plethora of data freely available today is greater than ever before. To contribute to the open data movement, the aggregated dataset compiled for this Master Thesis is also being made publicly available for anyone to use at the link given in Section 3.5. This also aids in the replicability of the presented results. Replication is a crucial part of the scientific process that is especially difficult when a lengthy data collection process is required. So providing the dataset should make it easier for anyone who is interested to repeat the analysis and since not all the collected data did end up being used, there might even be additional future findings possible from the same dataset.

Overall I have learned a lot about doing research, cryptocurrencies and new analysis methodologies and I will continue to pursue the objective of understanding cryptocurrency market dynamics for the foreseeable future as part of expanding the `Arkmon.eu` service. If you have any questions or remarks feel free to contact me at: `contact@arkmon.eu`.

References

Abbatemarco, N., De Rossi, L., & Salviotti, G. (2018). An econometric model to estimate the value of a cryptocurrency network. the bitcoin case.. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061317208&partnerID=40&md5=2dd77024b05eda2f14c9571d6682aa89 (cited By 1)

Aggarwal, A., Gupta, I., Garg, N., & Goel, A. (2019). Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction.. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073208227&doi=10.1109%2fIC3.2019.8844928&partnerID=40&md5=d3cdafe662abc3bad8ed85d84caab21b (cited By 5) doi: 10.1109/IC3.2019.8844928

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). Flair: An easy-to-use framework for state-of-the-art nlp. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics (demonstrations)* (pp. 54–59).

Akyildirim, E., Corbet, S., Cumming, D., Lucey, B., & Sensoy, A. (2020). Riding the wave of crypto-exuberance: The potential misusage of corporate blockchain announcements. *Technological Forecasting and Social Change*, *159*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087967517&doi=10.1016%2fj.techfore.2020.120191&partnerID=40&md5=db8591a6a79890110171a34f4eb7d4df (cited By 3) doi: 10.1016/j.techfore.2020.120191

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons and Fractals*, *126*, 325-336. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069566340&doi=10.1016%2fj.chaos.2019.07.011&partnerID=40&md5=e910c8061dba09317f7f81dce98a9e0b (cited By 100) doi: 10.1016/j.chaos.2019.07.011

Baig, A., Blau, B., & Sabah, N. (2019). Price clustering and sentiment in bitcoin. *Finance Research Letters*, *29*, 111-116. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062823903&doi=10.1016%2fj.frl.2019.03.013&partnerID=40&md5=d6e78d0dbc0ed8f6eff3d2f7fdad1a29 (cited By 13) doi: 10.1016/j.frl.2019.03.013

Bejaoui, A., Ben Sassi, S., & Majdoub, J. (2019). Market dynamics, cyclical patterns and market states: Is there a difference between digital currencies markets? *Studies in Economics and Finance*, *37*(4), 585-604. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076129465&doi=10.1108%2fSEF-08-2019-0302&partnerID=40&md5=be43ad4e20854d305154c182be8686db (cited By 1) doi: 10.1108/SEF-08-2019-0302

Boylston, C., Palacios, B., Tassev, P., & Bruckman, A. (2021). Wallstreetbets: Positions or ban. *arXiv preprint arXiv:2101.12110*.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic record*, *88*, 2–9.

coinmarketcap.com. (2021). Cryptocurrency market capitalisation. Retrieved from https://coinmarketcap.com/charts/ (accessed on May 12, 2021)

De Vries, A. (2018). Bitcoin's growing energy problem. *Joule*, *2*(5), 801–805.

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., & Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. *Royal Society open science*, *4*(11), 170623.

Eross, A., McGroarty, F., Urquhart, A., & Wolfe, S. (2019). The intraday dynamics of bitcoin. *Research in International Business and Finance*, *49*, 71-81. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062281467&doi=10.1016%2fj.ribaf.2019.01.008&partnerID=40&md5=f3e4fa0ff37031ab6b48e35a77e2b7f4 (cited By 24) doi: 10.1016/j.ribaf.2019.01.008

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, *48*(3), 432–435.

Giudici, P., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? a network var approach. *Finance Research Letters*, *28*, 309-318. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049298848&doi=10.1016%2fj.frl.2018.05.013&partnerID=40&md5=c569fba101bc65b29a250c013d3e1851 (cited By 46) doi: 10.1016/j.frl.2018.05.013

Holt, D. (2016). Branding in the age of social media. *Harvard business review*, *94*(3), 40–50.

Jesson, J., Matheson, L., & Lacey, F. M. (2011). *Doing your literature review: Traditional and systematic techniques*. Sage.

Jiang, S., Li, X., & Wang, S. (2021). Exploring evolution trends in cryptocurrency study: From underlying technology to economic applications. *Finance Research Letters*, *38*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084853547&doi=10.1016%2fj.frl.2020.101532&partnerID=40&md5=e53ed74403dcf253b4397284199dcd19 (cited By 0) doi: 10.1016/j.frl.2020.101532

King, T., & Koutmos, D. (2021). Herding and feedback trading in cryptocurrency markets. *Annals of Operations Research*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099478985&doi=10.1007%2fs10479-020-03874-4&partnerID=40&md5=ab08c2c66491045debd1827ecca6cd1a (cited By 0) doi: 10.1007/s10479-020-03874-4

Klarin, A. (2020). The decade-long cryptocurrencies and the blockchain rollercoaster: Mapping the intellectual structure and charting future directions. *Research in International Business and Finance*, *51*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069944517&doi=10.1016%2fj.ribaf.2019.101067&partnerID=40&md5=c1546f172e14ff0f57dd5bfe6ff1b487 (cited By 13) doi: 10.1016/j.ribaf.2019.101067

Kruschke, J. (2014). Doing bayesian data analysis: A tutorial with r, jags, and stan.

Li, T., Chamrajnagar, A., Fong, X., Rizik, N., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, *7*(JULY). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069882306&doi=10.3389%2ffphy.2019.00112&partnerID=40&md5=0ef805564c3e6cfeb4739b9ccbd6e025 (cited By 9) doi: 10.3389/fphy.2019.00112

Liang, J., Li, L., Chen, W., & Zeng, D. (2019). Towards an understanding of cryptocurrency: A comparative analysis of cryptocurrency, foreign exchange, and stock. In (p. 137-139). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072969911&doi=10.1109%2fISI.2019.8823373&partnerID=40&md5=08cc6416f2c23d01fa7943c8a70a8866 (cited By 2) doi: 10.1109/ISI.2019.8823373

Liang, J., Li, L., Zeng, D., & Zhao, Y. (2018). Correlation-based dynamics and systemic risk measures in the cryptocurrency market. In *2018 ieee international conference on* *intelligence and security informatics (isi)* (pp. 43–48).

Liu, Z., Ding, Z., Lv, T., Wu, J., & Qiang, W. (2019). Financial factors affecting oil price change and oil-stock interactions: a review and future perspectives. *Natural Hazards*, *95*(1-2), 207-225. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053467597&doi=10.1007%2fs11069-018-3473-y&partnerID=40&md5=d143c2d4aa8ea3f52ec5f941ceda7d2c (cited By 5) doi: 10.1007/s11069-018-3473-y

López-Cabarcos, M., Prez-Pico, A., Pieiro-Chousa, J., & evi, A. (2021). Bitcoin volatility, stock market and investor sentiment. are they connected? *Finance Research Letters*, *38*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077703057&doi=10.1016%2fj.frl.2019.101399&partnerID=40&md5=5885ebc5138947357edfdeda77dc4f3e (cited By 5) doi: 10.1016/j.frl.2019.101399

Makowski, D., Ben-Shachar, M. S., Chen, S., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in psychology*, *10*, 2767.

Maurer, S., & Scotchmer, S. (2006). *Open source software: the new intellectual property paradigm.* National Bureau of Economic Research Cambridge, Mass., USA.

Miles, J. (2005). R-squared, adjusted r-squared. *Encyclopedia of statistics in behavioral science*.

Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. *Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf*, *4*.

nasdaqtrader.com. (2021). NASDAQ Trader U.S. Market Volume. Retrieved from http://www.nasdaqtrader.com/ (accessed on May 12, 2021)

Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S., ... Gandomi, A. (2020). Data science in economics: Comprehensive review of advanced machine learning and deep learning methods. *Mathematics*, *8*(10), 1-25. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086140360&doi=10.3390%2fmath8101799&partnerID=40&md5=1c8af62eb552ab6906f2aef25ec16f1e (cited By 6) doi: 10.3390/math8101799

Panova, O., Leheza, Y., Ivanytsia, A., Marchenko, V., & Oliukha, V. (2019). International models of legal regulation and ethics of crypto currency use: Countr y review. *Journal of Legal, Ethical and Regulatory Issues*, *22*(Special Issue 2). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071442390&partnerID=40&md5=554e178129a5663b7bda161d25fe8136 (cited By 0)

Rahman, S., Hemel, J., Junayed Ahmed Anta, S., Muhee, H., & Uddin, J. (2019). Sentiment analysis using r: An approach to correlate cryptocurrency price fluctuations with change in user sentiment using machine learning. In (p. 492-497). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063224786&doi=10.1109%2fICIEV.2018.8641075&partnerID=40&md5=9c95d793f91b4f8263929d8f6c14c7f2 (cited By 2) doi: 10.1109/ICIEV.2018.8641075

Reitzig, M. (2004). Strategic management of intellectual property. *MIT Sloan Management Review*, *45*(3), 35.

Sachs, M., Balaji, S., Furtado, G., Sathyanarayanan, N., & Wong, E. (2021). Innovation diffusion analysis of the ethereum protocol. *MOT2421 Emerging and Breakthrough Technologies*.

Shah, A., Chauhan, Y., & Chaudhury, B. (2021). Principal component analysis based construction and evaluation of cryptocurrency index. *Expert Systems with Applications*, *163*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85088916238&doi=10.1016%2fj.eswa.2020.113796&partnerID=40&md5=e7c84cd1cd561bf8645b2387c50d924c (cited By 0) doi: 10.1016/j.eswa.2020.113796

Shanaev, S., Sharma, S., Ghimire, B., & Shuraeva, A. (2020). Taming the blockchain beast? regulatory implications for the cryptocurrency market. *Research in International Business and Finance*, *51*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85070747626&doi=10.1016%2fj.ribaf.2019.101080&partnerID=40&md5=04820c9dfa9efee1d6abb5e81e2c56a3 (cited By 12) doi: 10.1016/j.ribaf.2019.101080

Thampanya, N., Wu, J., Nasir, M., & Liu, J. (2020). Fundamental and behavioural determinants of stock return volatility in asean-5 countries. *Journal of International Financial Markets, Institutions and Money*, *65*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083648448&doi=10.1016%2fj.intfin.2020.101193&partnerID=40&md5=acb9dc022bc8e6b4fedd27c543a1530d (cited By 4) doi: 10.1016/j.intfin.2020.101193

Wang, S., & Vergne, J.-P. (2017). Buzz factor or innovation potential: What explains cryptocurrencies' returns? *PLoS ONE*, *12*(1). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85009420574&doi=10.1371%2fjournal.pone.0169556&partnerID=40&md5=acc76c1996b23aed67c58a57d92c43bf (cited By 31) doi: 10.1371/journal.pone.0169556

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems*, *37*(2). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075481457&doi=10.1111%2fexsy.12493&partnerID=40&md5=4521aea0c0e1b05f1fc51cc5884e8be9 (cited By 2) doi: 10.1111/exsy.12493

