Renato Pezzotti

Metcalfe's law to predict cryptocurrencies: a comparative analysis.

Curitiba - PR, Brasil

2019

Renato Pezzotti

Metcalfe's law to predict cryptocurrencies: a comparative analysis.

apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de mestre em Informática.

Pontifícia Universidade Católica do Paraná - PUCPR

Programa de Pós-Graduação em Informática - PPGIa

Supervisor: Prof. Dr. Julio Cesar Nievola

Co-supervisor: Prof. Dr. Emerson Cabrera Paraiso

Curitiba - PR, Brasil

2019

Renato Pezzotti
Metcalfe's law to predict cryptocurrencies: a comparative analysis.
Renato Pezzotti. – Curitiba - PR, Brasil, 2019-
83 p. : il. (algumas color.) ; 30 cm.

Supervisor: Prof. Dr. Julio Cesar Nievola

Dissertação OU Tese –
Pontifícia Universidade Católica do Paraná - PUCPR
Programa de Pós-Graduação em Informática - PPGIa, 2019.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador.

II. Universidade xxx. III. Faculdade de xxx. IV. Título

Acknowledgements

À minha querida esposa Cristina Miyadaira Pezzotti por toda paciência e apoio nos momentos divertidos e complicados que esses últimos anos tiveram. Pelos momentos silenciosos, onde só o meu teclado fazia barulho, você aparecia no quarto com um copo de café que só eu sei como revigorava.

Aos meus amados pais Luiz e Marisa Pezzotti que independente de qualquer coisa sempre me deram o apoio e a condição necessários para me tornar quem hoje sou.

Ao meu orientador Julio Nievola, a quem devo uma medalha por conta da paciência com que me guiou ao longo desse período.

Aos demais professores, colaboradores, colegas e amigos que de alguma forma participaram para este resultado.

A todos aqui mencionados deixo meu humilde e mais sincero agradecimento!

As pedras fundamentais da minha conquista também são suas! Contem sempre comigo!

"A gratidão não é somente a maior das virtudes; É também a mãe de todas as outras" Cícero

—

Renato Pezzotti

Abstract

In face of the current transformation pace of the society, its being observed great developments in the financial area, most recently one can observe the advent of a significant number of the so called cryptocurrencies

Such situation is leading institutions to adapt in order to overcome current limitations that may lead the decision to change from the current system to the new born, blockchain based one.

This dissertation's objective is to offer a better understanding about cryptocurrencies behaviour, by testing value prediction methods ARIMA, Metcalfe and investigating how sentiment analysis in social media relates to its movement.

In the course of this comparative study Bitcoin, Ether, Litecoin and Bitcoin Cash are evaluated by an unbiased analysis of the most recent and most used approaches in the trading decision making process by those who are the early adopters (investors) of this new technology.

Keywords: Blockchain, Bitcoin, Ethereum, Ether, Litecoin, Bitcoin Cash, Metcalfe, Prediction, Time Series, Cryptocurrency, Currency, ARIMA, Financial, Economics.

Resumo

Em face do ritmo em que a sociedade atual passa por transformações, eram previstos diversos desenvolvimentos relacionados com a área financeira. Um desses desenvolvimentos mais recentes da conta do surgimento das criptomoedas, gerenciadas em blockchains públicos, que vem ganhando força e perdendo estigmas como meio de pagamento, principalmente online.

Esta situação está levando instituições financeiras a se adaptar para evitar que seus clientes façam a opção de mudar do atual sistema para suas evoluções.

O objetivo desta dissertação é entender um pouco mais sobre criptomoedas, testando os métodos de predição de valor baseados em estatística (ARIMA neste caso), Lei de Metcalfe e investigando como a análise de sentimentos nas mídias sociais pode se relacionar com seu comportamento.

No decorrer deste estudo Bitcoin, Ether, Litecoin e Bitcoin Cash, foram os ativos estudados, sobre os quais foi reuninda uma visão imparcial das abordagens mais recentes e mais utilizadas no processo de tomada de decisão por aqueles que são os primeiros adeptos (investidores) desta nova tecnologia.

Palavras-chave: Blockchain, Bitcoin, Ethereum, Ether, Litecoin, Bitcoin Cash, Metcalfe, Predição, Séries Temporais, Criptomoédas, Moedas, ARIMA, Finanças, Economia.

List of Figures

List of abbreviations and acronyms

BTC	Bitcoin
ETH	Ether
LTC	Litecoin
BCH	Bitcoin Cash
ARIMA	Auto Regressive Integrated Moving Average
AR	Auto Regressive
MA	Moving Average
I	Integrated
ARMA	Auto Regressive Moving Average
RSS	Residual Sum Squared
MSE	Mean Squared Error
DAA	Daily Active Addresses

Contents

1 Introduction

1.1 Motivation

Bitcoin (BTC) emerged and became a impressive phenomenon over the past couple years, with prices reaching over 20.000,00 USD per BTC, and it brought questions even to the most experienced specialist, whether there was a bubble in the market or if we could be seeing a potential disruptive change in the markets.

But Bitcoin didn't arrive alone. A true avalanche of the so called cryptocurrencies[1] arrived to the market with the most diverse formats and motivations, introducing a fairly fertile field for research and study on how the market dynamics might change in the future with these new assets playing a role increasingly more important and significant for the investors and the population in general.

The motivation for this dissertation starts on the fact that although called currencies, some, if not most, of the existent techniques in place to forecast currencies fail Bitcoin and that lead people to look for alternatives, where some found out that mathematical laws which explain network connectivity were in some ways explaining the value model, assuming Bitcoin behaves as a "Fiat"[2] currency, even though it doesn't fulfill the definition of "Fiat" in a strict way.

The network valuation laws mentioned are mostly known by the computer science world, but not by the investments and financial worlds. Metcalfe's law (METCALFE, 2013) itself is relatively untested, and, although it was proposed in the 1980's, it was only proved by Metcalfe himself with recent data from Facebook and Tencent, in an article published years later, in 2015.

This work intends to shed more light on this topic, moved by the belief of the author that more understanding is needed on the behavior not only of the Bitcoin but of all the emerging dominant cryptocurrencies, and that would not happen without academic research and experiments.

1.2 Objective

The main objective of this dissertation is to offer a comparative analysis of results from a selection among well known time-series models to fit currency price prediction and the Metcalfe law for the top cryptocurrencies in the market today.

[1] Currencies based on tokens held and transacted in a blockchain technology
[2] Without intrinsic value and mostly backed by the government that issued it

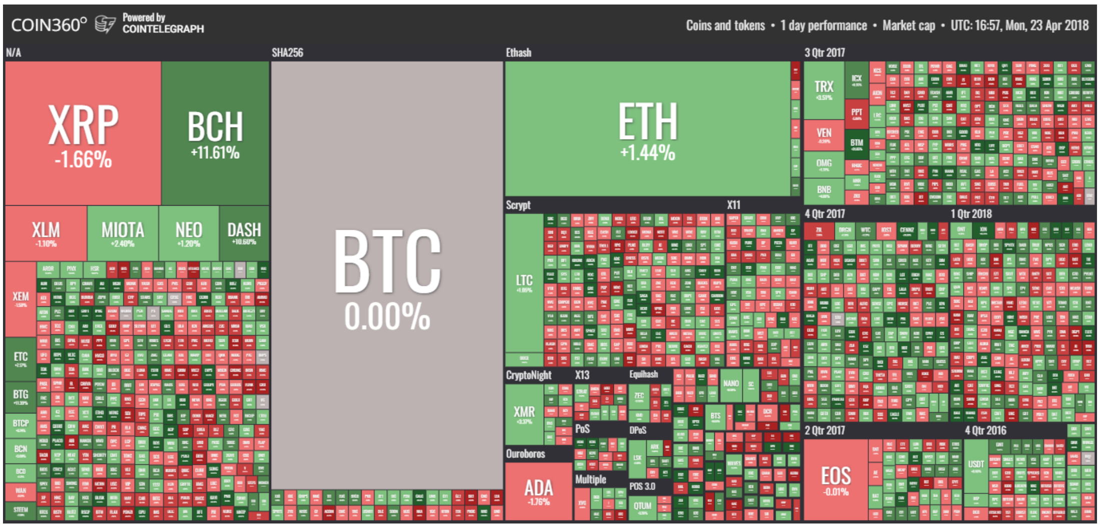

Figure 1 – Cryptocurencies Market Snapshot

Source: https://coin360.io

Also among this dissertation objective is to analyze the impact of sentiment analysis on social networks as an explanatory factor by itself and as a correction factor for the tested models in a specific time frame.

We are going to explore the laws and techniques applied to the most relevant cryptocurrencies in the market nowadays. The figure 1 shows a snapshot of the cryptocurrency market, with the size of the blocks showing the market slice of each currency.

This dissertation does not propose itself to be a review of all existent models, because that would be a nearly impossible job due to the number of variations used from a single model, but instead, the objective is to test the most common ones side by side with the most innovative ones, evaluate new approaches and offer further proof that each model work or not and deliver more mature perspectives to pave way to further research.

1.2.1 Hypothesis

Considering the characteristics of the underlying technology that enables the cryptocurrencies, this dissertation set as null hypothesis the idea that the network connectivity value model tested will better explain the value of the tokens.

According to a previous research by Peterson (PETERSON, 2017) and by Alabi (ALABI, 2017) Bitcoin price, in the medium to long-term, appears to follow Metcalfe's law, with high correlation depending on periods used. Therefore should be reasonable to assume the same is applicable to other cryptocurrencies held in similar structures.

If we manage to prove that the cryptocurrencies are behaving similarly, we then enable the development of models that can help to forecast growth, predict and avoid losses for the early adopters of these assets

Also, considering the entirely technology based, decentralized and independent characteristic of every cryptocurrency, this dissertation is assuming as null hypothesis that the impact the sentiment analysis in social networks can be a explanatory factor to the movements of the market, in a similar way that behavioral economics suggests that emotions affecting individual decision could impact in the market. It is therefore also reasonable to assume that the public mood and sentiment can drive market values as much as news (NOFSINGER, 2005).

A study described in an article by Kaminski (KAMINSKI, 2014) suggests that emotional sentiments rather mirror the market than that they make it predictable. Therefore this dissertation aims mainly to use sentiment data together with existing network models rather as a single feature to explain the prices.

1.3 Scientific Contribution

This dissertation will offer further proof that the cryptoassets behavior can be predicted and pave more way to new research in the cryptocurrency and blockchain subjects. The world is facing a potential new reality where the technology is allowing the resolution of certain problems, double expending for example, and are potentially threatening the existing financial structures.

These new currencies know no borders and have already scared and pushed governments to take actions, and implement regulations. While this is happening very fast, the research on this field is not demonstrating the same speed.

While in the economics field, predictive models are being explored for decades, these new assets are imposing challenges that, in a certain way, forces the research areas to adapt existing models and develop new ones to properly reflect the reality.

1.4 Scope

Within the scope of this document is:

- The extraction, cleaning and analysis of data from the most relevant cryptocurrencies in the market at the moment.
- Test and evaluate how the Auto Regressive Integrated Moving Average (ARIMA) model behave in predicting the price of these cryptocurrencies.
- Test and evaluate how the network valuation models behave in predicting the price and behaviour of these cryptocurrencies.

- The extraction, cleaning and analysis of data relevant social network communities.
- Evaluate the usage sentiment analysis on social media using Vader (HUTTO; GILBERT, 2014) lexicon to derive metrics to be used to improve the network models.
- Explore the potential correlation between the observed sentiment information and the cryptocurrencies studied.

Its NOT in the scope of this document:

- The development of a new models (not mentioned already) to do time series analysis.
- The development of a new lexicon dedicated to sentiment analysis.

1.5 Document Structure

Besides this brief introduction, this dissertation will contain the following chapters:

- **02 - State of The Art:** Brief literature review on the what is the the most advanced understanding of topics explored in this study.
- **03 - Methodology:** Explanation of the methods to be followed to analyze the data collected and present the results.
- **04 - Analysis:** Presentation of the results achieved.
- **05 - Conclusion:** Reevaluate the research questions and present the findings related to each one, proposing from there new opportunities for further research.

2 State of The Art

2.1 Time Series Forecasting

The definition of time series is a collection of data points in time order. Examples of time series are heights of ocean tides, counts of sunspots, stock prices or in our case currency pair values (the value of a currency in another currency).

A collection of techniques was developed seeking to extract meaningful useful information or statistics that explain the characteristics and behavior of time series. From that a number of models were developed seeking to forecast future values or gaps in existent data, all based on observed data.

The International Institute of Forecasters was founded in 1979 and on its Silver Jubilee (back in 2006) an article was published by De Gooijer (De Gooijer; HYNDMAN, 2006) reviewing the past research in journals managed by the institute, which is a great reference for each of these models.