Appendices

Appendix A Tracked Twitter user ids

The following users Tweets are being recorded as part of the Twitter sentiment analysis:

```
[813286, 27260086, 79293791, 25365536, 428333, 759251, 44409004, 807095, 44196397,
    ↪ 268414482, 145125358, 85603854, 11348282, 471741741, 23151437, 14872237, 96951800,
    ↪ 116362700, 822215679726100480, 627673190, 119509520, 19923144, 28706024, 209708391,
    ↪ 1339835893, 169686021, 113419517, 35094637, 939091, 115485051, 822215673812119553,
    ↪ 343627165, 558797310, 19895282, 92724677, 10671602, 224223563, 20609518, 16303106,
    ↪ 3004231, 500704345, 35936474, 525264466, 76294950, 1492538024, 866953267, 34613288,
    ↪ 268439864, 24929621, 16548023, 19583545, 2367911, 28785486, 112764971, 22910295,
    ↪ 24742040, 344634424, 18220175, 17461978, 207809313, 34743251, 37034483, 580097412,
    ↪ 27042513, 24886570, 56783491, 62513246, 140070953, 45709328, 46452468, 19028953,
    ↪ 185142711, 14159148, 1330457336, 51376979, 326359913, 18222378, 885487044243238912,
    ↪ 1636590253, 264107729, 15439395, 30313925, 30364057, 49573859, 23690344, 2097571,
    ↪ 452155423, 2883841, 130104041, 254274083, 7517222, 23605048, 155927976, 338084918,
    ↪ 16193578, 36412963, 818910970567344128, 403245020, 403246803, 15687962, 104249727,
    ↪ 185827887, 266336410, 122969292, 23186079, 152251488, 573299587, 44670915, 14573900,
    ↪  117778179, 803694179079458816, 884378714, 14499829, 153436029, 1072915220,
    ↪ 253508662, 18479513, 344801362, 75974281, 637313893, 38164846, 508932270, 166747718,
    ↪  33933259, 1196254680, 36008570, 348379865, 53853197, 1349149096909668363, 14399483,
    ↪  180463340, 11525172, 140750163, 138372303, 18927441, 789605024, 236636515,
    ↪ 186386857, 1429480158, 355708717, 163559269, 15872418, 1059327333532749824,
    ↪ 403255314, 1227253801, 34713362, 2500955780, 15764644, 9317502, 303862998, 50811932,
    ↪  14091091, 1264208004, 51742969, 274119641, 39344374, 54341363, 266135518, 15820664,
    ↪  368186169, 29840295, 119360445, 50004938, 95955871, 878204370403250176, 66369181,
    ↪ 439125710, 25421094, 38679388, 44627459, 202890266, 268361752, 54951973, 2541363451,
    ↪  9624742, 22203756, 1083198663424237569, 52422878, 14727959, 453507741, 106485105,
    ↪ 352145373, 66083959, 25069929, 181567076, 14224719, 393852070, 1725137533,
    ↪ 124839730, 121402638, 163730859, 156132825, 28412286, 754006735468261376, 131948686,
    ↪  30309979, 15290441, 84358766, 19018401, 65647594, 136479128, 773069256, 15057943,
    ↪ 1459223742, 130834357, 16302242, 86254626, 281766200, 1469065189, 16998020,
    ↪ 85772748, 27909036, 18936161, 16827333, 1016008321, 6519522, 178218286, 127483019,
    ↪ 69008563, 115622213, 92726549, 59808841, 1117317140, 19330043, 44728980, 22733444,
    ↪ 965741665, 2155215398, 15808765, 76348185, 25410548, 185449401, 72568426, 20106852,
    ↪ 171574926, 1173787748, 2803191, 1323730225067339784, 189283341, 85221650, 20255162,
    ↪ 75823831, 7428362, 139162440, 1353020900, 624462023, 8159472, 17681513, 116393932,
    ↪ 138600717, 162825036, 816412233488015360, 25087685, 17006157, 2916305152, 64444249,
    ↪ 70162012, 142295586, 185487499, 259758878, 2533070650, 25093616, 21904217,
    ↪ 203162990, 311775629, 3030158859, 27778893, 24920669, 36117822, 287834630, 17057271,
    ↪  740216334, 36042554, 38232487, 435299501, 27711339, 277444134, 96846955,
    ↪ 1903382054, 2253142422, 22009731, 391037985, 739417207, 2251588934, 39349894,
    ↪ 339040963, 20560294, 2559865245, 104969057, 471010707, 135901326, 23019705,
    ↪ 53556894, 109241992, 2936714848, 13201312, 43513715, 5120691, 14647570, 9161692,
    ↪ 450504440, 70179948, 819543872883884032, 124003770, 2495199115, 24002803, 300114634,
    ↪  152987149, 132355708, 27760317, 94360081, 59553554, 52939106, 71026122, 14857290,
    ↪ 165699676, 2892809359, 146569971, 204151028, 1205974434, 22179997, 2835653131,
    ↪ 87118217, 200163448, 44527648, 73171449, 2163267319, 52420819, 19426729, 22412376,
    ↪ 47033057, 78021221, 15007149, 17936793, 27860681, 158650448, 12804422, 41077287,
    ↪ 49153854, 14372726, 20646711, 14939981, 184004752, 15880163, 29307032, 86962432,
    ↪ 15962096, 54682125, 69231187, 79536438, 6273552, 17629860, 14700117, 78270570,
    ↪ 2359926157, 58913261, 50626996, 46104914, 66857043, 19802879, 27406401, 23970102,
    ↪ 128023009, 183253150, 2317351705, 15573174, 17243213, 240664869, 153403071,
    ↪ 17560096, 988573326376427520, 1430793200, 17939037, 7111412, 18653699, 97537762,
    ↪ 39975533, 278328390, 775887091, 18139619, 16213013, 227944601, 400636503, 166767883,
    ↪  2420020489, 2238689274, 26235265, 15143478, 78335735, 19882775, 125430181,
    ↪ 29501253, 2720868859, 31451582, 265205959, 1198406491, 108247668, 49751816,
    ↪ 90021519, 1296642259, 63662120, 98409905, 33512799, 105838056, 161418822, 158426909,
    ↪  316756361, 368863251, 19784831, 298959564, 14055301, 18164420, 108538006, 79915337,
    ↪  50184013, 24925573, 24275419, 14082692, 376378883, 144622747, 19604744, 16893886,
    ↪ 16581604, 176486829, 113317910, 28208983, 957027824, 2177710981, 11134252, 1440641,
    ↪ 495509842, 1150409755342573569, 73880716, 1675405508, 316577087, 627491854,
    ↪ 321866313, 301309392, 2489647370, 273538402, 38495835, 22736017, 130159846,
    ↪ 19122675, 332638587, 16561884, 309824558, 1628824092, 878247600096509952, 44874400,
    ↪ 126150285, 607281049, 15234407, 1561123663, 19154824, 14361155, 948444535, 16560657,
    ↪  1135806843773542401, 216065430, 54236340, 1728956976, 700568101, 64929124,
    ↪ 69610339, 568545739, 109387470, 1167010648410091520, 338304891, 2457691350,
    ↪ 624896628, 17074440, 56518880, 210938747, 45055696, 38127255, 24019308, 133448051,
```

```
↪ 710064516, 147975612, 1222639789, 2597666894, 14569869, 170424259, 8129122,
↪ 46134530, 8163442, 47319664, 15891449, 21866939, 44312605, 15651878, 17088779,
↪ 1349154719386775552, 28273230, 30602016, 16347506, 58545637, 35749264, 81216996,
↪ 1187330941, 56377143, 1638021792, 128487133, 632110916, 54711948, 19413393,
↪ 2584588824, 124450567, 45122598, 1139001474, 52833129, 85289376, 5768872, 105137891,
↪  45696879, 93957809, 111933542, 592361291, 17959104, 277983398, 24871896, 89014429,
↪ 1244325504, 558572709, 142296675, 42220353, 61452308, 85536744, 18272699,
↪ 2477802067, 568871711, 15647676, 56110169, 1545994664, 86395621, 14840869,
↪ 1562061097, 22799750, 92424604, 768019865777139712, 1903298402, 25489250, 441636351,
↪  2284802828, 3385125854, 33923443, 7563792, 126955629, 18101713, 73459349, 17899109,
↪  45363552, 701625379, 14836197, 1877831, 14377605, 2220938808, 27698202, 32372834,
↪ 173744348, 255427651, 2812768561, 139823781, 252709160, 25482356, 75749375,
↪ 301779034, 17004618, 89366493, 729468343, 19658936, 147227083, 17581754, 3390133696,
↪  15506669, 54730258, 225539878, 614754689, 298966311, 323319539, 2228878592,
↪ 399404516, 117087700, 342165503, 2955596087, 17137628, 16580226, 333800185,
↪ 145992645, 29860364, 342863309, 21870081, 286036879, 262454289, 293932241, 73181712,
↪  16712746, 39797433, 57016932, 116677732, 170719913, 17308994, 841211876776259585,
↪ 167421802, 21744554, 537004797, 3459051, 196610668, 2155649270, 772789437063368704,
↪ 1537083344, 2587658342, 2585481402, 575508800, 247217109, 567483109, 37338814,
↪ 40307393, 114035804, 399909209, 631701837, 18392906, 20321785, 18909186, 2463676838,
↪  44215023, 770120310, 27830610, 28420827, 68838777, 122392884, 32625314, 114721074,
↪ 788759365859835906, 497658491, 40839292, 90964854, 703558015862104064, 25098482,
↪ 120256278, 14575107, 1469402426, 75050496, 820521182, 7846, 50246444, 481156800,
↪ 471677441, 30023434, 34914032, 847238528, 222073627, 20518710, 267305045,
↪ 2282671736, 270176702, 19106719, 567167802, 2384071, 1712867886, 235582329,
↪ 1023096199, 25398075, 95092020, 342637804, 90086871, 46116615, 454873450, 192611378,
↪  347257032, 89556802, 1209342523, 190618660, 19256800, 91882544, 70006177, 86876958,
↪  373410448, 241747585, 90347940, 18055279, 8775672, 719324795167309827, 118739566,
↪ 32469566, 20661539, 46096164, 9594042, 28035260, 16273831, 62553744, 119844526,
↪ 14882900, 500713106, 35758259, 806179924407107585, 220388847, 1206216140, 18954398,
↪ 24691376, 47813521, 14589119, 28442098, 20596281, 546076148, 70903729, 384287774,
↪ 335501334, 20087934, 21323268, 17329538, 33537967, 194522360, 90921204, 20120877,
↪ 368948189, 101895924, 278057040, 207923746, 1033994695595065347, 2719829002,
↪ 6467332, 21059255, 18847632, 3097503906, 142198416, 114423418, 73402979, 5248441,
↪ 4035503243, 701667727, 16832632, 400177598, 279189207, 139930250, 27489253,
↪ 46220856, 429247237, 222516693, 591361197, 34570323, 15225275, 234092394,
↪ 2629502950, 92731878, 20225875, 1334928384, 95151029, 86437426, 823607244, 29856819,
↪  449539004, 72520597, 19537303, 965266179836252161, 140741639, 256136003, 116142776,
↪  81063712, 2389227352, 350224247, 2207520854, 54885400, 377723724, 25355724,
↪ 39455426, 398092540, 41259253, 987383070, 23795109, 3114898232, 67888012,
↪ 4284189436, 16029780, 14791162, 39822897, 10126672, 281890963, 2334193741,
↪ 435525090, 287858728, 67600610, 9109712, 738080573365702657, 2152685876, 28297965,
↪ 459968785, 128216887, 156380350, 435331179, 110783489, 87284779, 21313364,
↪ 871686942, 266025901, 2396018036, 2585657988, 17338082, 177886974, 334430754,
↪ 371797057, 462489802, 240458716, 1170817123, 50115087, 189868631, 167760471,
↪ 23702397, 56561449, 65493023, 20436059, 1349175483989188608, 153000691, 378197959,
↪ 244655353, 13353862, 139611058, 1059325677202157568, 167435593, 91966685,
↪ 4484449334, 2900350908, 441115772, 250495841, 27855118, 43989162, 41605989,
↪ 51809554, 18807671, 2246341620, 145924343, 20437286, 3324347603, 80676273, 93131556,
↪  21390437, 115644179, 43847197, 314120493, 16568227, 26548315, 31444922, 932163222,
↪ 46143892, 2557037323, 15263572, 279449435, 232428933, 21701757, 15134240, 21436663,
↪ 16588111, 43383705, 18667907, 17310915, 34082314, 19611483, 29403387, 28644174,
↪ 18759851, 69588362, 2495214973, 2880932962, 20241784, 56847778, 17292143, 19409270,
↪ 204002663, 1669854782, 95956928, 3025450914, 177656790, 21390674, 819685194,
↪ 41673710, 148529707, 920488039, 14663955, 52584545, 113613920, 159225370, 37344334,
↪ 251706984, 66515223, 37836873, 16576356, 302974524, 87239242, 109312021, 182033179,
↪ 197598287, 208711264, 65668018, 268941928, 28054681, 44256533, 281217033, 135112407,
↪  976781233907740674, 1653237805, 17732153, 287117871, 24423418, 33549858, 5391882,
↪ 95398297, 15092616, 38652395, 248666147, 225265639, 947425424, 842978276800835584,
↪ 29450962, 120965579, 50627903, 15830716, 214924885, 28165335, 7152572, 1124722303,
↪ 188006067, 18760392, 338906109, 364814031, 968652142750810114, 352881953, 59319830,
↪ 16245822, 252663106, 120852219, 630112482, 1349170292564905988, 489387676,
↪ 153264772, 34890810, 30282704, 585372692, 2220638376, 1112465706, 2811559122,
↪ 29701712, 1123232834376933376, 43769311, 38479920, 18576537, 90959399,
↪ 705694814612938752, 127431090, 484303440, 28123862, 3147289455, 16839878,
↪ 770781940341288960, 239819612, 296522382, 31641658, 909102066, 987221076, 23505298,
↪ 825670395714007041, 237348797, 201352564, 276540738, 184964515, 223622368,
↪ 405667557, 277920665, 967000437797761024, 32871086, 278175819, 89421672,
↪ 1169025294642102274, 16374678, 701635518, 875029068223315968, 1045110787, 70740986,
↪ 22917508, 953663868659798017, 23598103, 222022987, 90636188, 23489663, 4091551984,
↪ 926620369, 755953153, 586291040, 15460048, 225125248, 101986616, 14342564,
↪ 356442237, 16043308, 5405152, 473285651, 1568396827, 31428361, 39392910, 147491390]
```

Appendix B Hierarchical Bayesian Inference Results

B.1 Trace Auto-Correlation Plots

Here the auto-correlation plots of the traces from the hierarchical Bayesian model are shown. The traces are the sequence of sampled coefficient values tested. A low auto-correlation (as can be seen in the Figures 50 and 51) indicate that the inference solver of ***PyMC3*** has converged to a stable solution. If the auto-correlation would be high (largely outside of the visualised grey bars) this would indicate that more samples would have to be evaluated before a solution could be obtained.

Figure 50: Trace Auto-Correlation plots of 512 data

Figure 51: Trace Auto-Correlation of 86 400 data

B.2 Scatter Matrix Plots

To further visualise the data Fig. 52 and Fig. 53 show the scatter plots for the market of ETH_USD for the 512 and 86 400 data respectively. The remaining scatter plots for all of the analysed markets showed similar types of distributions per variable and were omitted from this appendix for brevity. An interested reader can however easily regenerate the plots for all markets by making use of the full dataset obtainable through the link in Section 3.5.

The first subplot in the first row shows the distribution of values of the market price. This is the endogenous variable we are trying to estimate. The remaining subplots have the market price on the x-axis and a variable on the y-axis.

It is difficult to interpret the relationships of variables to price from the scatter plots alone, that is why a hierarchical Bayesian analysis method is applied to the data. But at least some sort of clustering appears to exist, which may indicate the variables are not entirely independent from the price.

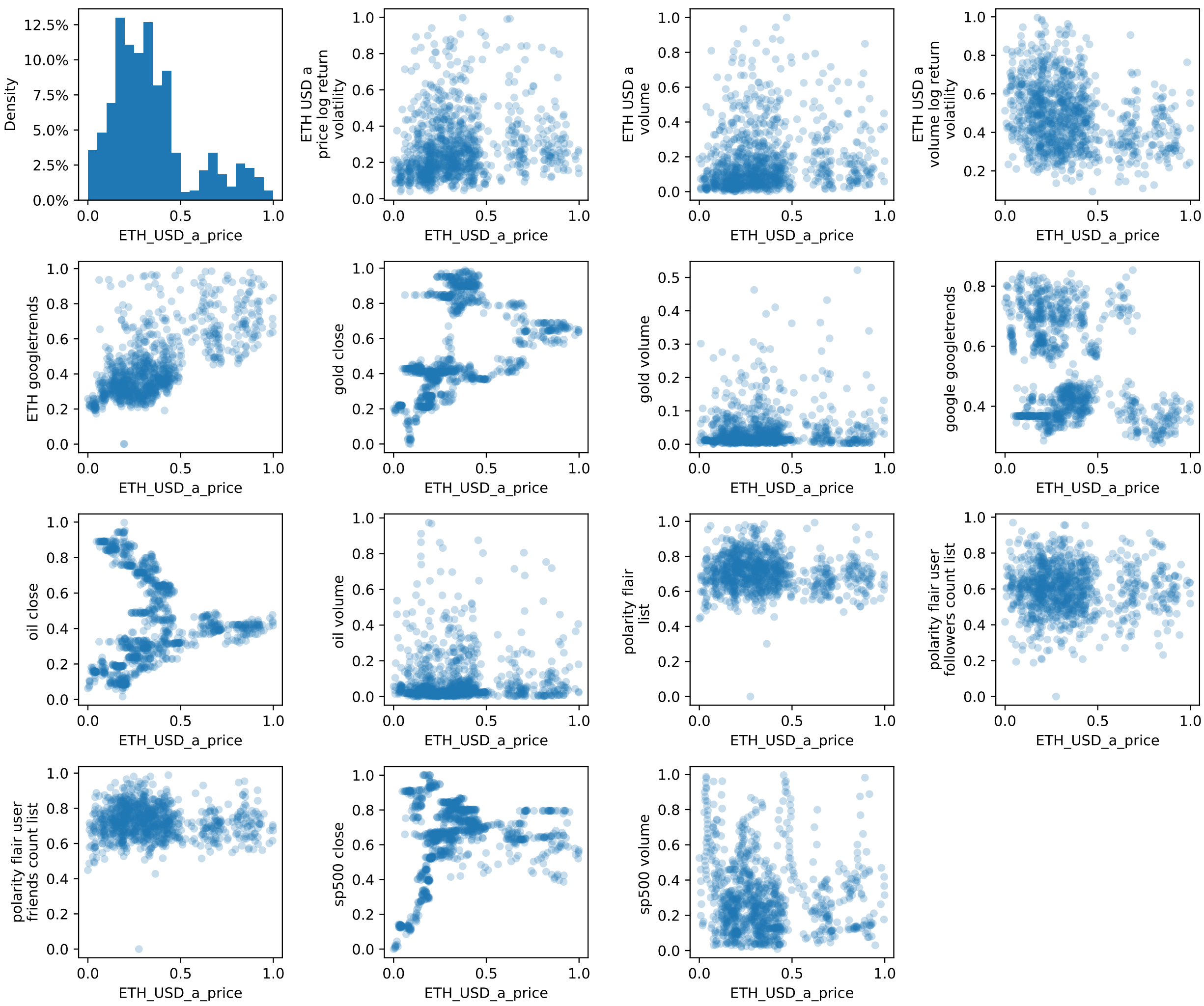

Figure 52: Scatter matrix ETH_USD of 512 data

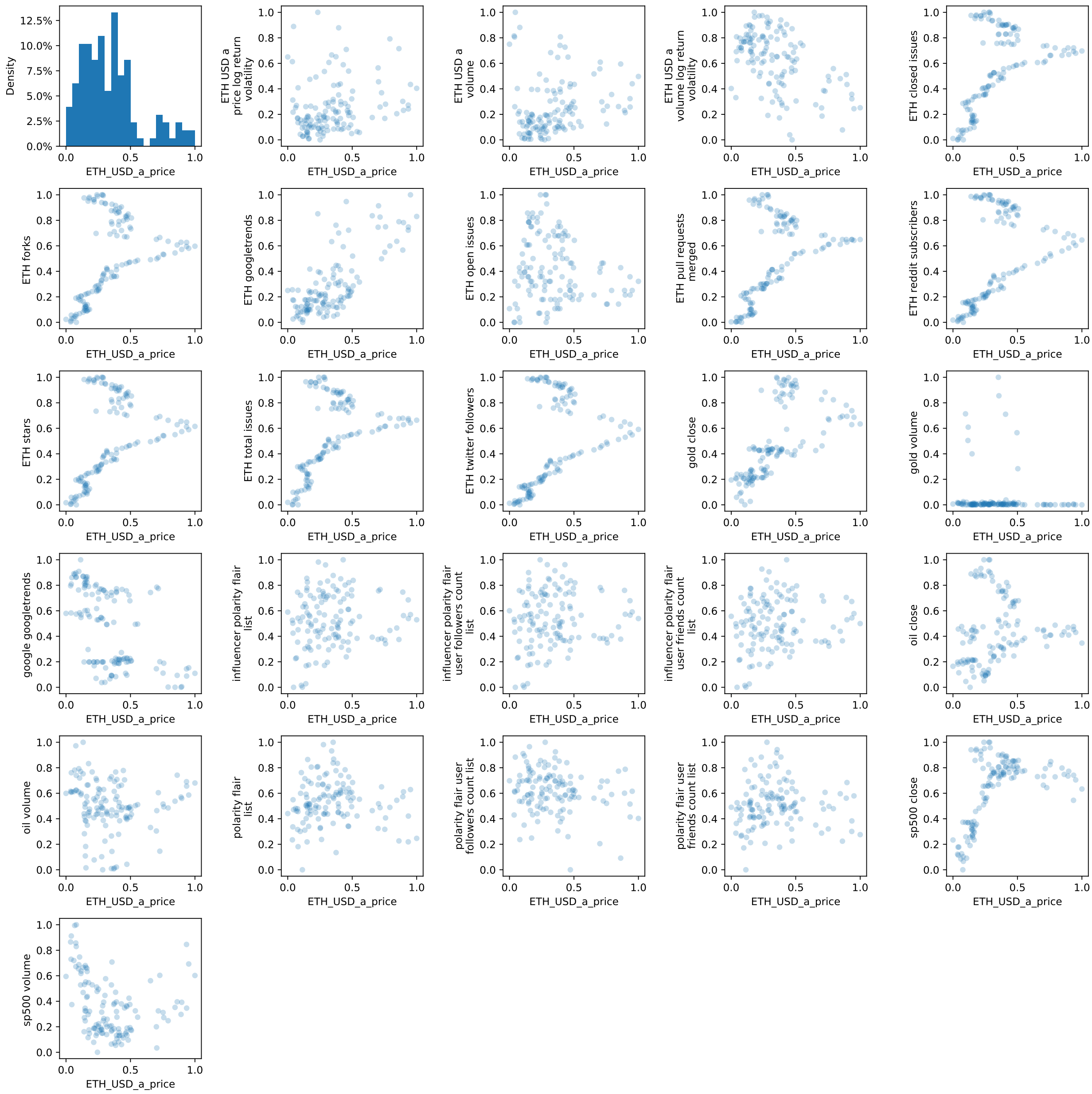

Figure 53: Scatter matrix ETH_USD of 86400 data

B.3 R^2 coefficients without control variable volume

Table 17 shows the R^2 values that are obtained when not including the control variable volume.

Table 17: R^2 values of the bayesian hierarchical model estimate, without control variable volume

	R^2 512 data		R^2 86400 data	
	standard	adjusted	standard	adjusted
ADA_EUR	0.687703	0.687532	0.822014	0.786978
ADA_USD	0.702358	0.702195	0.843145	0.812268
BTC_DAI	0.839685	0.839597	0.937673	0.925404
BTC_EUR	0.859653	0.859577	0.949191	0.939190
BTC_USD	0.847535	0.847451	0.944664	0.933771
BTC_USDC	0.837793	0.837704	0.938211	0.926047
BTC_USDT	0.853124	0.853044	0.943066	0.931858
ETH_EUR	0.611477	0.611265	0.917995	0.901852
ETH_USD	0.618693	0.618484	0.912187	0.894901
ETH_USDC	0.599010	0.598791	0.900359	0.880745
ETH_USDT	0.613668	0.613456	0.917052	0.900724
LTC_EUR	0.637072	0.636874	0.943503	0.932382
LTC_USD	0.629784	0.629582	0.950511	0.940769
LTC_USDT	0.632449	0.632248	0.943929	0.932892
XLM_USD	0.584594	0.584367	0.824923	0.790459
XRP_EUR	0.620057	0.619849	0.911480	0.894055
XRP_USD	0.623333	0.623126	0.907662	0.889485
XRP_USDT	0.625070	0.624864	0.905084	0.886400
mu	0.690170	0.690000	0.911814	0.894454
sigma	0.101207	0.101263	0.040098	0.047991

Appendix C Beta Coefficient Tables

I this section the mean (μ) and standard deviations (σ) for the hyper-prior and beta-coefficient distributions are provided. A larger magnitude of μ may be interpreted as more significant in terms of the variables influence. The magnitude of the ratio $\frac{\sigma}{\mu}$ indicates how large the standard deviation is compared to its mean. When this value is low this indicates a high level of confidence in the effect of this variable. The percentage of samples smaller than 0 out of the number of samples taken is given in the last column. When it has a value of 0 % this means that all values sampled were positive. If it has a value of 100 % this indicates that all values are positive. In the Bayesian inference methodology the null-hypothesis would assume values centred at 0. So variables that have a negative value ratio close to 50 % might be considered insignificant.