A few groups put together these models in a more organized an understandable way, but in general these models study the phenomena and try put in place different random (or stochastic) processes to enable forecasting of future values.

2.1.1 Auto Regressive (AR) Model

This model predicts the value of the variable in question based on a linear combination of previous values of the variable itself, thus AUTO regression. It is a very flexible and can handle a number of situations and normally its used to stationary data.

2.1.2 Moving Average (MA) Model

Moving average model is different from the moving average concept.

Its used to model time series of a single variable, assumed that the variable being predicted depends only on its present value and various past values of an error term.

2.1.3 Integrated (I) Model

The previous AR and MA models work well only with stationary time series, hence the Integrated model is introduced as a differencing step that can be applied one or more time to turn a non stationary series into stationary.

2.1.4 Combined models

A number of models are born from the mix up of the previous models mentioned. We are not going to describe each and every one of those, but the necessary focus will be given to the selected one in the further steps of this dissertation. To mention a few of these models:

- Auto Regressive Moving Average (ARMA)
- Auto Regressive Integrated Moving Average (ARIMA)
- Auto Regressive Fractionally Integrated Moving Average (ARFIMA)
- Vector Auto Regressive Moving Average(VARMA)
- Vector Auto Regressive Integrated Moving Average(VARIMA)

2.1.5 Non Linear Models

These models are designed to handle non-linear time series, or in other word, time series that the current observations might have non-linear relation to the past data points. These models mainly observe the changes of the variance, or heteroskedasticity of the series. Along these lines we have a number of models that derive or evolve one from the other. To mention a few:

- Auto Regressive Conditional Heteroskedasticity (ARCH)
- Generalized Auto Regressive Conditional Heteroskedasticity (GARCH)
- Integrated Generalized Auto Regressive Conditional Heteroskedasticity (IGARCH)

In summary, the idea of Time Series analysis is to explore the past and current behavior of something (some variable that changes over time like a stock price) and try to use such knowledge to make predictions and forecasts of the future movements of that variable.

2.2 Metcalfe's Law

2.2.1 Who is Metcalfe?

Bob Metcalfe, according to his own article (METCALFE, 2013), is Professor of Innovation and Murchison Fellow of Free Enterprise at The University of Texas at Austin. Metcalfe received a PhD in computer science from Harvard University. He is a member of the National Academy of Engineering and a Life Trustee of his alma mater, the Massachusetts Institute of Technology.

Metcalfe is recognized by the market and by the academy, having received IEEE Medal of Honor and the National Medal of Technology for his contribution leading the invention, standardization, and commercialization of Ethernet.

He published an article in 2013 where he offers further proof for his law and openly mention the criticisms his law received over time, mentioning that some considered the law wrong and dangerous. In the same article Metcalfe show that the law is valid and works well to fit Facebook data, results that were later reinforced with the work by Zhang (ZHANG; LIU; XU, 2015) where besides the Facebook data, fit the law to Tencent data (at the time was the biggest social network in China).

2.2.2 The Law

As described by Peterson (PETERSON, 2017):

Metcalfe (n^2): *Metcalfe's law (METCALFE, 2013) is based on the mathematical tautology describing connectivity among n users. As more people join a network, they add to the value of the network nonlinearly; i.e., the value of the network is proportional to the square of the number of users. The underlying mathematics for Metcalfe's law is based on pair-wise connections (e.g., telephony). If there are 4 people with telephones in a network, there could be a total of $3 + 2 + 1 = 6$ connections. This law, like most other laws, assumes equality among the members' network connections. The full math for Metcalfe's reasoning leads to the sum of all possible pairings between user, so the value of the network of size n is:*

$$\frac{n(n-1)}{2} \tag{2.1}$$

Which translates asympotically to:

$$n^2 \tag{2.2}$$

Metcalfe applies a proportionality factor (A), which he admits may decline over time. Metcalfe's law was originally designed to identify the break-even n where total network costs (cxn) are recouped. It is expressed more precisely as:

$$c \times n = M = A \times \frac{n(n-1)}{2} \tag{2.3}$$

2.2.3 The Network Effect

What the law proposed by Metcalfe looks to explain in essence is the "Network Effect", or the phenomena where a network has its value increased whether the number of participants grow. Like in a telephone network with very few users or the internet in its early days, compared to today's version of both.

For quite some time these models were used by companies to explain its own values, or the value of its customers.

A number of network models were proposed to describe the network effect, all of them proposing the proportionality in some way related to the number of users or nodes, to mention a few:

- Sarnoff's[1] law: $V \propto n$
- Odlyzko's[2] law: $V \propto nlog(n)$
- Reed's law: $V \propto 2^n$
- Metcalfe's law: $V \propto n^2$

These models were the ones tested by Zhang (ZHANG; LIU; XU, 2015) in a comparison, where the results shown that Metcalfe's law was the best fit.

Metcalfe uses a specific sigmoid function that he called "Netoid", limiting the data not as a function of N but as a function of time. This Netoid function is driven by a number of parameters that determines the growth, viral or not, of a network.

2.2.4 The sky is not the limit

Some of the critics to Metcalfe's law were attended by the acceptance of the fact that a network "do not grow to the sky". A team of researches,among them Briscoe and Odlyzko (BRISCOE; ODLYZKO; TILLY, 2006) published an article named "Metcalfe's Law is Wrong" where he extensively discuss the original Metcalfe proposition defending that his law overestimates the network value and proposing the already mentioned Odlyzko's Law.

This will be considered further in this dissertation, when evaluating the application of Metcalfe law over the cryptocurrencies data.

1 regarded as Father of American Television
2 one of the biggest critics of Metcalfe Law

2.2.5 So old yet so recent

The law of Metcalfe and its derivates by themselves mentioned as state of the art might sound somewhat strange. But an avalanche of explorations with these ideas trying to explain the cryptomarket behaviour is giving to these laws a new purpose. Its early to tell if these are going to really become some sort of standard, but the fact currently is that the recomendation for investors is to use this as one extra indicator in the arsenal to support trading decisions.

With this stated, the law is explored as state of the art on the subject of "network value laws explaining blockchains".

2.3 Cryptocurrency

A short definition to cryptocurrency, according to the Investopedia website[3], cryptocurrency is "a digital or virtual currency that uses cryptography for security. A cryptocurrency is difficult to counterfeit because of this security feature. A defining feature of a cryptocurrency, and arguably its most endearing allure, is its organic nature; it is not issued by any central authority, rendering it theoretically immune to government interference or manipulation. The first cryptocurrency to capture the public imagination was Bitcoin, which was launched in 2009 by an individual or group known under the pseudonym Satoshi Nakamoto. As of September 2015, there were over 14.6 million Bitcoins in circulation with a total market value of $3.4 billion. Bitcoin's success has spawned a number of competing cryptocurrencies, such as Ether, Ripple, Litecoin, Namecoin and PPCoin."

Cryptocurrency is in fact a implementation of blockchain, and quite regularly people mix up both, therefore, the next section intends to clarify the difference.

2.3.1 Blockchain

When Bitcoin was conceived by Satoshi Nakamoto in the "Bitcoin A Peer-to-Peer Electronic Cash System" (NAKAMOTO, 2008), even though it wasn't the first attempt to secure data into a chain of blocks cryptographicaly related, his (or their) intent was to support a currency based on a blockchain technology, and therefore was very common for people to call one by others name, but in fact, the blockchain technology itself evolved from the idea used only to support the currency (in fact its older than that).

Blockchain, according to Swan (SWAN, 2015) in her book, should be thought "as another class of thing like the internet. A comprehensive information technology with tiered levels and multiple classes of applications for any form of asset registry, inventory,

[3] https://www.investopedia.com/terms/c/cryptocurrency.asp

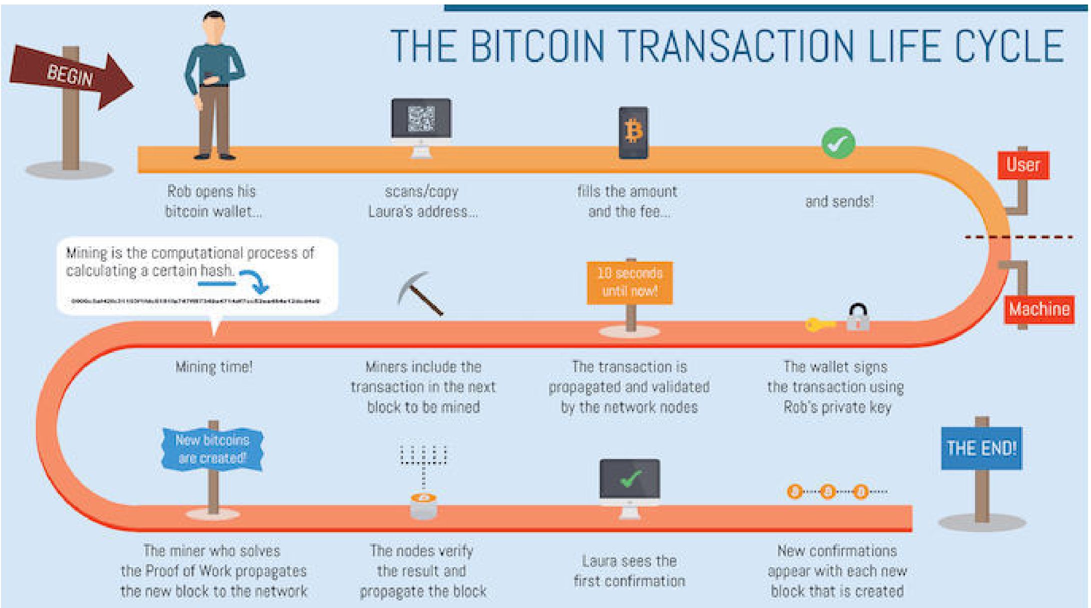

Figure 2 – Bitcoin lifecycle (ANKALKOTI; SANTHOSH, 2017).

and exchange, including every area of finance, economics and money". A shorter and simplistic definition would be a technology to store information in a de-centralized and secure way.

The implementation of blockchain for currency is known as "blockchain 1.0". The second version is known as the implementation for the so called "smart contracts" where the capabilities are extended beyond the simple money transactions (e.g. stocks, bonds, derivatives). The third and more recent version is taking the application even further, surpassing the currency, finance and markets applications, jumping into areas like health, science and government.

Common knowledge says that two key points are mandatory for a blockchain to exist:

- It should never have a single point of failure.
- Must be de-centralized, or in other words, cannot be controlled by a single entity or node.

2.3.2 Bitcoin

Bitcoin is the most well known example of cryptocurrency, and also is the first one, defined by Nakamoto (NAKAMOTO, 2008) as a "a system for electronic transactions without relying on trust". Nakamoto also define it as a framework of coins made from digital signatures, which provides strong control of ownership, with a peer-to-peer network using proof-of-work to record a public history of transactions that quickly becomes

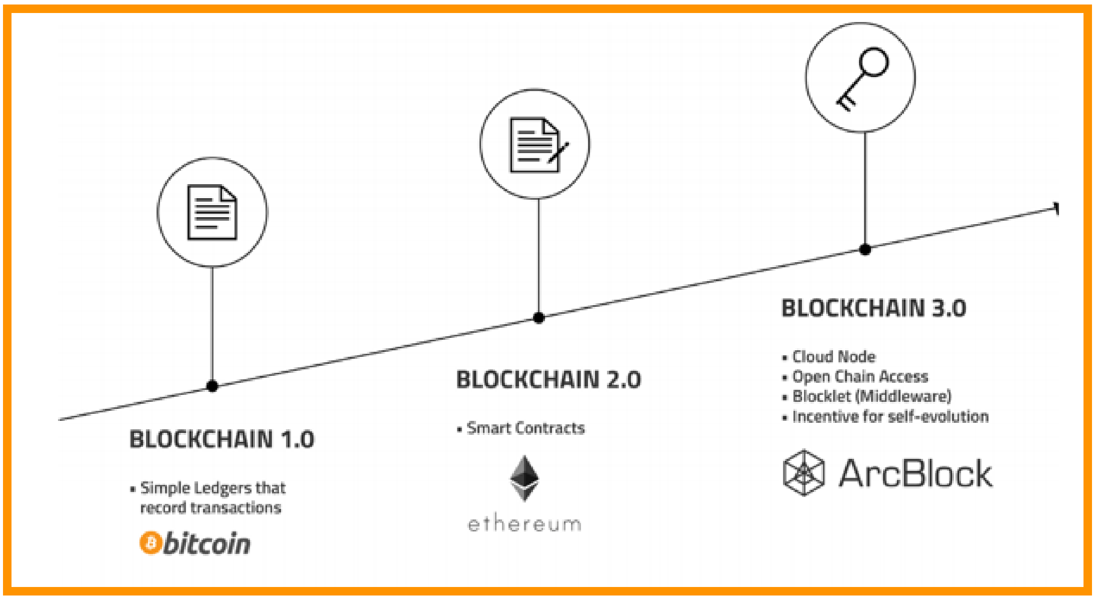

Figure 3 – Blockchain generations (KHATWANI, 2018).

computationally impractical for an attacker to change if honest nodes control a majority of CPU power, being that the mechanism to prevent double spending.