Each table is sorted according to significance. Where significance is defined by Eq. (C.1).

$$|\mu| \cdot \left| \frac{n_{negative}}{n} - \left(1 - \frac{n_{negative}}{n}\right) \right| \tag{C.1}$$

Table 18: Cluster A beta a price log return volatility distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_USDC	-0.842586	0.110987	0.131722	1.000000	0.000000
ETH_EUR	-0.444969	0.029840	0.067062	1.000000	0.000000
ETH_USD	-0.444123	0.032940	0.074168	1.000000	0.000000
ETH_USDT	-0.379437	0.034351	0.090531	1.000000	0.000000
BTC_USDT	-0.237434	0.037029	0.155954	1.000000	0.000288
BTC_EUR	-0.216081	0.034527	0.159786	1.000000	0.000288
BTC_USD	-0.210229	0.034961	0.166300	1.000000	0.001151
ADA_EUR	-0.176368	0.031135	0.176532	1.000000	0.006908
ADA_USD	-0.152637	0.029610	0.193991	1.000000	0.038284
LTC_EUR	-0.142305	0.027798	0.195340	1.000000	0.067358
LTC_USD	-0.131375	0.030252	0.230273	1.000000	0.142487
LTC_USDT	-0.119080	0.035850	0.301059	0.999424	0.294185
XRP_EUR	-0.105599	0.029316	0.277613	1.000000	0.433794
XRP_USDT	-0.100260	0.033917	0.338291	0.997697	0.490789
BTC_USDC	0.102811	0.053480	0.520175	0.943581	0.471790
XRP_USD	-0.032578	0.027964	0.858362	0.758204	0.994819
BTC_DAI	0.029075	0.031685	1.089751	0.630397	0.987334
XLM_USD	0.020390	0.031033	1.521968	0.462291	0.996258

Table 19: Cluster A beta a volume distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_USDC	-0.128565	0.030399	0.236448	1.000000	0.172136
XLM_USD	0.088700	0.029050	0.327514	0.997697	0.646805
BTC_USDC	-0.088376	0.028099	0.317946	0.998849	0.658319
XRP_USDT	0.072686	0.030986	0.426294	0.982729	0.803685
XRP_EUR	0.059890	0.029058	0.485187	0.959125	0.912781
ETH_USD	0.052685	0.029344	0.556966	0.928037	0.945596
ETH_USDT	-0.050250	0.028913	0.575376	0.925158	0.954231
LTC_USDT	-0.047624	0.029203	0.613190	0.900403	0.963731
LTC_EUR	0.035766	0.027438	0.767133	0.795049	0.991940
BTC_USDT	0.034554	0.029122	0.842791	0.767991	0.985895
BTC_USD	0.030996	0.029389	0.948158	0.703512	0.990213
BTC_EUR	0.030327	0.028723	0.947124	0.698330	0.991940
ADA_USD	-0.025433	0.027591	1.084829	0.640760	0.995394
LTC_USD	-0.025014	0.028437	1.136847	0.623489	0.994819
XRP_USD	-0.020569	0.030785	1.496718	0.496834	0.994531
ADA_EUR	0.013649	0.028858	2.114268	0.352907	0.998849
BTC_DAI	-0.012806	0.032837	2.564122	0.301094	0.994243
ETH_EUR	0.010555	0.028095	2.661915	0.295337	0.998561

Table 20: Cluster A beta a volume log return volatility distribution mean and significance 512 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XLM_USD	0.188838	0.028617	0.151546	1.000000	0.000576
BTC_USD	-0.167441	0.027798	0.166018	1.000000	0.006621
BTC_USDT	-0.165605	0.026483	0.159919	1.000000	0.006045
LTC_USD	0.070928	0.025652	0.361661	0.992516	0.870466
XRP_USDT	0.062565	0.024500	0.391602	0.990213	0.938975
BTC_EUR	-0.061428	0.024097	0.392274	0.990213	0.943581
LTC_EUR	0.060484	0.024933	0.412221	0.980426	0.940702
LTC_USDT	0.045612	0.021414	0.469478	0.971790	0.995107
ETH_USDC	0.041986	0.020776	0.494840	0.961428	0.996546
ADA_USD	-0.035952	0.027252	0.758016	0.804836	0.989925
ETH_USD	0.031391	0.027642	0.880558	0.750144	0.995682
ADA_EUR	0.025476	0.024654	0.967713	0.701209	0.998561
XRP_EUR	-0.025081	0.024746	0.986638	0.681059	0.998561
ETH_USDT	0.022758	0.025511	1.120978	0.625792	0.998273
BTC_DAI	0.016310	0.020070	1.230543	0.589522	1.000000
ETH_EUR	-0.016321	0.026348	1.614352	0.464594	0.999136
XRP_USD	0.006047	0.025019	4.137742	0.211860	1.000000
BTC_USDC	-0.001867	0.020245	10.842420	0.058722	1.000000

Table 21: Cluster A beta gold close distribution mean and significance 512 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ADA_USD	0.527565	0.025586	0.048499	1.000000	0.000000
ADA_EUR	0.492376	0.025060	0.050895	1.000000	0.000000
BTC_EUR	-0.420880	0.024768	0.058848	1.000000	0.000000
BTC_USDC	-0.388210	0.024719	0.063675	1.000000	0.000000
BTC_USD	-0.379652	0.024745	0.065179	1.000000	0.000000
BTC_USDT	-0.353596	0.025129	0.071066	1.000000	0.000000
BTC_DAI	-0.290651	0.024789	0.085287	1.000000	0.000000
ETH_USD	0.153052	0.023438	0.153139	1.000000	0.013529
ETH_USDT	0.147356	0.025130	0.170536	1.000000	0.029937
XRP_EUR	-0.146289	0.023354	0.159641	1.000000	0.022165
XRP_USDT	-0.132900	0.024809	0.186675	1.000000	0.094704
XRP_USD	-0.131833	0.024093	0.182756	1.000000	0.091537
ETH_EUR	0.130000	0.024062	0.185090	1.000000	0.103627
ETH_USDC	0.116914	0.023861	0.204090	1.000000	0.235751
XLM_USD	-0.086739	0.024671	0.284421	1.000000	0.707830
LTC_EUR	-0.085541	0.024141	0.282217	1.000000	0.726252
LTC_USD	-0.064395	0.024545	0.381160	0.994243	0.924870
LTC_USDT	-0.063305	0.024372	0.384996	0.990213	0.933218

Table 22: Cluster A beta gold volume distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
LTC_USDT	0.513071	0.064911	0.126515	1.000000	0.000000
LTC_USD	0.500071	0.063232	0.126446	1.000000	0.000000
LTC_EUR	0.497135	0.064026	0.128790	1.000000	0.000000
ADA_EUR	0.465603	0.063377	0.136117	1.000000	0.000000
ETH_EUR	0.457411	0.065582	0.143377	1.000000	0.000000
ADA_USD	0.456260	0.063282	0.138697	1.000000	0.000000
ETH_USD	0.450752	0.062884	0.139510	1.000000	0.000000
ETH_USDT	0.445855	0.065051	0.145903	1.000000	0.000000
XLM_USD	0.438538	0.063493	0.144782	1.000000	0.000000
ETH_USDC	0.424432	0.061460	0.144806	1.000000	0.000000
XRP_USD	0.370547	0.062767	0.169390	1.000000	0.000000
XRP_USDT	0.368402	0.061672	0.167405	1.000000	0.000000
XRP_EUR	0.363840	0.062464	0.171681	1.000000	0.000000
BTC_USD	0.295989	0.065237	0.220403	1.000000	0.002303
BTC_USDC	0.292603	0.064715	0.221170	1.000000	0.001439
BTC_EUR	0.288231	0.064818	0.224883	1.000000	0.001439
BTC_USDT	0.268268	0.067285	0.250815	1.000000	0.006908
BTC_DAI	0.232178	0.068336	0.294325	0.999424	0.030800

Table 23: Cluster A beta google googletrends distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ADA_EUR	-0.564954	0.043741	0.077424	1.000000	0.000000
ADA_USD	-0.544849	0.044404	0.081498	1.000000	0.000000
LTC_USDT	-0.501012	0.043326	0.086476	1.000000	0.000000
LTC_USD	-0.491728	0.043790	0.089052	1.000000	0.000000
LTC_EUR	-0.490301	0.043134	0.087975	1.000000	0.000000
XLM_USD	-0.423774	0.044686	0.105447	1.000000	0.000000
ETH_EUR	-0.345729	0.046605	0.134802	1.000000	0.000000
ETH_USD	-0.338028	0.045146	0.133558	1.000000	0.000000
ETH_USDT	-0.312229	0.045400	0.145406	1.000000	0.000000
ETH_USDC	-0.283609	0.046418	0.163668	1.000000	0.000000
XRP_USDT	-0.053253	0.043826	0.822970	0.773172	0.856074
XRP_EUR	-0.051589	0.042803	0.829693	0.762234	0.867876
XRP_USD	-0.045819	0.042830	0.934778	0.725964	0.896373
BTC_EUR	-0.040352	0.044709	1.107970	0.652850	0.910478
BTC_USD	-0.031366	0.044561	1.420680	0.507196	0.934657
BTC_USDC	0.025292	0.045023	1.780127	0.446172	0.949050
BTC_USDT	0.004342	0.045165	10.401294	0.085204	0.972654
BTC_DAI	0.000403	0.042717	105.870818	0.004030	0.980138

Table 24: Cluster A beta oil close distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_USDC	-1.090188	0.032924	0.030200	1.000000	0.000000
BTC_EUR	-1.039292	0.033455	0.032190	1.000000	0.000000
BTC_USD	-1.039056	0.033963	0.032687	1.000000	0.000000
BTC_USDT	-1.026031	0.033664	0.032809	1.000000	0.000000
XLM_USD	-0.885546	0.034985	0.039506	1.000000	0.000000
BTC_DAI	-0.808400	0.033029	0.040857	1.000000	0.000000
XRP_EUR	-0.807631	0.034228	0.042381	1.000000	0.000000
XRP_USD	-0.806483	0.036019	0.044662	1.000000	0.000000
XRP_USDT	-0.799843	0.035181	0.043985	1.000000	0.000000
LTC_USDT	-0.354411	0.039863	0.112477	1.000000	0.000000
LTC_USD	-0.348239	0.040940	0.117562	1.000000	0.000000
LTC_EUR	-0.337267	0.039257	0.116398	1.000000	0.000000
ETH_USD	-0.300688	0.033635	0.111859	1.000000	0.000000
ETH_USDT	-0.289111	0.032947	0.113959	1.000000	0.000000
ETH_EUR	-0.276520	0.034846	0.126017	1.000000	0.000000
ETH_USDC	-0.251206	0.034344	0.136717	1.000000	0.000000
ADA_EUR	0.121328	0.036785	0.303188	0.999424	0.287277
ADA_USD	0.098320	0.036574	0.371989	0.995394	0.522165

Table 25: Cluster A beta oil volume distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
XRP_EUR	-0.041133	0.025754	0.626114	0.900403	0.984744
XRP_USDT	-0.040601	0.025189	0.620402	0.901554	0.985032
XRP_USD	-0.041006	0.025978	0.633517	0.887737	0.986759
ETH_EUR	0.038327	0.023822	0.621558	0.892919	0.993667
ADA_USD	0.037569	0.023965	0.637904	0.894070	0.993667
ETH_USDT	0.036288	0.024851	0.684824	0.874496	0.991077
ETH_USD	0.036200	0.023932	0.661112	0.865285	0.995970
LTC_USD	0.033659	0.024106	0.716175	0.853771	0.994243
LTC_USDT	0.033426	0.023840	0.713198	0.847438	0.997697
ADA_EUR	0.031061	0.023601	0.759806	0.822107	0.997121
LTC_EUR	0.029309	0.024003	0.818953	0.781232	0.997409
ETH_USDC	0.026966	0.022900	0.849193	0.770294	0.998273
BTC_USD	-0.009897	0.023595	2.384063	0.316062	1.000000
BTC_EUR	-0.007828	0.023818	3.042795	0.249280	1.000000
BTC_DAI	-0.006832	0.023301	3.410769	0.216465	1.000000
BTC_USDT	-0.005156	0.023958	4.646512	0.142775	0.999136
BTC_USDC	-0.004618	0.023699	5.131425	0.149108	1.000000
XLM_USD	-0.004196	0.023382	5.572821	0.130685	1.000000

Table 26: Cluster A beta polarity flair list distribution mean and significance 512 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XLM_USD	-0.507188	0.050481	0.099531	1.0	0.000000
BTC_USDC	-0.491832	0.049208	0.100051	1.0	0.000000
BTC_EUR	-0.476451	0.051115	0.107282	1.0	0.000000
BTC_USDT	-0.460856	0.049903	0.108283	1.0	0.000000
XRP_USDT	-0.451596	0.048943	0.108378	1.0	0.000000
XRP_USD	-0.450808	0.050949	0.113017	1.0	0.000000
XRP_EUR	-0.448400	0.049320	0.109991	1.0	0.000000
BTC_USD	-0.447964	0.049379	0.110230	1.0	0.000000
BTC_DAI	-0.413347	0.048374	0.117030	1.0	0.000000
ETH_USDC	-0.336761	0.049249	0.146242	1.0	0.000288
ETH_USDT	-0.328109	0.049319	0.150314	1.0	0.000000
ETH_USD	-0.323928	0.050146	0.154805	1.0	0.000000
ETH_EUR	-0.322948	0.050491	0.156344	1.0	0.000000
LTC_USDT	-0.308668	0.050640	0.164060	1.0	0.000000
LTC_USD	-0.307037	0.050204	0.163510	1.0	0.000000
LTC_EUR	-0.304562	0.048860	0.160429	1.0	0.000000
ADA_USD	-0.275011	0.053157	0.193292	1.0	0.002303
ADA_EUR	-0.270484	0.052715	0.194890	1.0	0.002015

Table 27: Cluster A beta polarity flair user followers count list distribution mean and significance 512 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ETH_USDT	0.049237	0.017007	0.345414	0.999424	0.987047
ETH_USDC	0.047720	0.016281	0.341182	0.998273	0.992228
ETH_EUR	0.047320	0.015767	0.333192	0.997697	0.995970
ETH_USD	0.046970	0.015530	0.330649	0.997697	0.995107
LTC_EUR	0.043021	0.014647	0.340463	0.991940	0.997985
BTC_USDC	0.042755	0.015109	0.353397	0.986183	0.998849
BTC_USDT	0.042467	0.014567	0.343025	0.989062	0.999136
BTC_EUR	0.042360	0.014694	0.346889	0.986183	0.998561
LTC_USD	0.041964	0.014449	0.344319	0.990213	0.998849
LTC_USDT	0.041779	0.014838	0.355161	0.987910	0.997985
BTC_DAI	0.040948	0.014454	0.352991	0.987334	0.999712
BTC_USD	0.039995	0.014746	0.368704	0.982153	0.999424
XRP_USD	0.037935	0.015329	0.404087	0.962579	1.000000
XRP_USDT	0.037677	0.015135	0.401692	0.966609	0.999424
XLM_USD	0.037541	0.015030	0.400375	0.963731	0.999712
XRP_EUR	0.037292	0.015516	0.416082	0.960276	1.000000
ADA_USD	0.032751	0.017338	0.529390	0.894646	1.000000
ADA_EUR	0.031649	0.017083	0.539756	0.890616	1.000000

Table 28: Cluster A beta polarity flair user friends count list distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_USDT	0.165545	0.036122	0.218199	1.000000	0.023316
LTC_EUR	0.165445	0.034719	0.209854	1.000000	0.027634
BTC_USDT	0.163423	0.035851	0.219376	1.000000	0.033679
LTC_USDT	0.163109	0.036394	0.223128	1.000000	0.037133
ETH_USDC	0.162888	0.035001	0.214876	1.000000	0.035406
ETH_EUR	0.162628	0.035954	0.221079	0.999424	0.036269
BTC_EUR	0.162428	0.034670	0.213445	1.000000	0.033967
BTC_USDC	0.162195	0.035467	0.218671	1.000000	0.031952
ETH_USD	0.162124	0.035551	0.219283	1.000000	0.034542
LTC_USD	0.161965	0.035396	0.218542	0.999424	0.037997
BTC_USD	0.161403	0.035288	0.218630	1.000000	0.041739
BTC_DAI	0.161035	0.035697	0.221674	0.999424	0.037709
XRP_EUR	0.153215	0.035715	0.233102	0.998849	0.070812
XRP_USD	0.152448	0.036937	0.242291	0.997121	0.073690
XLM_USD	0.151941	0.036114	0.237684	0.999424	0.080599
XRP_USDT	0.151501	0.035827	0.236483	0.997697	0.075130
ADA_EUR	0.145556	0.039252	0.269667	0.992516	0.110823
ADA_USD	0.145396	0.039669	0.272835	0.990789	0.105066

Table 29: Cluster A beta quote googletrends distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_EUR	0.828131	0.034194	0.041290	1.0	0.000000
ETH_USDT	0.820351	0.031803	0.038768	1.0	0.000000
ETH_USD	0.805924	0.032440	0.040252	1.0	0.000000
LTC_USD	0.728338	0.034175	0.046922	1.0	0.000000
LTC_USDT	0.724682	0.033323	0.045983	1.0	0.000000
LTC_EUR	0.724682	0.032206	0.044441	1.0	0.000000
ETH_USDC	0.683332	0.029764	0.043557	1.0	0.000000
ADA_EUR	0.538592	0.032737	0.060783	1.0	0.000000
ADA_USD	0.528002	0.033197	0.062873	1.0	0.000000
XRP_USD	0.494598	0.023401	0.047314	1.0	0.000000
XRP_USDT	0.488888	0.023302	0.047663	1.0	0.000000
XRP_EUR	0.484079	0.023895	0.049361	1.0	0.000000
BTC_USDC	-0.245466	0.025548	0.104081	1.0	0.000000
XLM_USD	-0.212709	0.047986	0.225596	1.0	0.010363
BTC_USD	-0.197642	0.027049	0.136858	1.0	0.000576
BTC_DAI	-0.192836	0.024860	0.128917	1.0	0.000000
BTC_USDT	-0.186480	0.028047	0.150403	1.0	0.000576
BTC_EUR	-0.174973	0.029281	0.167346	1.0	0.007484

Table 30: Cluster A beta sp500 close distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
XRP_USD	0.799968	0.038761	0.048453	1.000000	0.000000
XRP_USDT	0.799085	0.037958	0.047501	1.000000	0.000000
XRP_EUR	0.795094	0.038684	0.048653	1.000000	0.000000
XLM_USD	0.483164	0.035614	0.073711	1.000000	0.000000
BTC_USDC	0.392412	0.035004	0.089203	1.000000	0.000000
BTC_USD	0.366458	0.035414	0.096638	1.000000	0.000000
BTC_USDT	0.365566	0.035432	0.096924	1.000000	0.000000
BTC_EUR	0.316060	0.035662	0.112835	1.000000	0.000000
BTC_DAI	0.293706	0.034787	0.118440	1.000000	0.000000
ETH_USD	0.270362	0.034934	0.129213	1.000000	0.000000
ETH_USDT	0.262791	0.034644	0.131831	1.000000	0.000000
ETH_EUR	0.256229	0.034886	0.136152	1.000000	0.000000
ETH_USDC	0.225040	0.035432	0.157449	1.000000	0.000000
LTC_USDT	0.191651	0.038411	0.200419	1.000000	0.010363
LTC_USD	0.187191	0.039029	0.208499	1.000000	0.012090
LTC_EUR	0.174326	0.037815	0.216920	1.000000	0.023892
ADA_EUR	-0.090646	0.034954	0.385612	0.987910	0.607369
ADA_USD	-0.061404	0.034757	0.566031	0.922855	0.867876

Table 31: Cluster A beta sp500 volume distribution mean and significance 512 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_USDC	-0.038567	0.013781	0.357324	0.998273	1.0
XLM_USD	-0.036573	0.013129	0.358987	0.997697	1.0
BTC_EUR	-0.035719	0.012404	0.347274	0.998273	1.0
XRP_USD	-0.034940	0.012658	0.362293	0.994243	1.0
XRP_EUR	-0.033787	0.012406	0.367195	0.994243	1.0
XRP_USDT	-0.033756	0.012731	0.377157	0.993092	1.0
BTC_DAI	-0.032931	0.012229	0.371350	0.993092	1.0
BTC_USD	-0.032090	0.012053	0.375586	0.987910	1.0
ADA_EUR	-0.030844	0.011882	0.385214	0.984456	1.0
ADA_USD	-0.030249	0.012126	0.400867	0.978123	1.0
BTC_USDT	-0.029443	0.011618	0.394583	0.978699	1.0
ETH_USDC	-0.025975	0.012219	0.470406	0.937824	1.0
ETH_USDT	-0.024478	0.012438	0.508131	0.924007	1.0
ETH_USD	-0.023665	0.012498	0.528116	0.907887	1.0
ETH_EUR	-0.021951	0.013031	0.593631	0.873345	1.0
LTC_USDT	-0.019200	0.013917	0.724854	0.793898	1.0
LTC_USD	-0.018890	0.014136	0.748336	0.793898	1.0
LTC_EUR	-0.018669	0.014052	0.752690	0.789292	1.0

Table 32: Cluster A intercept const distribution mean and significance 512 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
BTC_EUR	1.325993	0.050437	0.038037	1.0	0.000000
BTC_USD	1.314693	0.047981	0.036496	1.0	0.000000
BTC_USDT	1.274829	0.046944	0.036824	1.0	0.000000
BTC_USDC	1.202280	0.047998	0.039923	1.0	0.000000
XLM_USD	0.968560	0.049929	0.051550	1.0	0.000000
BTC_DAI	0.890536	0.045549	0.051148	1.0	0.000000
LTC_EUR	0.529595	0.050559	0.095467	1.0	0.000000
LTC_USDT	0.525909	0.051486	0.097899	1.0	0.000000
LTC_USD	0.508859	0.050172	0.098596	1.0	0.000000
XRP_EUR	0.384502	0.047229	0.122833	1.0	0.000000
ADA_EUR	0.359152	0.054726	0.152374	1.0	0.000000
ADA_USD	0.349263	0.054926	0.157263	1.0	0.000000
XRP_USD	0.348666	0.048061	0.137842	1.0	0.000000
XRP_USDT	0.333876	0.048087	0.144027	1.0	0.000000
ETH_USDC	0.279496	0.052156	0.186607	1.0	0.000000
ETH_EUR	0.249410	0.056624	0.227033	1.0	0.005469
ETH_USD	0.213101	0.055123	0.258669	1.0	0.019574
ETH_USDT	0.202421	0.053391	0.263763	1.0	0.027058

Table 33: Cluster B beta a price log return volatility distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
BTC_USDT	-0.341608	0.083217	0.243605	1.000000	0.001545
ETH_EUR	-0.335059	0.076172	0.227340	1.000000	0.000515
BTC_EUR	-0.311845	0.072334	0.231955	1.000000	0.002060
ADA_EUR	-0.308884	0.078277	0.253420	1.000000	0.002403
LTC_USD	-0.280453	0.069572	0.248070	1.000000	0.004978
ETH_USD	-0.259120	0.071092	0.274361	1.000000	0.012187
ETH_USDT	-0.248393	0.068762	0.276827	0.999657	0.015448
BTC_USD	-0.220510	0.067156	0.304550	0.998970	0.037075
LTC_EUR	-0.217744	0.070388	0.323262	0.996910	0.046001
ADA_USD	-0.197265	0.058332	0.295703	0.998627	0.049262
LTC_USDT	-0.139632	0.062452	0.447260	0.976313	0.268623
ETH_USDC	-0.138242	0.058211	0.421080	0.983179	0.254034
BTC_USDC	-0.048741	0.060928	1.250037	0.583934	0.800378
XLM_USD	-0.039717	0.054727	1.377941	0.535530	0.859938
XRP_USD	-0.037065	0.067021	1.808172	0.426365	0.802781
XRP_EUR	-0.030956	0.062175	2.008478	0.392722	0.853587
BTC_DAI	0.001409	0.047578	33.777862	0.028150	0.964641
XRP_USDT	0.000373	0.061905	165.793652	0.001716	0.890319

Table 34: Cluster B beta a volume distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XRP_USD	0.254123	0.079424	0.312543	0.998970	0.022657
XRP_EUR	0.215047	0.064718	0.300946	0.998627	0.034672
XRP_USDT	0.170927	0.063398	0.370906	0.993821	0.134741
ETH_USDC	-0.168848	0.050975	0.301900	1.000000	0.090113
ADA_EUR	0.150528	0.081874	0.543911	0.945074	0.276519
BTC_USDT	0.147777	0.075586	0.511483	0.951253	0.254549
LTC_USDT	-0.131459	0.061160	0.465238	0.972537	0.301751
BTC_USDC	-0.125547	0.048475	0.386109	0.992104	0.300034
XLM_USD	0.119471	0.068335	0.571981	0.915894	0.387058
ADA_USD	-0.118567	0.092006	0.775985	0.815311	0.427223
BTC_EUR	0.090273	0.072051	0.798148	0.794027	0.552523
BTC_USD	0.082490	0.075692	0.917591	0.725369	0.585651
ETH_USD	0.065726	0.067400	1.025473	0.669413	0.683831
ETH_EUR	0.054850	0.074354	1.355583	0.534844	0.714384
LTC_USD	0.050020	0.071949	1.438404	0.516993	0.735324
ETH_USDT	-0.046188	0.076092	1.647439	0.461723	0.726227
BTC_DAI	-0.027679	0.044254	1.598867	0.453828	0.944902
LTC_EUR	0.024211	0.068016	2.809252	0.282870	0.837796

Table 35: Cluster B beta a volume log return volatility distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XLM_USD	0.323985	0.057829	0.178492	1.000000	0.000000
ADA_USD	-0.259594	0.069995	0.269632	0.999657	0.010127
LTC_USD	0.104092	0.043924	0.421971	0.983522	0.470649
XRP_EUR	-0.095973	0.040192	0.418787	0.983522	0.538792
LTC_EUR	0.091207	0.053045	0.581593	0.918297	0.570374
BTC_USDC	0.079974	0.047981	0.599960	0.905596	0.662890
ETH_USDC	0.073492	0.053236	0.724375	0.832475	0.693615
LTC_USDT	0.074153	0.055260	0.745222	0.824923	0.677309
BTC_USD	-0.080197	0.067376	0.840135	0.760384	0.617576
BTC_DAI	-0.062882	0.065538	1.042239	0.661174	0.704943
ETH_USD	0.059764	0.072319	1.210072	0.604531	0.699622
XRP_USD	0.048257	0.045849	0.950089	0.706488	0.872297
XRP_USDT	-0.030920	0.056541	1.828634	0.418812	0.880364
ETH_USDT	0.024663	0.051733	2.097652	0.364229	0.917954
ETH_EUR	-0.023931	0.070647	2.952126	0.271198	0.826296
BTC_EUR	0.016003	0.065251	4.077480	0.188465	0.861655
ADA_EUR	0.006674	0.045252	6.780632	0.108823	0.971850
BTC_USDT	0.003224	0.058301	18.083924	0.034672	0.915208