Kroll (KROLL; DAVEY; FELTEN, 2013) gives a "lower" level definition as follows: A Bitcoin is a fixed-value cryptographic object represented as a chain of digital signatures over the transactions in which the coin was used. A coin can be checked for validity simply by checking the cryptographic validity of the signatures that constitute its history.

Each Bitcoin is owned by a Bitcoin address, which consists of a public key. The owner of a Bitcoin (that is, the holder of the corresponding private key) can create a transaction (acting as the sender) by signing an assertion that Bitcoins are being transferred from one address to another. A transaction may involve many input identities and many output identities. Occasionally an extra output value will appear in a transaction for change to transfer back to the sender, since fixed-value coins must be transferred in an all-or-nothing manner. If the total value of the input Bitcoins exceeds the value of the output Bitcoins, the difference is interpreted as a transaction fee, which is paid to the player who successfully appends that transaction to the blockchain.

The Bitcoin distributed ledger, implemented through blockchain, provides perhaps the most robust transaction data set in history (PETERSON, 2017). Even though this dissertation shall not look exclusively to Bitcoin, most cryptocurrencies work in a very similar way to it, and therefore the understanding of its mechanics shall be helpful going forward.

Bitcoin is quite normally referred as a FIAT currency, but that is not entirely correct as explained in the Figure 4. Its a fact that Bitcoin, as any other FIAT currency does not have intrinsic value, but the similarities stop close to that, where Bitcoin is not issued or backed by a Government.

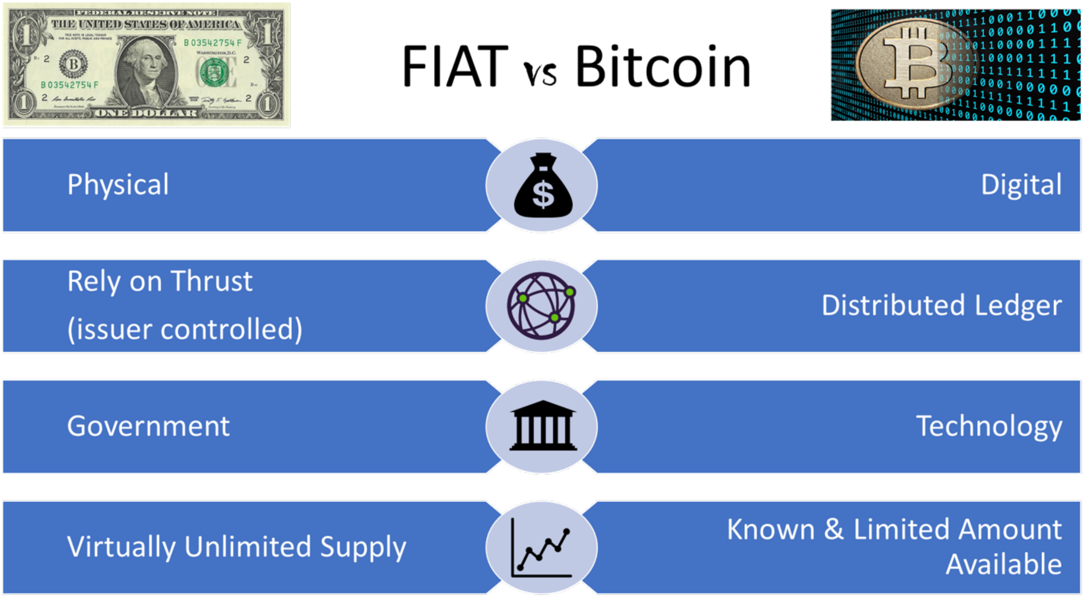

Figure 4 – Bitcoin versus Fiat Money.

2.4 Sentiment Analysis

The following excerpt from Hutto (HUTTO; GILBERT, 2014) defines in a very summarized way the sentiment analysis task and why the lexicon is a critical part of it:

> Sentiment analysis, or opinion mining, is an active area of study in the field of natural language processing that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions via the computational treatment of subjectivity in text. (...) an endeavor (of reviewing all the literature on the Sentiment Analysis subject) would not be possible within the limited space available (such treatments are available in Liu (LIU, 2012) and Pang & Lee (PANG; LEE, 2008))
>
> (...)
>
> A substantial number of sentiment analysis approaches rely greatly on an underlying sentiment (or opinion) lexicon. A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative (LIU, 2010). Manually creating and validating such lists of opinion-bearing features, while being among the most robust methods for generating reliable sentiment lexicons, is also one of the most time-consuming. For this reason, much of the applied research leveraging sentiment analysis relies heavily on preexisting manually constructed lexicons.

Every time you pick up a newspaper and start reading the business section, your brain automatically starts, in some ways, calculating a sentiment score that composes for you a feeling if what you just read was positive, negative or whatever. This is what sentiment and opinion mining is all about, the ability to derive such emotion understanding from written text. The collection of evaluations you take from a newspaper will compose your opinion in the moment you decide or not to buy a house or invest your money for example. its understandable that the same would happen with social media, specially when we are talking about a digital asset, that demands some sort of tech understanding to get someone interested into it.

3 Methodology

Cryptocurrencies are a fairly new phenomenon and the research on the topic shown itself so far very limited. As the popularity grows, so does the number of articles, some with doubtful quality. There is yet some skepticism from the market and researchers around the possibility of it being a investment bubble that will blow some time from now or if we are facing a real potential disruption in the financial markets that came to stay. Having that said, besides looking for formal articles, this phase review will have to cover specialists opinions and listen to the voice of the real players of the cryptocurrency world.

During the development of this dissertation we explored ARIMA and Network laws performance over the price of four selected cryptocurrencies:

- Bitcoin
- Ether
- Litecoin
- Bitcoin Cash

The selection of these assets took into consideration the availability of the data and presence in the market.

This dissertation is conducting a comparative study with multiple cryptoassets, using known methods as well as some experimental theories. The dependent variable on our work will be the price of the cryptocurrencies selected for the study and the independent variables will be the variables that compose the models being tested (see figure 5).

3.1 Data

For this work, the following sources of data were used:

- **Coinbase.com:** Founded in June of 2012, Coinbase is a digital currency wallet and platform where merchants and consumers can transact with new digital currencies like Bitcoin, Ether(eum), and Litecoin.
- **bitinfocharts.com:** Public platform of statistics, with no paid versions, log-ins or whatsoever, regarded as reliable by the community where you can get up-to-date information of how top cryptocurrencies are performing in the market making it easy make informed decisions.

Figure 5 – Cause (Dependent Variable) and Effects (Independent Variables).

- **blockchain.info (now blockchain.com** Blockchain.info was launched in August 2011. The service provides data on recent transactions, mined blocks in the Bitcoin blockchain, charts on the Bitcoin economy, and statistics and resources for developers.
- **Reddit.com:** Social network founded in 2005, self defined as "home to thousands of communities, endless conversation, and authentic human connection. Whether you're into breaking news, sports, TV fan theories, or a never-ending stream of the internet's cutest animals, there's a community on Reddit for you."

3.1.1 Cryptocurrencies Data

Data was extracted using the available APIs and tools from the already mentioned websites.

A dataset was created for each currency containing:

1. **Date:** Current date.
2. **Open Price in US Dollars:** Token price valuated at the market open.
3. **High Price in US Dollars:** Highest token price valuated at that specific day.
4. **Low Price in US Dollars:** Lowest token price valuated at that specific day.
5. **Close Price in US Dollars:** Last token valuated price of the day.
6. **Volume of transactions:** The estimated transaction value in US Dollars or any other currency, also to be used as a proxy of number of node. Although most websites

acknowledge that this information is hard to estimate because the impossibility to measure transactions outside the exchange.

7. **Market Cap:** Total amount of cryptoasset transacted on a certain day, multiplied by an average of the parity Crypto/USD from the main exchanges available. (This is our main equivalent to Network Value for a Blockchain)

8. **Number unique addresses transacting in a day :** The number unique of wallet addresses, or, in other words the number of nodes in the network operating in a specific date.

Assumptions made over data:

- Both the volume and the daily unique addresses could be used as a proxy for number of nodes in the network. After assessing the quality of both indicators we decided to move forward using the daily number of addresses. Such metric will be the foundation when applying Metcalfe law or any variation of network valuation law (METCALFE, 2013).

- The curve of the price, as it happens in the currency market, is assumed to consist of bubbles and bursts (ALABI, 2017) and in order to have a reliable model, such noise must be filtered by using known statistics techniques (e.g. moving averages, exponentiation and logarithmic scales to smooth the lines and facilitate visualization).

- The growth of the network is assumed to begin from the point of critical mass, as stated by (ALABI, 2017). Its observed in the initial values of the data set some values that are zero or very close to zero. To avoid over-fitting, whenever necessary, these records are filtered out.

Statistics on the collected data:

- Bitcoin: 2192 records starting in 28 April 2013 up until 28 April 2019.
- Ether: 1361 records starting in 07 August 2015 up until 28 April 2019.
- Litecoin: 2192 records starting in 28 April 2013 up until 28 April 2019.
- Bitcoin Cash: 645 records starting in 23 July 2017 up until 28 April 2019.

At any moment, missing data in a column without a visible distortion in the sequence was handled with linear interpolation impacting only Volume values in less than 1% of the records.

3.1.2 Social Media Data

Data was extracted using the available APIs and tools from the already mentioned websites.

1. **id:** Unique identifier of the postage within the social network.
2. **created:** Date and time the postage was created.
3. **subreddit:** Which community that postage was placed.
4. **title:** Title of the postage.
5. **body:** Body of the postage (normally just a link or image).
6. **comms_num:** Number of comments made under that postage.
7. **ups:** Number of upvotes used to keep that post on the top of the list.
8. **downs:** Number of downvotes used to push that post down on the list.

We originally started with the idea of using, Twitter micro-blog but limitations of the API and a loss of a hard drive during the course of this development led us to use Reddit instead. This is more of an information rather than a downside. Reddit is regarded as the frontpage of the internet, and known place for many of the key cryptocurrencies advocates around the world.

Data on social network is somewhat tricky to get specially in times that data protection and data ownership are subjects taken more and more into consideration. Reddit offered some ease to get the data because its already provided without poster information. Even with that there are some limitations on how deep in the past you can go.

Our data set compiled 47065 posts that were used further in the development of this dissertation.

Data was gathered, cleansed, any non English and non textual posts were removed, and the final data set ended up with 38119 posts.

Vader lexicon (HUTTO; GILBERT, 2014) was used to generate sentiment analysis metrics, in contrast to what was done by Kaminski (KAMINSKI, 2014), who used search strings with positive or negative words and counted the results (e.g. ”Bitcoin AND feel OR happy OR great OR love OR awesome OR lucky OR good OR sad OR bad OR upset OR unhappy OR nervous -bot” and ”Bitcoin AND hope OR fear OR worry”).