Table 36: Cluster B beta gold close distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ADA_USD	0.664058	0.052435	0.078961	1.000000	0.000000
ADA_EUR	0.596453	0.056636	0.094955	1.000000	0.000000
XLM_USD	0.286413	0.092744	0.323811	0.998627	0.019224
XRP_USDT	0.269140	0.057329	0.213008	1.000000	0.002231
XRP_USD	0.257803	0.053855	0.208901	1.000000	0.001716
XRP_EUR	0.247673	0.054012	0.218078	1.000000	0.003605
ETH_USDC	0.139427	0.082600	0.592428	0.905939	0.310676
ETH_USD	0.129844	0.083301	0.641548	0.878819	0.355304
ETH_USDT	0.118439	0.082924	0.700137	0.856162	0.418641
ETH_EUR	0.103994	0.081421	0.782935	0.794027	0.470992
BTC_EUR	-0.049733	0.056049	1.127014	0.634741	0.812564
BTC_USD	-0.031177	0.057555	1.846074	0.407140	0.872640
LTC_USDT	0.027163	0.057762	2.126516	0.374871	0.886543
BTC_USDT	-0.022315	0.055330	2.479513	0.301751	0.913148
BTC_DAI	-0.012620	0.054513	4.319677	0.178853	0.927738
BTC_USDC	-0.012999	0.056419	4.340196	0.166838	0.911432
LTC_USD	0.009063	0.060915	6.720911	0.106076	0.890663
LTC_EUR	-0.002559	0.061294	23.948211	0.031239	0.898043

Table 37: Cluster B beta gold volume distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ETH_USDC	-0.164929	0.027512	0.166814	1.000000	0.001373
ETH_USDT	-0.161812	0.025722	0.158962	1.000000	0.003433
ETH_USD	-0.161390	0.025292	0.156714	1.000000	0.002060
XLM_USD	-0.161182	0.025339	0.157205	1.000000	0.003605
ETH_EUR	-0.158231	0.024555	0.155181	1.000000	0.004119
BTC_USDC	-0.148306	0.023340	0.157380	1.000000	0.026777
BTC_USD	-0.147343	0.022992	0.156046	1.000000	0.026777
BTC_DAI	-0.147246	0.023095	0.156848	1.000000	0.025232
BTC_USDT	-0.147192	0.023506	0.159695	1.000000	0.032269
BTC_EUR	-0.146590	0.023530	0.160512	1.000000	0.031926
LTC_USDT	-0.145298	0.022879	0.157466	1.000000	0.033642
ADA_USD	-0.145264	0.023206	0.159750	1.000000	0.034501
LTC_USD	-0.142769	0.023439	0.164173	1.000000	0.043769
XRP_USD	-0.141743	0.023801	0.167916	1.000000	0.054240
XRP_EUR	-0.140009	0.024099	0.172125	1.000000	0.058187
LTC_EUR	-0.140006	0.023473	0.167659	1.000000	0.058359
XRP_USDT	-0.139751	0.024010	0.171805	1.000000	0.068143
ADA_EUR	-0.136101	0.025077	0.184250	0.999657	0.089598

Table 38: Cluster B beta google googletrends distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ADA_EUR	-0.246551	0.053857	0.218440	1.000000	0.001373
ADA_USD	-0.203331	0.050903	0.250345	1.000000	0.015791
ETH_USD	-0.125805	0.047245	0.375544	0.991761	0.290422
BTC_USD	-0.125277	0.046065	0.367705	0.992791	0.291624
ETH_USDC	-0.118210	0.046933	0.397034	0.989701	0.354274
ETH_USDT	-0.115302	0.046837	0.406213	0.986268	0.371953
ETH_EUR	-0.114070	0.047456	0.416027	0.983179	0.380536
BTC_USDC	-0.093725	0.044298	0.472632	0.961208	0.557158
BTC_USDT	-0.088269	0.044876	0.508407	0.950566	0.611397
BTC_DAI	-0.086823	0.044860	0.516688	0.940954	0.610882
LTC_USD	-0.089192	0.050121	0.561949	0.915894	0.580330
BTC_EUR	-0.083809	0.045087	0.537969	0.933059	0.636801
LTC_EUR	-0.077950	0.050206	0.644083	0.872640	0.669928
XLM_USD	-0.068080	0.051968	0.763347	0.802609	0.730003
LTC_USDT	-0.061366	0.049484	0.806375	0.777892	0.779609
XRP_EUR	-0.039058	0.048048	1.230145	0.594233	0.899588
XRP_USDT	-0.022405	0.047978	2.141393	0.374528	0.943529
XRP_USD	-0.022139	0.049245	2.224324	0.365946	0.941126

Table 39: Cluster B beta influencer polarity flair list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
LTC_USDT	-0.695713	0.121436	0.174549	1.0	0.0
LTC_EUR	-0.693696	0.121324	0.174895	1.0	0.0
LTC_USD	-0.693412	0.121347	0.175001	1.0	0.0
ETH_USDC	-0.693036	0.121181	0.174856	1.0	0.0
ETH_USD	-0.692721	0.121116	0.174841	1.0	0.0
ETH_EUR	-0.692107	0.120783	0.174515	1.0	0.0
ETH_USDT	-0.691624	0.120981	0.174923	1.0	0.0
ADA_EUR	-0.690931	0.120954	0.175059	1.0	0.0
ADA_USD	-0.690396	0.121066	0.175357	1.0	0.0
BTC_USDT	-0.687548	0.121079	0.176103	1.0	0.0
BTC_EUR	-0.687469	0.121210	0.176313	1.0	0.0
XRP_EUR	-0.687388	0.121444	0.176675	1.0	0.0
XLM_USD	-0.687013	0.121184	0.176392	1.0	0.0
BTC_USDC	-0.686652	0.121151	0.176437	1.0	0.0
XRP_USDT	-0.686500	0.120984	0.176234	1.0	0.0
XRP_USD	-0.686086	0.121481	0.177064	1.0	0.0
BTC_DAI	-0.685891	0.121101	0.176560	1.0	0.0
BTC_USD	-0.685381	0.121177	0.176802	1.0	0.0

Table 40: Cluster B beta influencer polarity flair user followers count list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
BTC_USD	0.414048	0.084268	0.203521	1.0	0.000172
XRP_USDT	0.413713	0.084135	0.203366	1.0	0.000000
XRP_USD	0.413558	0.083852	0.202757	1.0	0.000000
XLM_USD	0.412947	0.084038	0.203508	1.0	0.000172
BTC_DAI	0.412851	0.083836	0.203066	1.0	0.000000
XRP_EUR	0.412416	0.083702	0.202955	1.0	0.000343
BTC_USDC	0.412393	0.084199	0.204171	1.0	0.000000
BTC_EUR	0.411897	0.084126	0.204240	1.0	0.000000
BTC_USDT	0.411021	0.084105	0.204624	1.0	0.000000
ADA_USD	0.409261	0.084108	0.205511	1.0	0.000000
ADA_EUR	0.407446	0.084141	0.206509	1.0	0.000172
ETH_USDT	0.407404	0.083859	0.205838	1.0	0.000172
ETH_EUR	0.406669	0.084311	0.207322	1.0	0.000000
ETH_USD	0.406206	0.084141	0.207138	1.0	0.000172
LTC_USD	0.406016	0.084376	0.207815	1.0	0.000172
LTC_EUR	0.405623	0.084332	0.207908	1.0	0.000172
ETH_USDC	0.405583	0.084131	0.207433	1.0	0.000343
LTC_USDT	0.404133	0.084343	0.208701	1.0	0.000172

Table 41: Cluster B beta influencer polarity flair user friends count list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XLM_USD	0.355220	0.074287	0.209128	1.0	0.000343
BTC_USD	0.354626	0.074239	0.209343	1.0	0.000172
XRP_USDT	0.354428	0.074293	0.209615	1.0	0.000515
XRP_USD	0.354372	0.074567	0.210419	1.0	0.000343
BTC_DAI	0.354274	0.074510	0.210317	1.0	0.000343
BTC_USDC	0.353672	0.074636	0.211032	1.0	0.000343
XRP_EUR	0.353437	0.074713	0.211391	1.0	0.000515
BTC_EUR	0.352776	0.073956	0.209641	1.0	0.000343
BTC_USDT	0.352676	0.074403	0.210967	1.0	0.000343
ADA_USD	0.351056	0.074106	0.211094	1.0	0.000343
ADA_EUR	0.349619	0.073874	0.211298	1.0	0.000343
ETH_USDT	0.349351	0.074342	0.212800	1.0	0.000515
ETH_EUR	0.348706	0.074031	0.212303	1.0	0.000687
ETH_USD	0.348347	0.073985	0.212388	1.0	0.000343
ETH_USDC	0.348133	0.073887	0.212239	1.0	0.000343
LTC_EUR	0.347481	0.073847	0.212522	1.0	0.000343
LTC_USD	0.347192	0.074166	0.213616	1.0	0.000515
LTC_USDT	0.345606	0.074177	0.214629	1.0	0.000515

Table 42: Cluster B beta oil close distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ETH_USDT	0.088053	0.054046	0.613794	0.934089	0.621353
LTC_USDT	0.089284	0.058978	0.660566	0.913835	0.616203
ETH_USD	0.082108	0.053348	0.649727	0.909372	0.665808
ETH_EUR	0.077336	0.050887	0.658000	0.907999	0.699107
LTC_EUR	0.079576	0.057942	0.728139	0.869894	0.666495
BTC_DAI	0.076639	0.052000	0.678511	0.888088	0.698764
ETH_USDC	0.072889	0.051061	0.700531	0.875386	0.721593
LTC_USD	0.073718	0.057512	0.780171	0.838654	0.708720
BTC_USDT	0.070763	0.051826	0.732395	0.858222	0.742362
BTC_USD	0.064676	0.052750	0.815601	0.813594	0.769482
BTC_USDC	0.055023	0.050772	0.922757	0.748713	0.823893
XRP_USD	0.047423	0.055770	1.176008	0.640920	0.835565
BTC_EUR	0.043586	0.049337	1.131937	0.630621	0.878647
ADA_EUR	-0.045625	0.064303	1.409371	0.480604	0.800549
XRP_EUR	0.027407	0.055027	2.007773	0.434604	0.903364
ADA_USD	0.013149	0.049553	3.768738	0.263646	0.956746
XRP_USDT	0.011576	0.058233	5.030338	0.234123	0.919670
XLM_USD	0.005618	0.062304	11.090732	0.180227	0.898730

Table 43: Cluster B beta oil volume distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XRP_USDT	-0.148939	0.036804	0.247109	1.000000	0.087024
XRP_USD	-0.141886	0.035777	0.252157	1.000000	0.119636
XRP_EUR	-0.137918	0.035284	0.255835	1.000000	0.144010
XLM_USD	-0.119853	0.033535	0.279801	0.999657	0.278235
LTC_EUR	-0.116919	0.032799	0.280523	1.000000	0.312564
LTC_USD	-0.115111	0.032741	0.284429	0.999657	0.328871
LTC_USDT	-0.104413	0.032790	0.314043	0.998970	0.449022
BTC_USD	-0.086348	0.031191	0.361221	0.992448	0.675249
BTC_USDT	-0.085307	0.031350	0.367495	0.992448	0.680227
BTC_USDC	-0.084150	0.032139	0.381933	0.984552	0.687436
BTC_EUR	-0.079945	0.032037	0.400732	0.984552	0.735153
BTC_DAI	-0.077326	0.031362	0.405587	0.985239	0.757638
ADA_USD	-0.069140	0.032385	0.468397	0.963955	0.829214
ADA_EUR	-0.064212	0.033855	0.527232	0.935118	0.856162
ETH_USDC	-0.042642	0.032883	0.771130	0.795400	0.967559
ETH_USD	-0.036095	0.033993	0.941770	0.714727	0.978030
ETH_EUR	-0.031612	0.034687	1.097265	0.630621	0.981634
ETH_USDT	-0.029715	0.034397	1.157548	0.620323	0.984380

Table 44: Cluster B beta polarity flair list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ETH_USDC	-0.044714	0.042872	0.958810	0.718503	0.900618
ADA_USD	-0.042975	0.043202	1.005301	0.688980	0.905252
ETH_USD	-0.043029	0.042781	0.994250	0.684174	0.912633
ETH_USDT	-0.041952	0.041825	0.996962	0.689667	0.917096
ADA_EUR	-0.041898	0.042911	1.024180	0.681428	0.914178
ETH_EUR	-0.040658	0.042068	1.034672	0.675249	0.925678
XLM_USD	-0.039498	0.042169	1.067636	0.655681	0.926365
BTC_USD	-0.035302	0.042715	1.210006	0.601442	0.933917
LTC_EUR	-0.033596	0.042522	1.265686	0.577068	0.940783
LTC_USD	-0.033227	0.042533	1.280046	0.575009	0.943014
BTC_USDC	-0.033012	0.042240	1.279529	0.577755	0.943529
LTC_USDT	-0.032705	0.042091	1.286999	0.580158	0.945589
XRP_EUR	-0.032486	0.042323	1.302808	0.572949	0.943357
BTC_DAI	-0.031003	0.042495	1.370699	0.539307	0.947134
XRP_USDT	-0.028869	0.042672	1.478141	0.509784	0.951596
BTC_EUR	-0.028592	0.042665	1.492202	0.506694	0.955372
BTC_USDT	-0.027481	0.043642	1.588064	0.500172	0.953484
XRP_USD	-0.018936	0.046938	2.478782	0.352558	0.953141