The posts were extracted from the following subset of communities, which are commonly used by investors from which the mood, we infer, can cause an impact in the cryptocurrencies price:

- r/Bitcoin
- r/BitcoinMarkets
- r/btc
- r/CanadianInvestor
- r/CryptoCurrency
- r/dashpay
- r/Daytrading
- r/dividends
- r/economy
- r/ethereum
- r/EthereumClassic
- r/ethtrader
- r/finance
- r/financial
- r/FinancialPlanning
- r/Forex
- r/investing
- r/investing_discussion
- r/InvestmentClub
- r/litecoin
- r/news
- r/options
- r/pennystocks
- r/personalfinance
- r/portfolios
- r/Ripple
- r/SecurityAnalysis
- r/StockMarket
- r/stocks
- r/thewallstreet
- r/ValueInvesting
- r/wallstreetbets
- r/worldnews

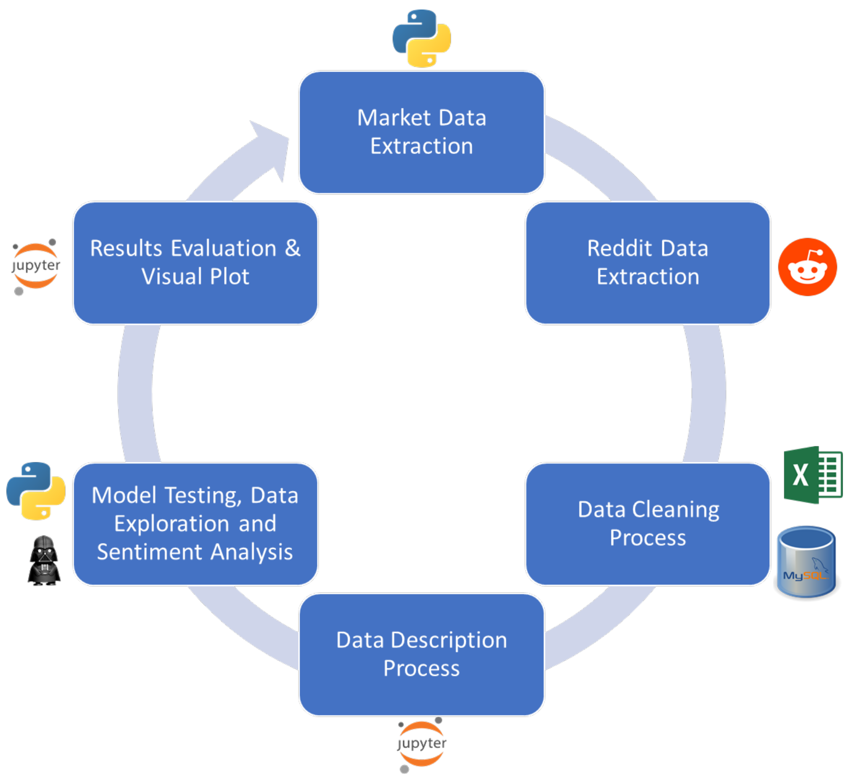

Figure 6 – Data processing cycle

3.1.3 Language and Data Storage

Python was the main language used to develop the models found in this dissertation. The main packages used were:

- **NumPy:** Algebra.
- **Pandas:** Data manipulation.
- **Matplotlib:** Charts.
- **Seaborn:** Charts.
- **Scipy:** Statistical models and regression.
- **SciKit Learn:** Statistical models and regression.
- **NLTK:** Handle natural language and sentiment analysis.

Data storage, due to the low volume was mainly text files and Excel spreadsheets.

3.2 Experimentation

3.2.1 Sentiment Analysis

Valence Aware Dictionary for sEntiment Analysis (VADER) is the gold standard lexicon used to produce the information about sentiment analysis.

The execution was conducted be done using the Python implementation of VADER that is deployed with the Natural Language Toolkit (NLTK).

The process followed for the sentiment analysis piece is described in the figure 7

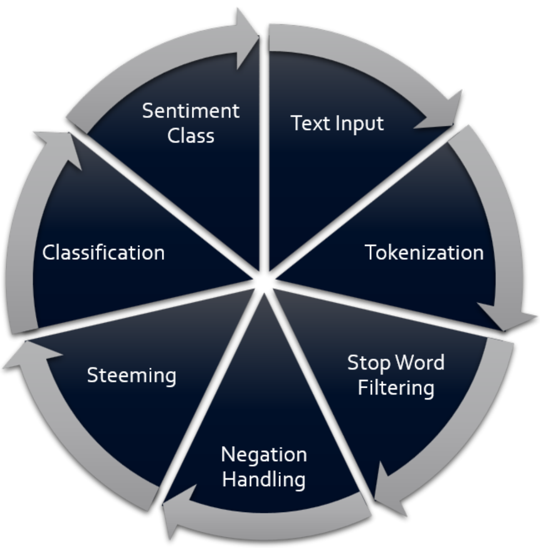

Figure 7 – Sentiment Analysis Process Cycle

3.2.2 Models and Data Exploration

During the course of the development an extensive descriptive analysis was conducted in the datasets produced. Plots of the data will be presented in the upcoming sections.

Due to the fact that some of the numbers we are dealing with are in the billions or even trillions scale, as its common in econometrics we uses fairly frequently visualizations of data in logarithmic scale. Every time we use log scale, we refer to the natural logarithm.

Initial investigations used the Close price for each cryptocurrency and evaluated the ARMA(p,q) model, also known as Box and Jenkings (BOX; JENKINS, 1976) with the formula described bellow, where $y_t, \phi 1, \phi 2..., \theta 1, \theta 2...$ are as defined previously:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

The ARIMA model introduces to the above formula one parameter of differentiation to handle non stationary time series, but uses the same formula structure.

We proceeded to explore a number of alternatives to turn the input data stationary and differentiation showed the best results, in our case with a 7 days lagged version of the variable. Methods used to identify the ARIMA p and q parameters are discussed in the analysis section of this document.

In a second moment we took the same dataset and explored the Metcalfe law using the original formula and the variations plotted bellow, to mention Metcalfe (METCALFE, 2013), Odlyzko (BRISCOE; ODLYZKO; TILLY, 2006) and Generalized Metcalfe proposed in an article by Clearblocks (CLEARBLOCKS, 2018). These formulas are defined as follows:

$$Metcalfe = a * n^2$$

$$Odlyzko = a * n * log(n)$$

$$MetcalfeGeneralized = a * n^{1.5}$$

For the purpose of this work lets consider "a" as a proportionality constant (Metcalfe's coefficient), obtained by fitting the formula to the data, and "n" being the number of nodes in the network, for which we used the number of Daily Active Addresses (DAA) as a proxy.

Metcalfe formula models the network value, in our case the Close Price or the Market Cap (both highly correlated in every observation), but in order to forecast we should be able to predict how the number of users grow within the studied blockchain.

We briefly mention that Metcalfe proposed in his article (METCALFE, 2013) a generalization of the sigmoid function, denominated netoid, in order to model the growth of a network in function of time. The formula is described as follows:

$$N(t) = \frac{p}{1 + e^{(-v*(t-h)))}}$$

Where p is the peak population, t is the moment in time, v is the virality (how fast is the adoption) and h is the center of the sigmoid curve (growth peak).

We cover in the next section more details about the issues of the applicability of it to the existing data. To summarize, by the time Alabi (ALABI, 2017) wrote his article, the blockchains investigated seemed to be yet to reach a point of stagnating value. This reality seems to persist even now that we are nearly 2 years later.

Metcalfe's law and network models are the main motivators of this dissertation, and therefore we also explored a few derivations of the laws and potential metrics being explored by the market.

Lastly we explore the sentiment analysis data together with the datasets produced and at this stage we seek for correlation that support pilot tests conducted prior to this dissertation. For that we aim to force an impact of the identified sentiment, positive or negative, of the day and introduce an error factor to the network model formula, pushing the predicted value up or down respectively.

The data explored shows a huge variation in the course of time. Such uncertainty brings problem when dealing with the model, so for every model executed we transformed the data to remove this stochastic component. To do so we used moving averages in different time spans.

Here are the tests we performed and discuss results in the upcoming section:

1. ARIMA.

2. Metcalfe for value prediction.

3. Metcalfe for bubble prediction.

4. Metcalfe + Sentiment Analysis.

3.3 Evaluation

A great number of metrics commonly used to measure accuracy of forecasting are listed in the article by De Gooijer (De Gooijer; HYNDMAN, 2006). From the papers evaluated the most popular is undoubtedly the Mean Squared Error (MSE) for the type of research proposed here, therefore, unless a better reason to change is found during the evolution of this work, it will be used.

The figure 8 shows the available metrics.

Commonly used forecast accuracy measures

MSE	Mean squared error	$=\text{mean}(e_t^2)$
RMSE	Root mean squared error	$=\sqrt{\text{MSE}}$
MAE Mean	Absolute error	$=\text{mean}(\lvert e_t \rvert)$
MdAE	Median absolute error	$=\text{median}(\lvert e_t \rvert)$
MAPE	Mean absolute percentage error	$=\text{mean}(\lvert p_t \rvert)$
MdAPE	Median absolute percentage error	$=\text{median}(\lvert p_t \rvert)$
sMAPE	Symmetric mean absolute percentage error	$=\text{mean}(2\lvert Y_t - F_t \rvert/(Y_t+F_t))$
sMdAPE	Symmetric median absolute percentage error	$=\text{median}(2\lvert Y_t - F_t \rvert/(Y_t+F_t))$
MRAE	Mean relative absolute error	$=\text{mean}(\lvert r_t \rvert)$
MdRAE	Median relative absolute error	$=\text{median}(\lvert r_t \rvert)$
GMRAE	Geometric mean relative absolute error	$=\text{gmean}(\lvert r_t \rvert)$
RelMAE	Relative mean absolute error	$=\text{MAE}/\text{MAE}_b$
RelRMSE	Relative root mean squared error	$=\text{RMSE}/\text{RMSE}_b$
LMR	Log mean squared error ratio	$=\log(\text{RelMSE})$
PB	Percentage better	$=100\ \text{mean}(I\{\lvert r_t \rvert<1\})$
PB(MAE)	Percentage better (MAE)	$=100\ \text{mean}(I\{\text{MAE}<\text{MAE}_b\})$
PB(MSE)	Percentage better (MSE)	$=100\ \text{mean}(I\{\text{MSE}<\text{MSE}_b\})$

Here $I\{\text{u}\}=1$ if u is true and 0 otherwise.

Figure 8 – Commonly used measures of accuracy by (De Gooijer; HYNDMAN, 2006)

Existing works also provide a great deal of information on how these models should be evaluated. Since the objective here is obviously compare the results and identify causal relations, the usage of the same techniques is something expected. For Cryptocurrencies and Metcalfe's law the results can be compared to Peterson (PETERSON, 2017), Alabi (ALABI, 2017) or Van Vliet (Van Vliet, 2018) among others. For ARIMA, one of the many examples can be Tlegenova (TLEGENOVA, 2015) and Azari (AZARI, 2018).

4 Analysis

4.1 General considerations

As explained in the previous section, this study was conducted by firstly evaluating the statistical and Metcalfe formulas to model and forecast cryptocurrencies independently, then the introduction of sentiment analysis is attempted to improve the existing results. Lately the comparative results will be discussed together.

The currencies initially selected for this study are Bitcoin, Ether, Litecoin and Bitcoin Cash.

1. Bitcoin: Had its genesis block mined in the network in January 3rd 2009 by Satoshi Nakamoto, is monitored under the symbol BTC, was the first cyptocurrency to be launched and have a hard cap of 21 MM bitcoins, of which, by the time this dissertation was being written, 17 MM was already mined in circulation.
2. Ether: Lauched in 2014 by Vitalik Buterin, is monitored unther the symbol ETH.
3. Litecoin: Was an early bitcoin spin off launched in 2011 by Charlie Lee.
4. Bitcoin Cash: Most recent of the evaluated cryptos, its a recent spin off from the Bitcoin, started when a hard fork took place in Bitcoin blockchain in August 2017. It also have a hard cap of 21 MM bitcoins, of which, by the time this dissertation was being written, also 17 MM was already mined in circulation.

Before we move further on the analysis, the following conventions must be made:

1. Close Price: refers to the close price of the cryptoasset being evaluated in United States Dollar.
2. Market Cap: refers to the total amount of cryptoasset transacted on a certain day, multiplied by an average of the parity Crypto/USD from the main exchanges available. The exchanges are not revealed by the data provider and therefore we have to make an assumption that this information is correct to move forward.
3. Unique Addresses: refers to the total number of unique wallet addresses operating on that blockchain per day. Since we don‘t have a reliable source for the amount traded off chain (outside the exchanges), this will then be used as a proxy for the number of nodes in the network. This may be refered as Daily Active Addresses (DAA) or Daily Unique Addresses (DUA) as well.

The imported data from the 4 cryptocurrencies being evaluated can be seen in the figures 9, 10, 11 and 12.

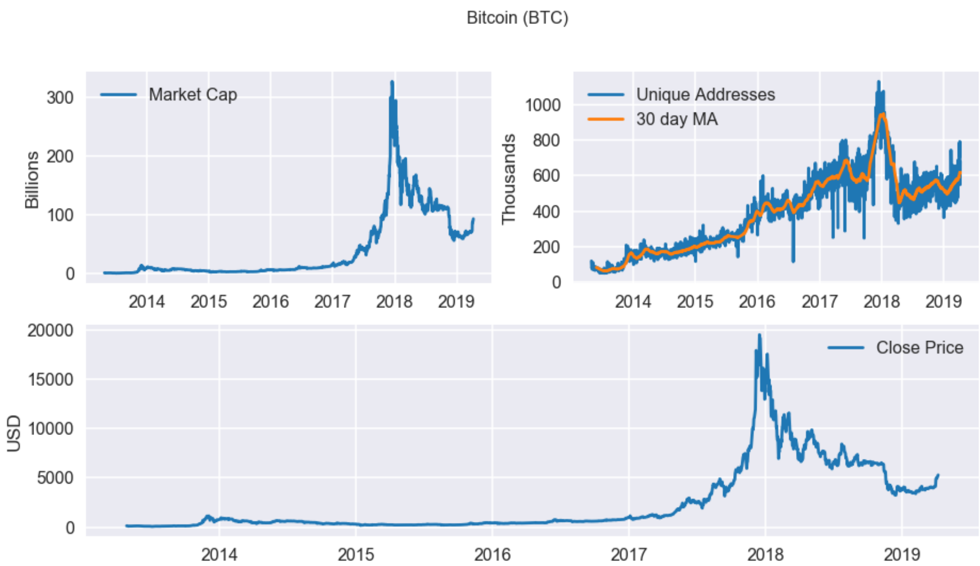

Figure 9 – Bitcoin Data

Source: Author(2019), with data extracted from https://bitinfocharts.com/

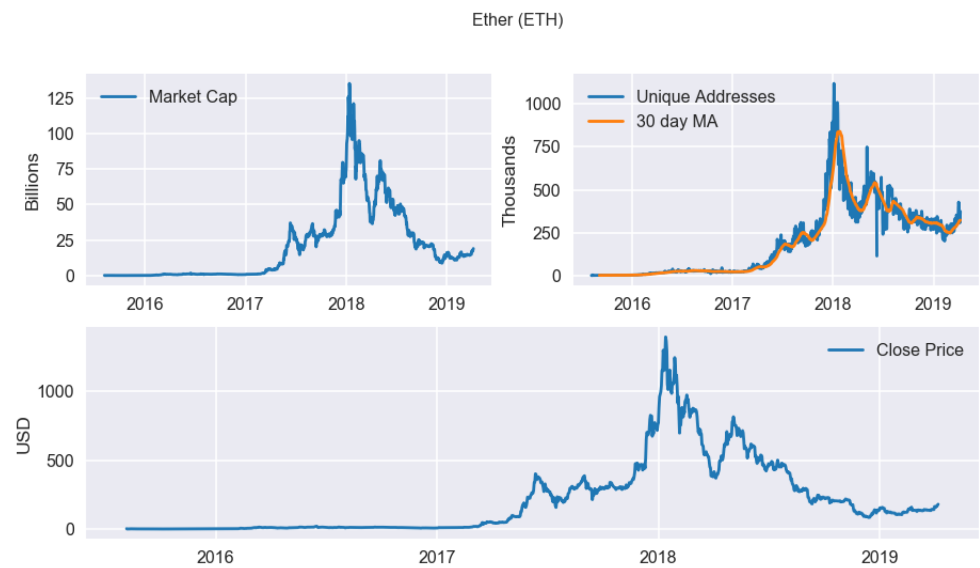

Figure 10 – Ether Data

Source: Author(2019), with data extracted from https://bitinfocharts.com/

From this data by itself, a few observations can already be made. All 4 cryptocurrencies selected show a spike between second half of 2017 and first half of 2018. This is

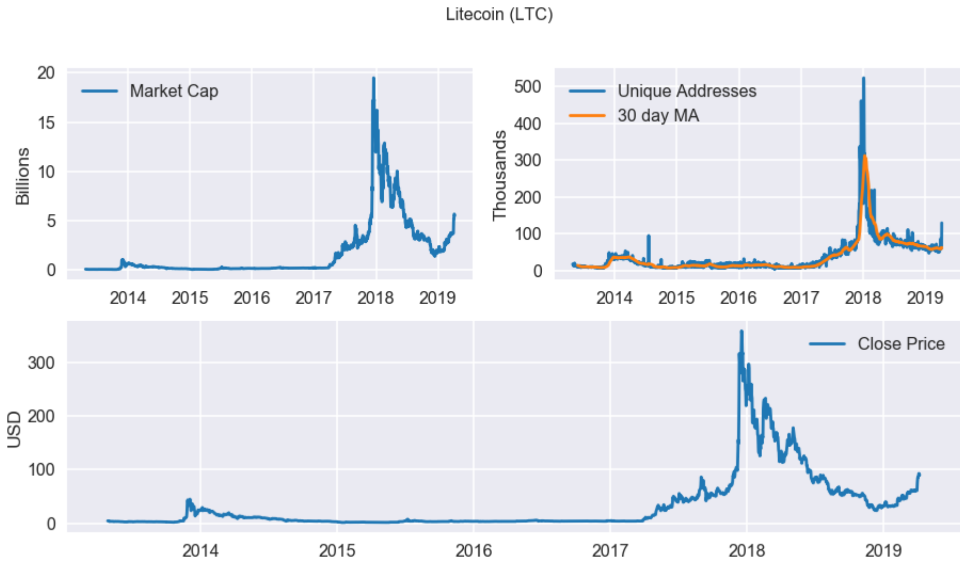

Figure 11 – Litecoin Data

Source: Author(2019), with data extracted from https://bitinfocharts.com/

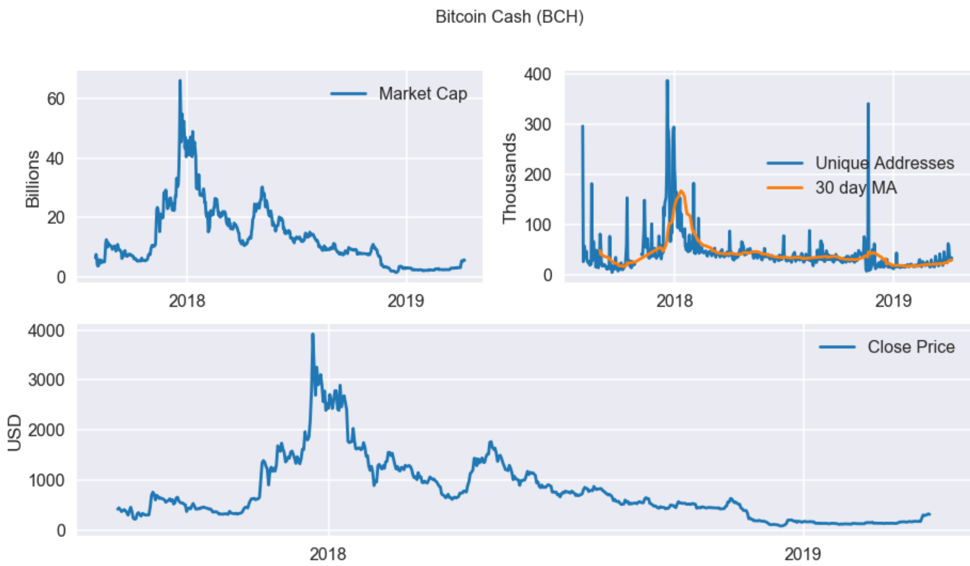

Figure 12 – Bitcoin Cash Data

Source: Author(2019), with data extracted from https://bitinfocharts.com/

present in all indicators observed, and it represents the bubble that happened in the end of 2017 when the cryptocurrencies gained popularity as an investment and speculation asset, and the world observed a run to participate on this market seeking for quick gains, followed closely by some uncertainty from the investor, leading to a proportional correction in the upcoming months.

In the figure 12 you can also observe a spike in the unique addresses by the second half of 2018, which is most likely related to a second hard fork that was announced in that blockchain. Such type of distortion will be eliminated whenever possible, to avoid having a tendency in the forecast.

Although some may argue that Bitcoin is becoming mature, the fact is that none of the blockchains evaluated reached maturity by any means in terms of adoption by the users. To have an order of magnitude to this, the process called "mining" consists of providing processing capacity to the network and receive tokens in exchange for is expected to continue until the last Bitcoin is generated, and this shall not happen until the year 2140. After that, is expected that the network will be self sufficient and there will be no need for more reward in exchange for computer power.

Therefore, for every cryptocurrencies studied here, we will assume that we are in the early phases of adoption. This assumption will be important during the discussion of the Metcalfe law tests, since the proxy for the network value in that case is the number of nodes or users in the network itself.

4.2 ARIMA

ARIMA (BOX; JENKINS, 1976) is a well known model for forecasting time series. Its a generalization of the Auto-Regressive Moving Average (ARMA) model, but with a parameter added to consider an order of differentiation transformation of the time series being analyzed. This was developed to allow the analysis of time series that show trends and are non stationary.

The model by itself is well known and widely used, and can be explored in a number of ways, by changing the input data and looking for results. This entire article could be subject to find the best tweaks to obtain the best ARIMA model to fit the data. This could also lead not only to advantages but to overfitting and and to a number of other problems.

In order to obtain the results for this work, we took one common approach that consists in transforming the data, fitting a model by adopting common sense approaches in defining parameters and comparing the results. The techniques applied here are commonly used by investors, not only for cryptocurrencies but for other investment assets like stocks, derivatives and foreign exchange.

A number of works can be cited when we are looking exclusively for the ARIMA model and its variations like Seasonal ARIMA (SARIMA) among others, seeking for the best ways to improve the models using the most different metrics available. For this section we are mainly inspired in the work conducted by Azari (AZARI, 2018) in his published

Dickey-Fuller Test (p-value)			
	ln(Close Price)	ln(Close Price) - 1 day diff	ln(Close Price) - 7days diff
BTC	0.307828873	0.000000000	0.000152417
ETH	0.014716534	0.000000000	0.000815131
LTC	0.189453532	0.000000000	0.000669862
BCH	0.586798094	0.000000000	0.000014275

Table 1 – Source: Authors (2019)

article, with a few adjustments based on models and techniques widely available and used by investors.

4.2.1 Transformation

The dataset available was split in 2 moments, prior to 2017, where the value of all available cryptocurrencies was near zero and after that when we started seeing adoption. For this step of the study we are considering only from 2017 onward.

We observed that all 3 cryptocurrencies evaluated had some sort of trend visible when plotted. We also performed a Dickey-Fuller test before and after the difference transformation to remove seasonal component. In order to do that we experimented gaps of 1, 7 and 30 days, obtaining best results with 7 days. In other words, we remove seasonality by subtracting current value from the value from 7 days ago.

The results of the stationary test are presented in the table 1 and the plots (figures 13, 14, 15 and 16)show how the data was transformed before and after the process of differentiation.

4.2.2 Parameter Selection

The model selected expects 3 parameters where lies most of the exploration around the tuning and tweaking to make sure the model fits the data correctly and provides good forecasting, adequately representing the original data.

For the purpose of this dissertation we explored 2 methods to select those parameters, being the first the visual evaluation of the Auto Correlation (ACF) and Partial Auto Correlation (PACF) plots of the data, as it can be seen in the figures 17, 19, 18 and 20.

Autocorrelation plots (also known as ACF or the auto correlation function) are a useful visual tool in determining whether a series is stationary. These plots can also help to choose the order parameters for ARIMA model. If the series is correlated with its lags then, generally, there are some trend or seasonal components and therefore its statistical properties are not constant over time.