Table 45: Cluster B beta polarity flair user followers count list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XRP_USD	0.096571	0.036409	0.377019	0.996910	0.574150
LTC_USD	0.092008	0.035847	0.389609	0.992448	0.621867
LTC_EUR	0.091071	0.034386	0.377575	0.993134	0.628562
XLM_USD	0.090678	0.034768	0.383417	0.992448	0.633539
XRP_USDT	0.088987	0.034098	0.383180	0.992104	0.652763
BTC_USD	0.087493	0.035949	0.410877	0.980776	0.661861
XRP_EUR	0.086081	0.033963	0.394553	0.986955	0.673532
LTC_USDT	0.085709	0.034005	0.396746	0.981119	0.679025
ETH_USD	0.084904	0.033814	0.398267	0.986612	0.686749
BTC_USDT	0.085455	0.035025	0.409868	0.980089	0.683831
BTC_DAI	0.084874	0.034215	0.403133	0.983522	0.677995
ETH_EUR	0.079217	0.034252	0.432381	0.970134	0.741160
ETH_USDT	0.078907	0.034346	0.435278	0.972880	0.742705
BTC_EUR	0.078868	0.035264	0.447122	0.966358	0.739959
BTC_USDC	0.076614	0.034767	0.453788	0.958805	0.757467
ETH_USDC	0.075783	0.034646	0.457172	0.953999	0.770340
ADA_USD	0.055157	0.042056	0.762482	0.796773	0.878991
ADA_EUR	0.048113	0.044380	0.922410	0.718160	0.901304

Table 46: Cluster B beta polarity flair user friends count list distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
ADA_EUR	-0.112042	0.054306	0.484692	0.975970	0.430827
ADA_USD	-0.104735	0.051816	0.494731	0.968417	0.480776
ETH_USDC	-0.097821	0.050042	0.511568	0.958462	0.524545
ETH_USDT	-0.092919	0.048744	0.524588	0.949537	0.559732
ETH_USD	-0.092263	0.049472	0.536203	0.941984	0.570889
ETH_EUR	-0.090580	0.048676	0.537387	0.946104	0.581188
XLM_USD	-0.090715	0.049142	0.541718	0.940954	0.585822
LTC_EUR	-0.084304	0.048667	0.577285	0.915551	0.628390
XRP_EUR	-0.084204	0.048680	0.578121	0.914864	0.632166
LTC_USDT	-0.082125	0.048512	0.590706	0.915551	0.650189
LTC_USD	-0.080382	0.048910	0.608466	0.896327	0.653793
XRP_USDT	-0.079935	0.048978	0.612718	0.895984	0.664435
BTC_USD	-0.077789	0.049828	0.640550	0.882252	0.667868
BTC_USDC	-0.075959	0.049209	0.647841	0.883968	0.684861
BTC_EUR	-0.075865	0.050305	0.663088	0.867834	0.692585
BTC_DAI	-0.074374	0.050174	0.674618	0.857878	0.693272
BTC_USDT	-0.072549	0.050445	0.695313	0.847580	0.697906
XRP_USD	-0.068359	0.051463	0.752828	0.808788	0.724854

Table 47: Cluster B beta quote closed issues distribution mean and significance 86 400 data

	μ	σ	$\left\|\frac{\sigma}{\mu}\right\|$	p-direction	ROPE
XLM_USD	0.887177	0.468184	0.527724	0.959492	0.026262
XRP_USDT	-0.466636	0.314630	0.674252	0.854102	0.077412
XRP_EUR	-0.458826	0.317247	0.691432	0.840371	0.075695
LTC_USDT	-0.402615	0.270805	0.672616	0.860281	0.091830
LTC_EUR	-0.348923	0.267102	0.765503	0.812221	0.124614
LTC_USD	-0.285782	0.272684	0.954168	0.726399	0.156025
ETH_USDC	-0.309896	0.427129	1.378298	0.557844	0.134398
BTC_USD	-0.307993	0.453204	1.471474	0.521456	0.139375
XRP_USD	-0.234686	0.312179	1.330197	0.576725	0.167697
ADA_EUR	-0.159370	0.555134	3.483304	0.274288	0.140577
BTC_USDC	-0.146479	0.451806	3.084452	0.292482	0.180742
ETH_USDT	-0.132832	0.440042	3.312777	0.276347	0.168211
ADA_USD	-0.143773	0.568865	3.956687	0.250601	0.140577
ETH_EUR	-0.129352	0.443171	3.426086	0.266735	0.169585
ETH_USD	-0.124463	0.442513	3.555370	0.261586	0.176965
BTC_DAI	0.094931	0.452775	4.769523	0.130793	0.177652
BTC_USDT	-0.021112	0.453547	21.482674	0.062479	0.179197
BTC_EUR	0.018804	0.448762	23.864914	0.006522	0.179712

Table 48: Cluster B beta quote forks distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
XLM_USD	-2.085298	0.310128	0.148721	1.000000	0.000000
LTC_USDT	0.526842	0.556559	1.056405	0.658084	0.098181
LTC_EUR	0.517828	0.559464	1.080406	0.658084	0.100240
BTC_EUR	-0.385964	0.255441	0.661825	0.874356	0.101270
BTC_DAI	-0.340233	0.249739	0.734025	0.814624	0.127188
BTC_USDT	-0.335507	0.252786	0.753445	0.813594	0.141092
BTC_USDC	-0.321950	0.244180	0.758441	0.808788	0.144696
ETH_USD	0.424621	0.601220	1.415897	0.515620	0.108308
ETH_USDT	0.395312	0.607264	1.536165	0.472709	0.112599
BTC_USD	-0.256185	0.238038	0.929165	0.707861	0.191212
ETH_USDC	0.327956	0.600458	1.830911	0.401648	0.116718
LTC_USD	0.291220	0.561339	1.927541	0.403021	0.130793
ADA_USD	-0.189420	0.237308	1.252813	0.570203	0.244422
XRP_EUR	0.177132	0.302718	1.708998	0.451081	0.221421
XRP_USDT	0.168736	0.306334	1.815466	0.407484	0.223653
ETH_EUR	0.226004	0.584764	2.587412	0.290422	0.130450
ADA_EUR	0.124202	0.239173	1.925686	0.391006	0.290766
XRP_USD	-0.009409	0.305607	32.478798	0.011672	0.254034

Table 49: Cluster B beta quote googletrends distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
LTC_USDT	0.488128	0.084949	0.174031	1.000000	0.000000
LTC_USD	0.406090	0.088573	0.218111	1.000000	0.000172
ADA_EUR	0.405521	0.075138	0.185288	1.000000	0.000000
ADA_USD	0.404986	0.074683	0.184409	1.000000	0.000000
LTC_EUR	0.396598	0.084722	0.213622	1.000000	0.000000
ETH_USDT	0.354115	0.067147	0.189619	1.000000	0.000000
ETH_USD	0.326194	0.071660	0.219686	1.000000	0.001373
ETH_EUR	0.290629	0.072521	0.249533	1.000000	0.003433
XLM_USD	-0.290426	0.065268	0.224731	1.000000	0.001202
ETH_USDC	0.286262	0.058763	0.205278	1.000000	0.001030
XRP_USD	-0.212442	0.081622	0.384209	0.991761	0.087195
XRP_USDT	-0.177221	0.079935	0.451047	0.974253	0.160144
XRP_EUR	-0.164281	0.076665	0.466672	0.964641	0.198764
BTC_DAI	-0.120161	0.060923	0.507014	0.949880	0.372297
BTC_USD	-0.089063	0.072673	0.815979	0.776862	0.555956
BTC_USDC	-0.072735	0.057645	0.792534	0.781668	0.688637
BTC_EUR	-0.078775	0.072138	0.915743	0.712667	0.604703
BTC_USDT	-0.041667	0.066545	1.597061	0.468246	0.794370

Table 50: Cluster B beta quote open issues distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_EUR	-0.325819	0.066640	0.204530	1.000000	0.000343
BTC_USDT	-0.303207	0.062204	0.205153	0.999657	0.001202
BTC_USDC	-0.293146	0.063421	0.216346	0.999657	0.003090
BTC_DAI	-0.284279	0.062216	0.218855	1.000000	0.004291
BTC_USD	-0.276701	0.063047	0.227852	1.000000	0.005149
ETH_USDC	-0.266800	0.050095	0.187762	1.000000	0.001202
ETH_USD	-0.261458	0.050962	0.194915	1.000000	0.001716
ETH_USDT	-0.260927	0.050691	0.194274	1.000000	0.003090
ETH_EUR	-0.247148	0.051877	0.209901	1.000000	0.003948
ADA_USD	-0.253308	0.100659	0.397379	0.963611	0.061277
ADA_EUR	-0.253285	0.100754	0.397788	0.961895	0.061277
LTC_USD	-0.246591	0.099380	0.403017	0.956402	0.073292
LTC_USDT	-0.245256	0.101063	0.412072	0.950566	0.073292
XRP_USD	-0.236467	0.085278	0.360635	0.974940	0.068829
LTC_EUR	-0.240415	0.102957	0.428247	0.937521	0.079300
XRP_USDT	-0.225953	0.086310	0.381985	0.970477	0.086337
XRP_EUR	-0.220794	0.087252	0.395174	0.964298	0.097151
XLM_USD	-0.134016	0.129980	0.969888	0.673189	0.311535

Table 51: Cluster B beta quote pull requests merged distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_USDT	2.375477	0.590983	0.248785	1.000000	0.000000
BTC_DAI	2.360080	0.555063	0.235188	1.000000	0.000000
BTC_USDC	2.252086	0.538347	0.239044	1.000000	0.000000
BTC_EUR	2.242213	0.562227	0.250747	1.000000	0.000000
BTC_USD	2.016296	0.522359	0.259069	0.999657	0.000172
XLM_USD	-0.916144	0.511385	0.558192	0.925506	0.028150
XRP_USD	-0.820514	0.170535	0.207839	1.000000	0.000000
LTC_USD	-0.799754	0.183986	0.230053	1.000000	0.000000
LTC_EUR	-0.785386	0.185300	0.235935	1.000000	0.000000
LTC_USDT	-0.773898	0.187224	0.241923	1.000000	0.000172
XRP_USDT	-0.668096	0.157884	0.236319	1.000000	0.000343
XRP_EUR	-0.664690	0.157042	0.236263	1.000000	0.000172
ETH_USD	-0.330009	0.398531	1.207637	0.590800	0.144696
ETH_USDT	-0.295692	0.385183	1.302650	0.561277	0.155853
ETH_EUR	-0.242291	0.389585	1.607921	0.467216	0.168211
ADA_EUR	0.276528	1.506919	5.449434	0.150017	0.049262
ADA_USD	0.255787	1.525111	5.962425	0.139719	0.052180
ETH_USDC	0.024740	0.385302	15.574009	0.068314	0.201167

Table 52: Cluster B beta quote reddit subscribers distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_USD	5.719340	0.959778	0.167813	1.000000	0.000000
ETH_EUR	5.668145	0.929466	0.163981	1.000000	0.000000
ETH_USDT	5.633906	0.960799	0.170539	1.000000	0.000000
ETH_USDC	4.825280	0.995300	0.206268	1.000000	0.000000
BTC_DAI	-4.586834	0.560109	0.122112	1.000000	0.000000
XRP_USDT	4.423541	0.859987	0.194411	1.000000	0.000000
XRP_EUR	4.267279	0.829579	0.194405	1.000000	0.000000
BTC_USDC	-4.083642	0.552606	0.135322	1.000000	0.000000
BTC_EUR	-4.070753	0.614948	0.151065	1.000000	0.000000
BTC_USDT	-4.034149	0.600741	0.148914	1.000000	0.000000
XRP_USD	3.865317	0.859480	0.222357	1.000000	0.000000
BTC_USD	-3.360615	0.587058	0.174688	1.000000	0.000000
ADA_USD	2.175823	0.811844	0.373121	0.995881	0.002746
ADA_EUR	1.622331	0.824571	0.508263	0.960522	0.011843
LTC_USDT	1.012234	0.472522	0.466811	0.965328	0.016135
LTC_USD	0.712756	0.441742	0.619765	0.896327	0.048404
LTC_EUR	0.676213	0.459524	0.679554	0.851356	0.056986
XLM_USD	-0.062884	0.929399	14.779693	0.047031	0.084277