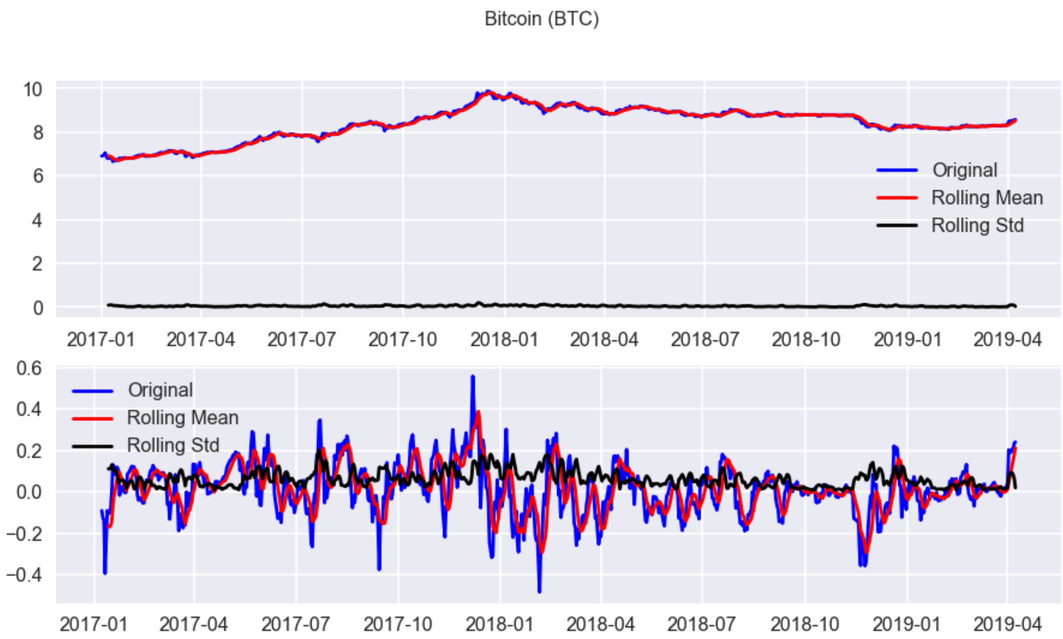

Figure 13 – Bitcoin Dataset Transformation into stationary

Source: Author(2019)

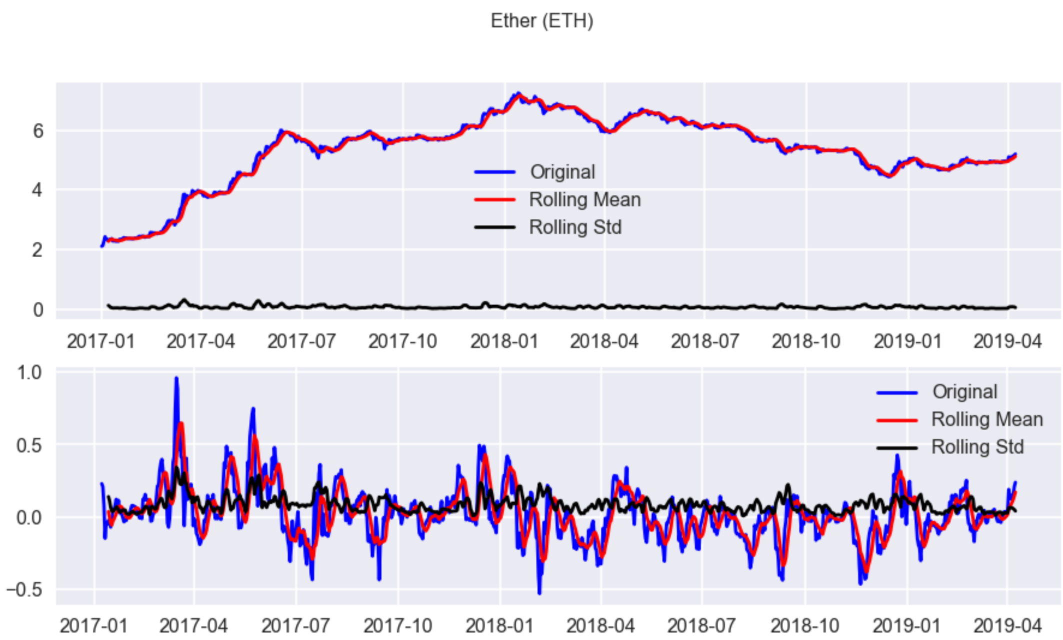

Figure 14 – Ether Dataset Transformation into stationary

Source: Author(2019)

ACF plots display correlation between a series and its lags. In addition to suggesting the order of differencing, ACF plots can help in determining the order of the MA(q) model. Partial autocorrelation plots (PACF), as the name suggests, display correlation between a variable and its lags that is not explained by previous lags. PACF plots are useful when

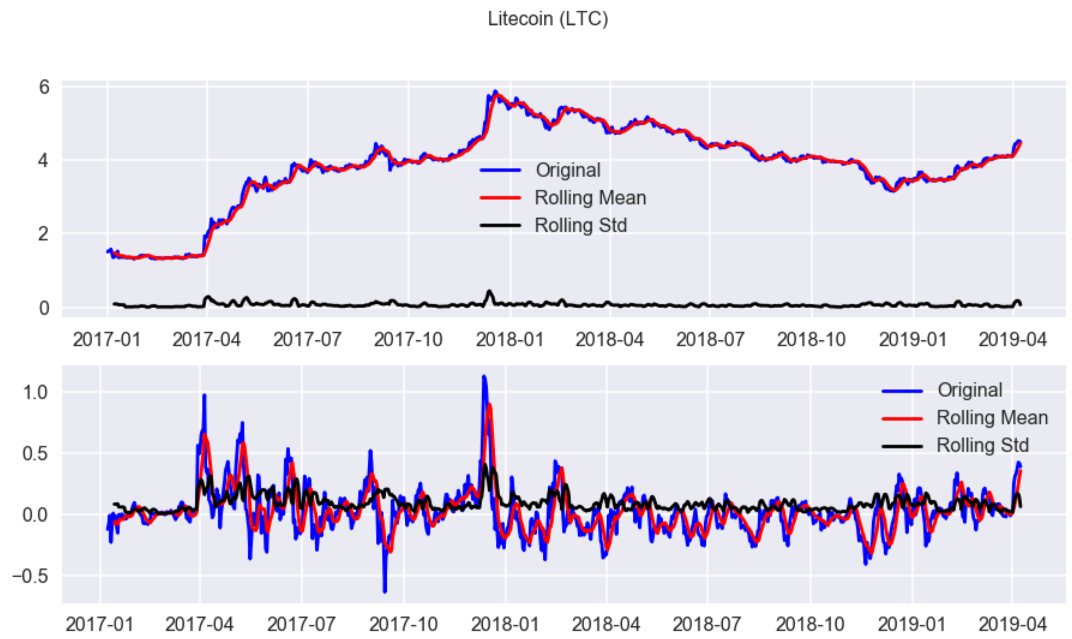

Figure 15 – Litecoin Dataset Transformation into stationary

Source: Author(2019)

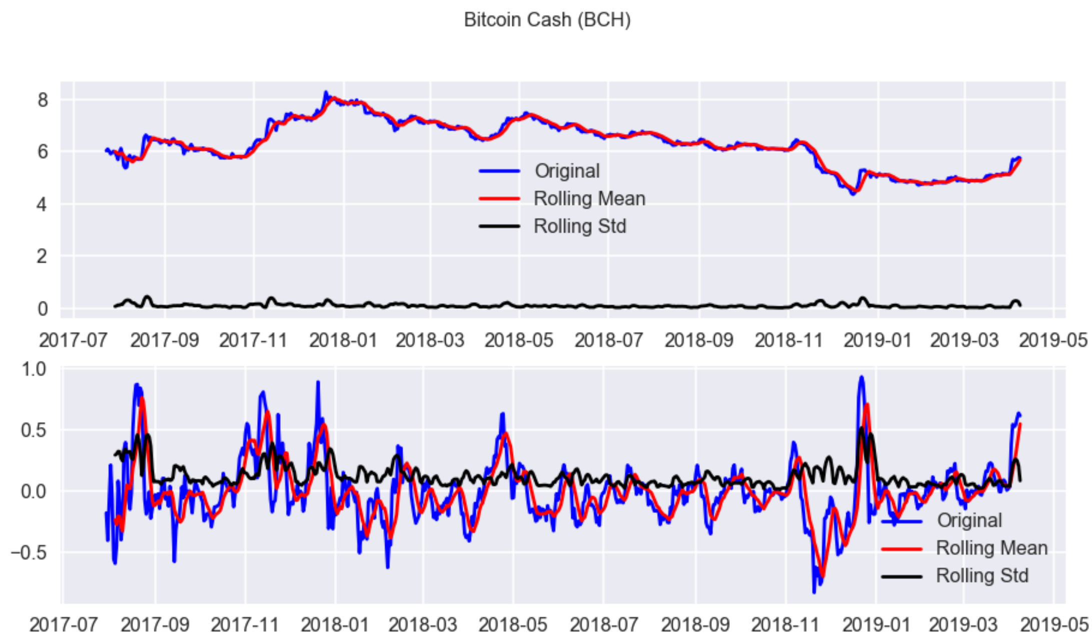

Figure 16 – Bitcoin Cash Dataset Transformation into stationary

Source: Author(2019)

determining the order of the AR(p) model.

For the Auto Regressive (AR) parameter p, we observed in the PACF plot a significant lag on day 2 for all samples. For the Moving Average (MA) parameter q, we observed that for BTC, LTC and BCH it is significant on the 8th lag and for ETH in the 9th lag, observing the ACF plot. These are our first guesses for parameters while we observe still in the PACF plot other possibilities of lag, suggesting that the data may still have some seasonality present.

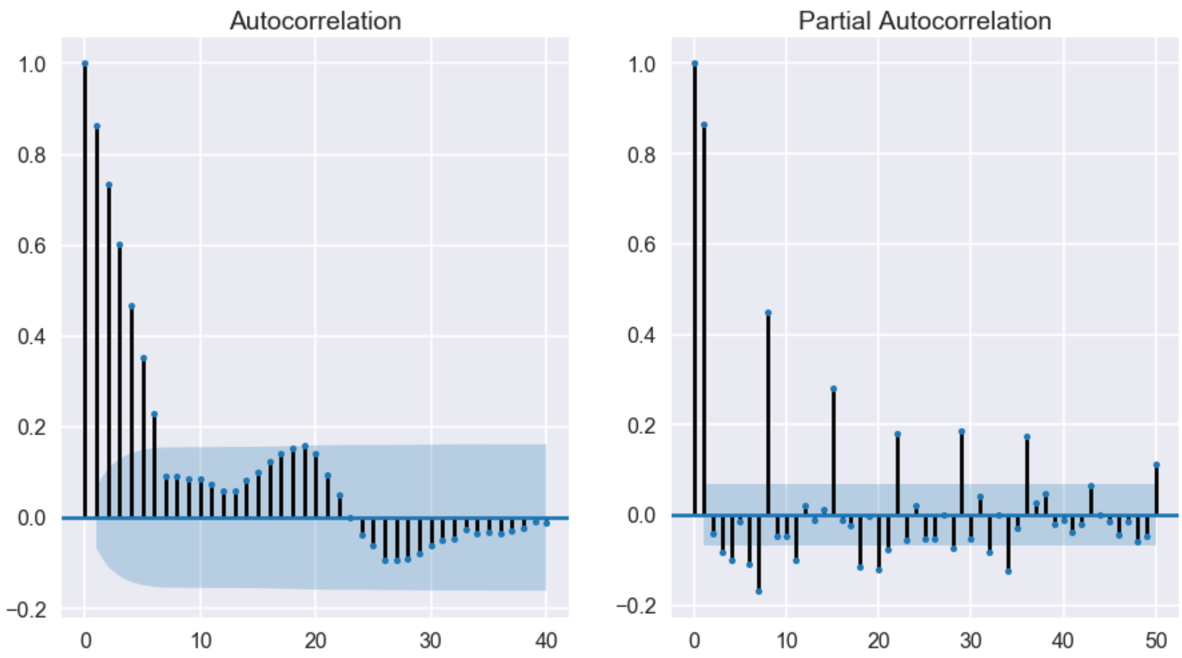

Figure 17 – Bitcoin ACF and PACF

Source: Author(2019)

Besides this visual approach we also used a grid search in order to identify the best set of parameter p, q and d aiming on minimize the RSS of the result. This grid search and the model execution will be presented in next section.

4.2.3 Model Execution

We applied the model to fit the whole data, even aware that due to the huge spike between 2017 and 2018 and fairly different behaviours in two different moments, we might end up with a weak model. That is a risk we had to take in order to proceed with this dissertation and stays as an opportunity for further work to segment the dataset into different moments in time, and train models individually to investigate interactions between those.

Acording to experiments executed by Azari (AZARI, 2018), he observed, when testing the models aiming to minimize the Mean Squared Error (MSE) in different sized windows, randomly generated, for BTC, that the MA (q) parameters increase slowly, and still according to him this is due to the fact that using moving average for prediction requires initialization of the model with random prediction error, and the impact of this

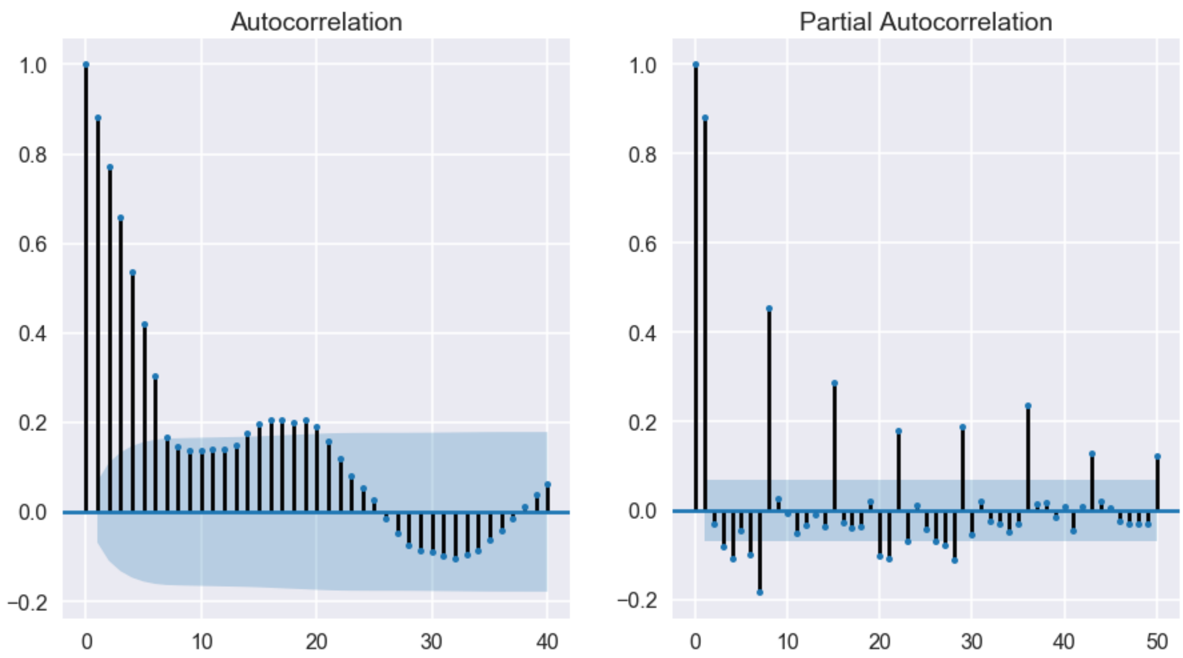

Figure 18 – Ether ACF and PACF

Source: Author(2019)

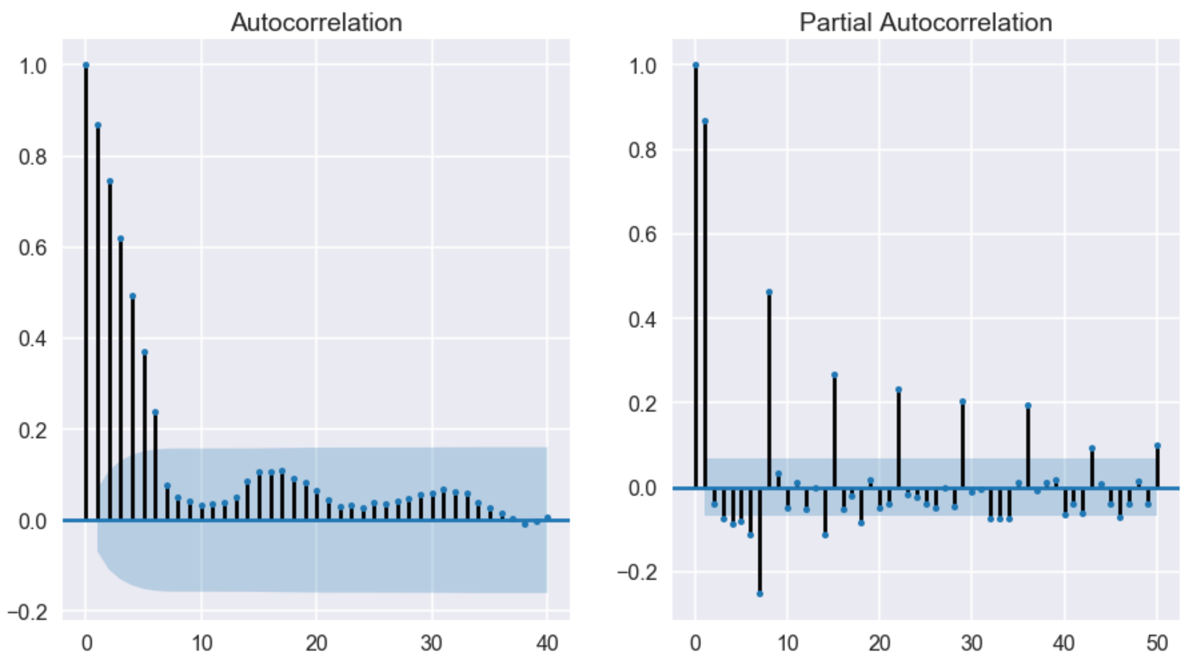

Figure 19 – Litecoin ACF and PACF

Source: Author(2019)

initialization disappear slowly by increasing the window size. On the other hand, one sees that the MSE-minimizing ARIMA model accepts some level of RSS in fitting to the data in order to be able to capture fluctuations in the price.

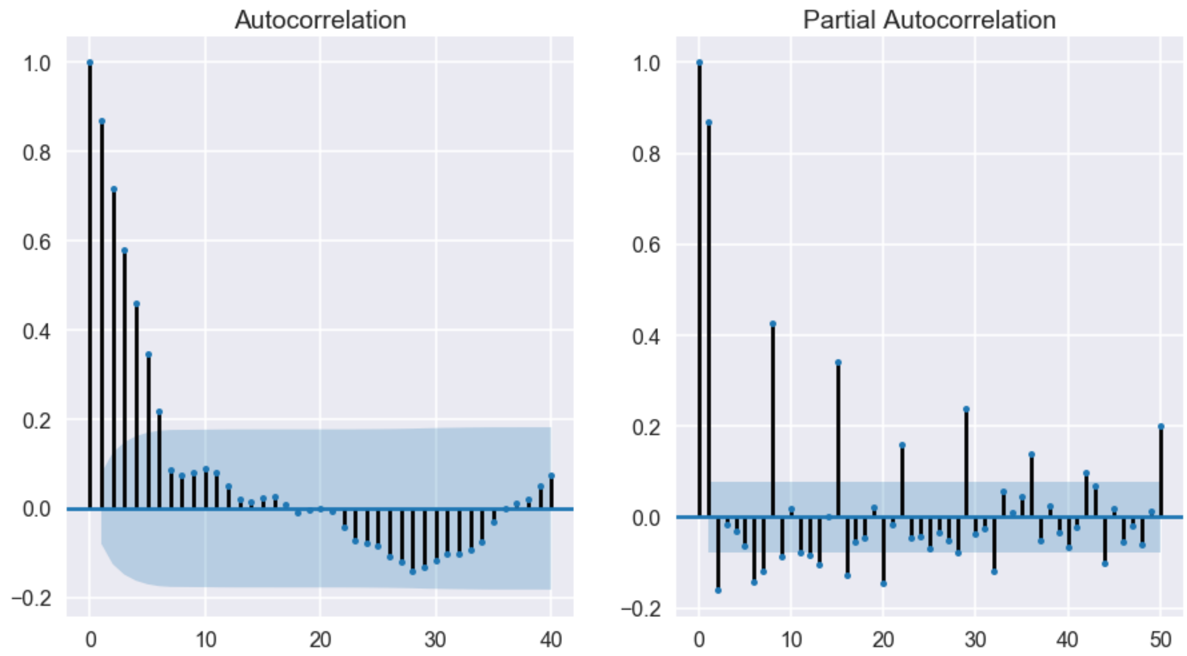

Figure 20 – Bitcoin Cash ACF and PACF

Source: Author(2019)

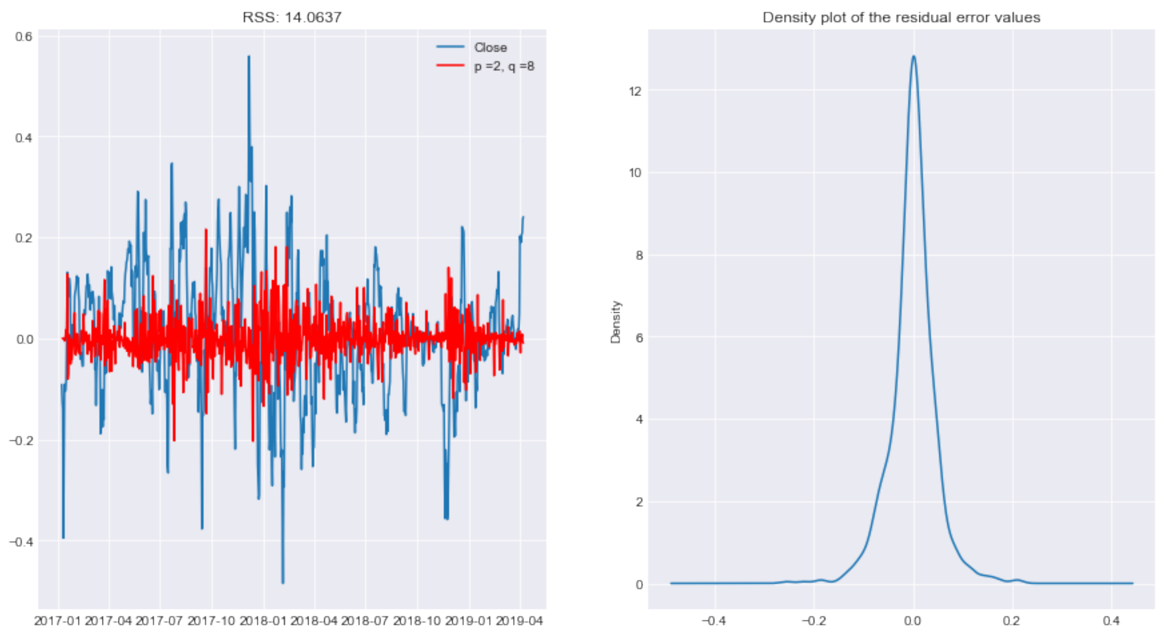

Figure 21 – Bitcoin ARIMA Model (Visual Identified Parameters)

Source: Author(2019)

The plots in the figures 21, 22, 23 and 24 are the executions of the models with the parameters initially identified by visual analysis of the charts ACF and PACF. The output of each model can be found in the appendix B

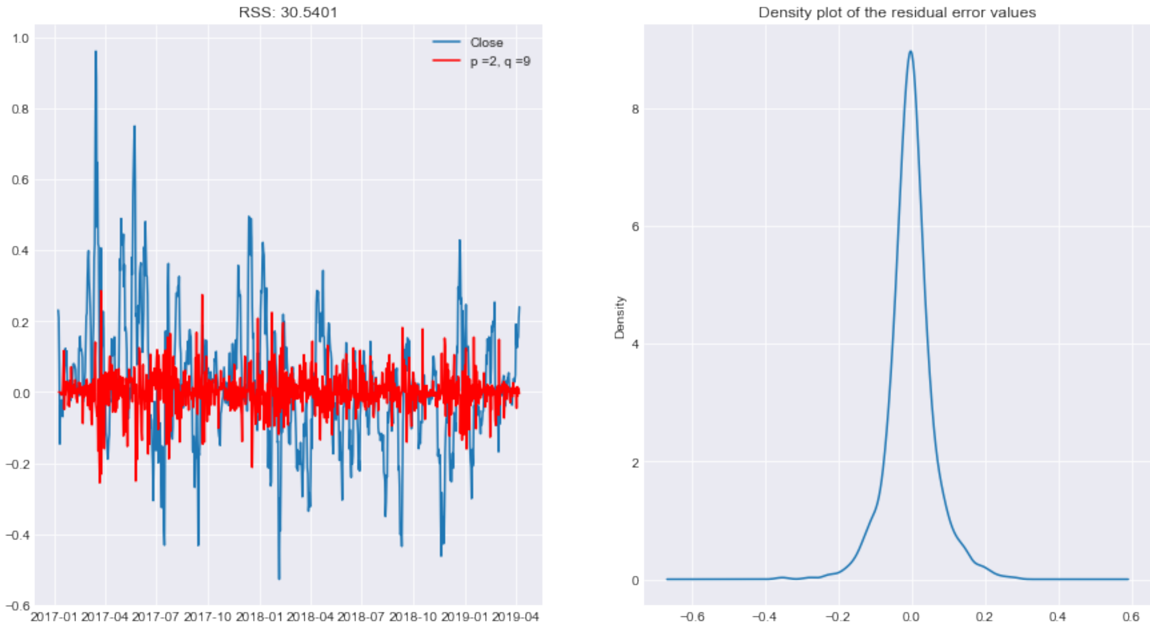

Figure 22 – Ether ARIMA Model (Visual Identified Parameters)

Source: Author(2019)