Table 53: Cluster B beta quote stars distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ADA_USD	-1.556580	0.648709	0.416753	0.991761	0.006179
ADA_EUR	-1.124376	0.620970	0.552280	0.943357	0.023172
LTC_USDT	-0.955442	0.378541	0.396194	0.987642	0.009784
LTC_EUR	-0.672024	0.361422	0.537811	0.937865	0.040508
XRP_USD	0.788039	0.628548	0.797610	0.792997	0.058187
XLM_USD	0.652406	0.461367	0.707177	0.845520	0.063337
BTC_EUR	0.682864	0.580832	0.850583	0.759355	0.075867
BTC_DAI	0.656959	0.557470	0.848562	0.768967	0.075867
LTC_USD	-0.542904	0.355720	0.655217	0.871267	0.066941
BTC_USDC	0.609794	0.561488	0.920782	0.727085	0.085479
ETH_USDT	-0.667230	0.783085	1.173635	0.623069	0.074665
ETH_USD	-0.646341	0.804938	1.245378	0.589427	0.082218
XRP_USDT	0.554982	0.608060	1.095640	0.637830	0.091658
ETH_EUR	-0.615323	0.790796	1.285173	0.571232	0.084277
ETH_USDC	-0.588757	0.769061	1.306246	0.556471	0.082046
BTC_USDT	0.461939	0.578358	1.252021	0.568486	0.108994
XRP_EUR	0.259809	0.584749	2.250688	0.343632	0.132166
BTC_USD	0.230996	0.534953	2.315854	0.323035	0.138517

Table 54: Cluster B beta quote total issues distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
XLM_USD	1.442484	0.461225	0.319744	0.999313	0.001373
XRP_USD	1.115117	0.342648	0.307275	0.995537	0.002231
XRP_EUR	1.101272	0.349183	0.317072	0.992791	0.003605
ETH_EUR	1.096034	0.401467	0.366291	0.986612	0.007037
LTC_EUR	1.080341	0.355932	0.329463	0.994507	0.003605
XRP_USDT	1.078286	0.347718	0.322473	0.995194	0.003090
BTC_DAI	1.072925	0.389588	0.363108	0.990045	0.008067
ETH_USDT	1.071267	0.392175	0.366085	0.982149	0.006179
LTC_USD	1.058445	0.360590	0.340679	0.992448	0.003776
ETH_USD	1.063733	0.390380	0.366991	0.986955	0.008754
LTC_USDT	1.045346	0.356995	0.341509	0.991761	0.005321
ADA_EUR	1.065659	0.443163	0.415858	0.970820	0.012358
ADA_USD	1.058629	0.433689	0.409670	0.973223	0.010470
ETH_USDC	1.000381	0.396721	0.396570	0.975970	0.011329
BTC_USDT	0.978204	0.407476	0.416555	0.966014	0.014246
BTC_USDC	0.972650	0.401057	0.412334	0.971164	0.013732
BTC_USD	0.953434	0.412736	0.432894	0.965328	0.015105
BTC_EUR	0.950990	0.415140	0.436534	0.958119	0.018366

Table 55: Cluster B beta quote twitter followers distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ETH_USD	-5.858657	0.957018	0.163351	1.000000	0.000000
ETH_USDT	-5.748200	0.960580	0.167110	1.000000	0.000000
ETH_EUR	-5.692123	0.923362	0.162217	1.000000	0.000000
XRP_USDT	-5.052417	0.470665	0.093156	1.000000	0.000000
ETH_USDC	-5.013680	0.993368	0.198132	1.000000	0.000000
XRP_USD	-4.719234	0.458400	0.097134	1.000000	0.000000
XRP_EUR	-4.669728	0.479113	0.102600	1.000000	0.000000
ADA_EUR	-0.493881	0.416868	0.844066	0.762788	0.093718
LTC_EUR	-0.273444	0.053796	0.196736	1.000000	0.000343
LTC_USD	-0.255350	0.054972	0.215280	1.000000	0.002918
LTC_USDT	-0.239434	0.054395	0.227181	1.000000	0.006351
ADA_USD	-0.321498	0.378762	1.178117	0.605218	0.150875
BTC_USDT	-0.187753	0.089263	0.475426	0.960179	0.161861
BTC_EUR	-0.177068	0.090610	0.511727	0.951596	0.198078
BTC_DAI	-0.153171	0.088160	0.575570	0.911775	0.272228
BTC_USD	-0.139088	0.087402	0.628390	0.886715	0.326639
BTC_USDC	-0.127761	0.087099	0.681730	0.862685	0.380879
XLM_USD	0.285670	1.104535	3.866473	0.205630	0.067800

Table 56: Cluster B beta sp500 close distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_USD	0.510049	0.104683	0.205240	1.000000	0.000000
BTC_DAI	0.508410	0.096724	0.190248	1.000000	0.000000
BTC_USDC	0.491401	0.099109	0.201687	1.000000	0.000000
XRP_USD	0.430661	0.106523	0.247348	0.999657	0.001202
BTC_EUR	0.428447	0.100075	0.233577	1.000000	0.000515
XRP_EUR	0.425823	0.106033	0.249007	1.000000	0.001202
XRP_USDT	0.416831	0.104462	0.250611	1.000000	0.001202
BTC_USDT	0.416252	0.099986	0.240204	1.000000	0.000687
ETH_USD	0.170338	0.124175	0.728989	0.833162	0.263302
ETH_USDT	0.164255	0.121818	0.741637	0.817027	0.269653
ETH_EUR	0.153872	0.122096	0.793490	0.789907	0.308960
XLM_USD	0.158875	0.136870	0.861493	0.761758	0.302952
ADA_USD	0.120871	0.097138	0.803651	0.786818	0.406454
LTC_USDT	0.117349	0.106607	0.908464	0.728802	0.406111
ADA_EUR	0.103490	0.086927	0.839957	0.768280	0.473910
LTC_USD	0.109966	0.107215	0.974984	0.694473	0.428424
ETH_USDC	0.112483	0.125219	1.113229	0.629935	0.408514
LTC_EUR	0.100083	0.106270	1.061813	0.659114	0.454686

Table 57: Cluster B beta sp500 volume distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
ADA_EUR	0.190080	0.034444	0.181210	1.000000	0.000172
ADA_USD	0.189736	0.035563	0.187432	1.000000	0.000172
XLM_USD	0.179274	0.030963	0.172712	1.000000	0.002231
BTC_USDT	0.169819	0.028444	0.167497	1.000000	0.009955
BTC_DAI	0.168258	0.028084	0.166908	1.000000	0.012015
BTC_EUR	0.164831	0.028637	0.173733	1.000000	0.019739
BTC_USD	0.164769	0.027704	0.168139	1.000000	0.017851
BTC_USDC	0.164471	0.028514	0.173370	1.000000	0.018366
LTC_EUR	0.160789	0.028740	0.178742	1.000000	0.030724
XRP_USDT	0.160617	0.028649	0.178368	0.999657	0.029180
XRP_EUR	0.159215	0.028513	0.179084	1.000000	0.031754
ETH_USDT	0.159115	0.028783	0.180894	1.000000	0.033127
XRP_USD	0.158876	0.030040	0.189081	0.999657	0.039478
ETH_EUR	0.157504	0.028725	0.182375	0.999657	0.040336
LTC_USD	0.157202	0.029811	0.189636	0.998970	0.044628
LTC_USDT	0.156403	0.029360	0.187720	1.000000	0.042053
ETH_USD	0.155487	0.029589	0.190297	0.999657	0.046172
ETH_USDC	0.153219	0.030428	0.198594	0.999657	0.058187

Table 58: Cluster B intercept const distribution mean and significance 86 400 data

	μ	σ	$\left\lvert\frac{\sigma}{\mu}\right\rvert$	p-direction	ROPE
BTC_USD	0.785356	0.081521	0.103801	1.000000	0.000000
BTC_EUR	0.736349	0.078396	0.106465	1.000000	0.000000
BTC_USDT	0.710311	0.071924	0.101257	1.000000	0.000000
BTC_DAI	0.698710	0.066626	0.095356	1.000000	0.000000
BTC_USDC	0.671672	0.073251	0.109058	1.000000	0.000000
XRP_USD	-0.179991	0.080310	0.446192	0.974940	0.154995
XRP_USDT	-0.167057	0.087994	0.526728	0.946790	0.219876
XRP_EUR	-0.108822	0.077745	0.714420	0.847923	0.456231
LTC_USDT	-0.084065	0.084632	1.006742	0.688980	0.562307
ETH_USD	-0.074102	0.097895	1.321072	0.543769	0.570717
ADA_EUR	-0.053585	0.096977	1.809767	0.414693	0.624271
ADA_USD	0.047488	0.099170	2.088330	0.363543	0.634912
ETH_USDT	-0.036337	0.082442	2.268789	0.342945	0.726056
ETH_USDC	0.029259	0.074207	2.536201	0.303124	0.793169
LTC_EUR	-0.020383	0.075000	3.679484	0.208720	0.806042
LTC_USD	-0.015800	0.074195	4.696016	0.169928	0.812221
XLM_USD	0.004027	0.074328	18.456546	0.057329	0.821490
ETH_EUR	-0.002045	0.091211	44.592364	0.022314	0.730690

Appendix D Draft Timeline

This draft timeline was created as part of the planning process at the beginning of work on this Master thesis. This plan was largely followed, but multiple iterations of re-writes and some restructuring required the buffer time to be used. The fixed deadlines of Kickoff, Greenlight and Defense were all met.

The complexity of some of the data aggregation and pre-processing is high and needs to be concluded before any analysis of the hypothesis is possible. It is likely that the data collection phase will be a substantial part of the time spent on this thesis. Especially sentiment evaluation solutions need to be tested and applied correctly to be useful.

The general pre-processing steps would be cleaning the data and formatting it into a consistent schema, followed by descriptive statistics and understanding the dataset. Next the correlation with the price data can be tested using some off-the-shelf approaches. After the kickoff in March the literature study and data aggregation phase will begin. This is expected to take until roughly mid to end of April. Real-time Twitter recording will likely continue beyond this point as well, the longer the time range covered the more significant any found relations are expected to be. During the aggregation phase, findings of the literature study and the methodology will be documented in the thesis draft.

In the next step data pre-processing can begin and first review of datasets can occur. Also sentiment analysis approaches are applied to the obtained tweets. This is expected to take roughly two to three weeks. After which the datasets should be in a state to facilitate easy and quick analysis of relations between various data series. The ease of use is important to allow for flexibility in the final analysis, as potentially new relations might be thought up as a result of reviewing the datasets in the pre-processing phase.

Next the main analysis can be performed. First results are document in the draft report. Based on these results an iteration of the pre-processing with a second analysis phase might be appropriate. In the ideal case a model could be developed linking the effects of the investigated variables together in a single model, to explain both the effect on cryptocurrency market capitalisation and the effects of the variables on each other. The analysis phase is expected to take 1 to 4 weeks time. This would likely be concluded in end of June to mid July.

Finally, this is followed by a final write-up and discussion of the findings and creation of presentation material. Which is expected to allow for presentation of the thesis at some point in July or August.

This is of course only a high level breakdown of the time-line, but should at least illustrate the expected distribution of time expenditure on the different phases. Figure 54 and Figure 55 show a visual breakdown of the expected timeline.

Name	Begin date	End date
Research Proposal	01/02/2021	02/03/2021
Kickoff	03/03/2021	03/03/2021
Literature Reserach	03/03/2021	30/03/2021
Data Aggregation	03/03/2021	15/04/2021
Data Preprocessing	31/03/2021	30/04/2021
Correlation Analysis	16/04/2021	14/05/2021
Discussion and Analysis Iteration	03/05/2021	02/06/2021
Feedback and Improvements	17/05/2021	09/06/2021
Writeup and Documentation *This process happens throughout but is finalised here.*	03/06/2021	16/06/2021
Buffer Time	17/06/2021	14/07/2021
Finalisation and Presentation	15/07/2021	11/08/2021
Green Light	15/07/2021	15/07/2021
Defense	30/08/2021	30/08/2021

Figure 54: Tasks and expected corresponding dates

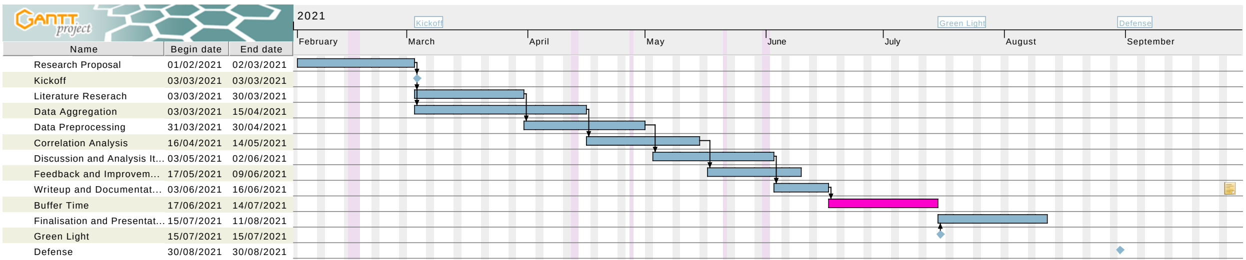

Figure 55: Gantt Chart