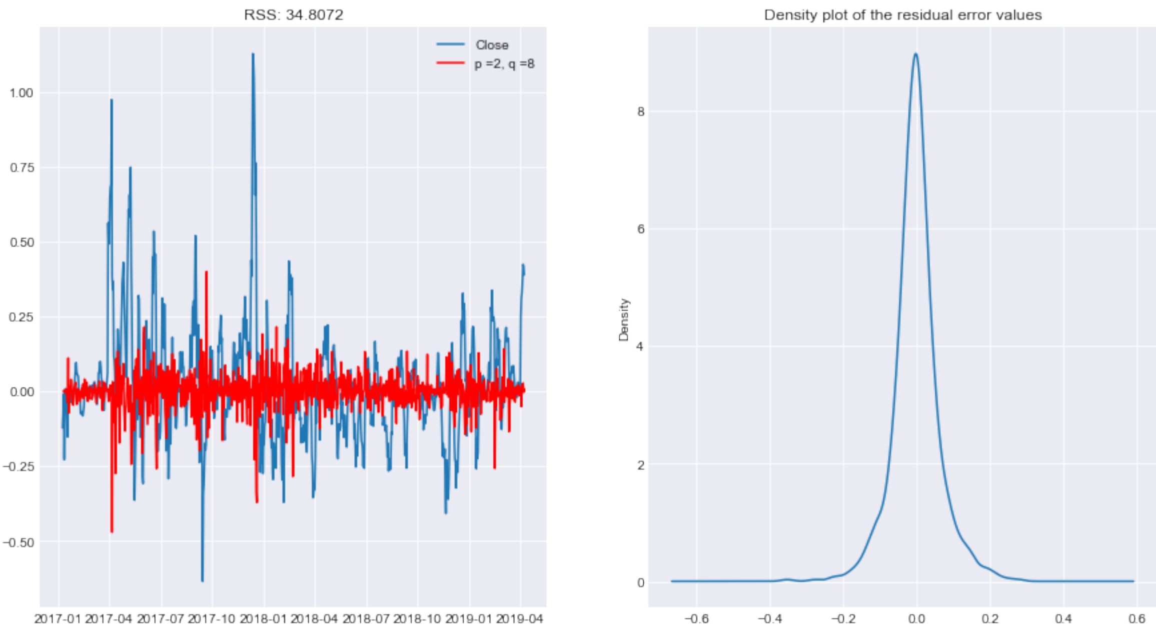

Figure 23 – Litecoin ARIMA Model (Visual Identified Parameters)

Source: Author(2019)

After that, following the same approach from Azari (AZARI, 2018) we ran a grid search finding the following best results with great improvements in the RSS:

- BTC: Best ARIMA(9, 1, 4) RSS=0.002

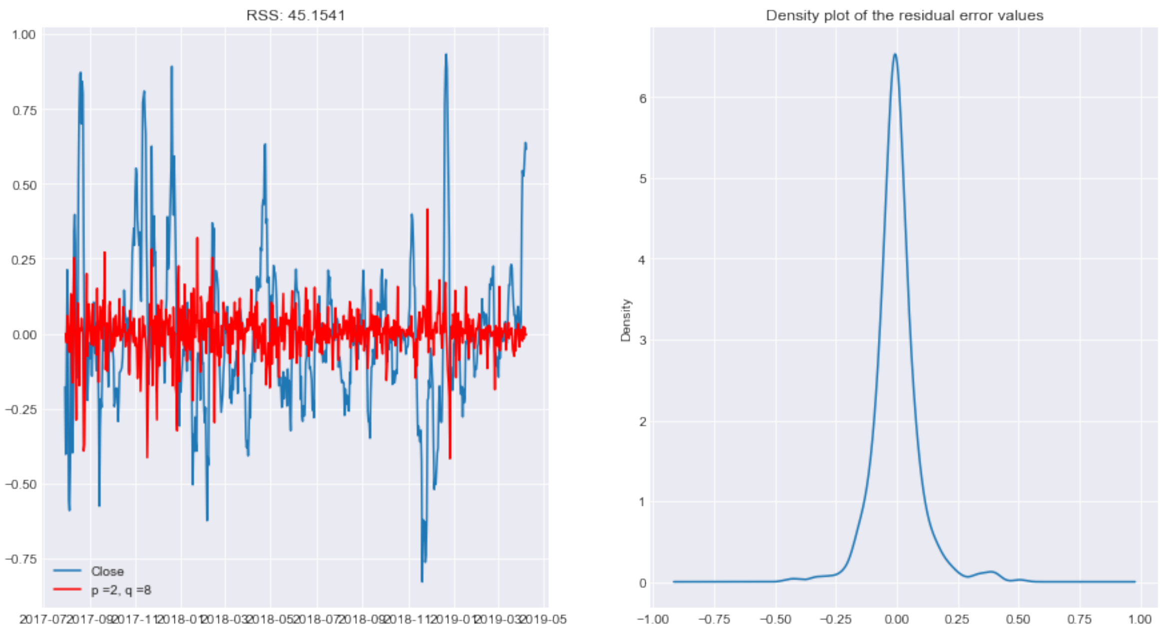

Figure 24 – Bitcoin Cash ARIMA Model (Visual Identified Parameters)

Source: Author(2019)

- ETH: Best ARIMA(6, 1, 6) RSS=0.004
- LTC: Best ARIMA(9, 1, 9) RSS=0.004
- BCH: Best ARIMA(8, 1, 6) RSS=0.007

We also explored with the grid search proposed by the same cited author, where we selected random start locations for estimation window of size k days searching for the best model minimizing the MSE, having achieved consistent results to what he found, for all cryptoassets. Its possible to find better models by exploring multiple time windows, and 10 days presented the best results for the data we used, although, due to the nature of the assets studied, the MSE results were still high.

ARIMA seems to be viable for cryptocurrencies but not on its own. Most likely we have a great opportunity in the future to test combined models and use ARIMA as one of those.

4.3 Metcalfe

In this section we are going to discuss the usage of Metcalfe‘s law and its variations into predicting the value of the cryptocurrency being analyzed.

Worth mentioning that when this work started, the initial idea was to train a model to forecast the cryptocurrency value in a certain point of time, although, during the past year a number of studies and articles were published (mostly outside the academic world though) supporting that instead of predicting the value of the currency itself, the data extracted from the network model was in fact allowing us to see the whenever that asset was moving away from its normal behaviour. An article published at Forbes magazine by Willy Woo (WOO, 2017) supports that in fact the network value was actually predicting (or at best telling the story while it happens) of bubbles in the Bitcoin market.

To understand that we must understand that network models assume the blockchain is a network by itself, and try to value it based on the number of users connected to it. When Metcalfe proposed his law (METCALFE, 2013) by 2013 and Zhang (ZHANG; LIU; XU, 2015) successfully tested it against data from Facebook and Tencent, what they did was basically assume that those companies were valued based on the size of their network, and they were able to fit a model to predict how this value will behave moving forward.

4.3.1 Formula Selection and Model Fit

A lot of study was already dedicated on this topic and several variations of the formula were proposed as it can be seen in charts of this dissertation. Metcalfe‘s law originally proposed that the network value is proportional to the square number of active users, which was interpreted, mainly for Bitcoin, as the number of active addresses in a day (as a proxy).

Kalichkin (KALICHKIN, 2018) defends in one appendix of of his article that is not possible to find one law to rule them all, and in order to prove that he shows a matrix of Metcalfe’s and Odlyzko’s formulas correlation tables against the price. Such data was updated and recalculated for this dissertation and can be found in the figures of the appendix A.

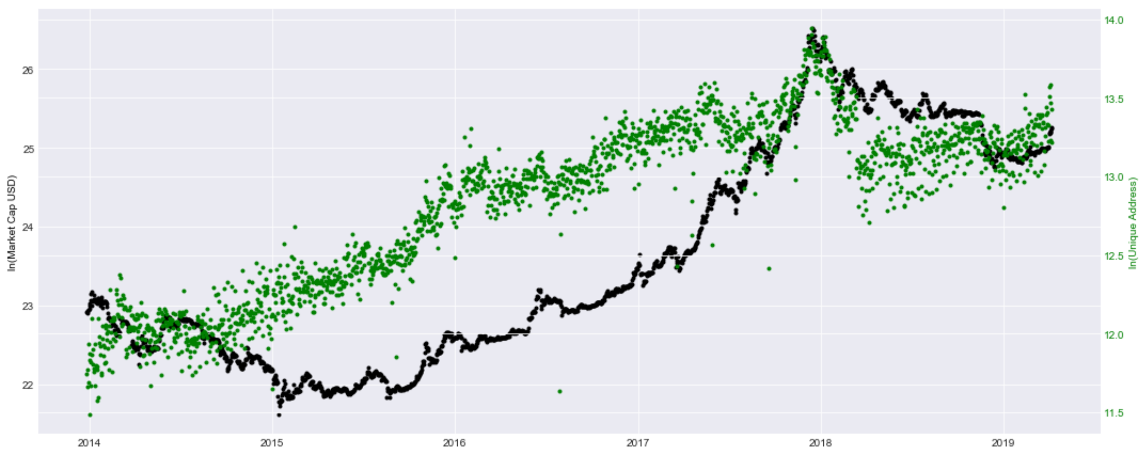

Figure 25 – Bitcoin Market Cap and Daily Active Addresses Overlap

Source: Author(2019)

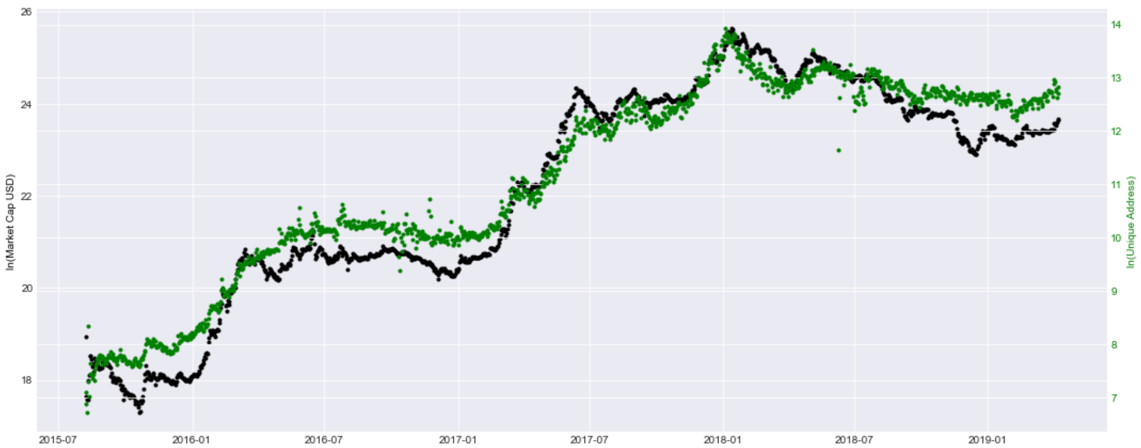

Figure 26 – Ether Market Cap and Daily Active Addresses Overlap

Source: Author(2019)

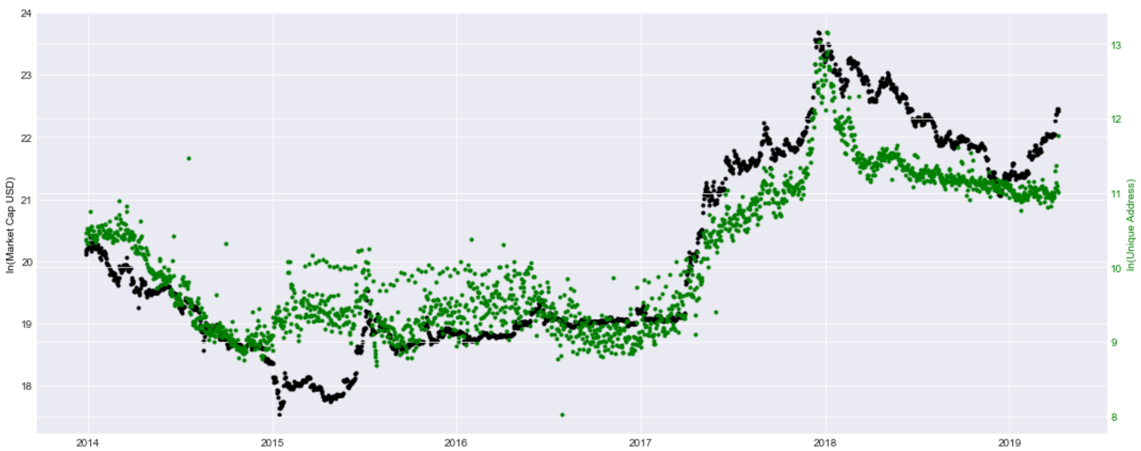

Figure 27 – Litecoin Market Cap and Daily Active Addresses Overlap

Source: Author(2019)

From the data presented on those tables we can extract some bits of information, first being that, compared to other works in the area, Metcalfe law correlation to the main indicators (Close, Market Cap and Volume) is a bit lower here in this work. One of the propositions to explain such behavior comes from Wheatley (WHEATLEY et al., 2018) who proposed that Metcalfe law could instead be predicting the Bitcoin bubble and not the value of the network. This topic is very recent and is still being studied, and will be further explored in this dissertation.

Tables also showed that over time the 30-day Moving Average of the value predicted by the Metcalfe and Odlyzko model seem to have the best correlations with the variables we are trying to predict. Therefore we chose to use 30-day MA for the analysis going further

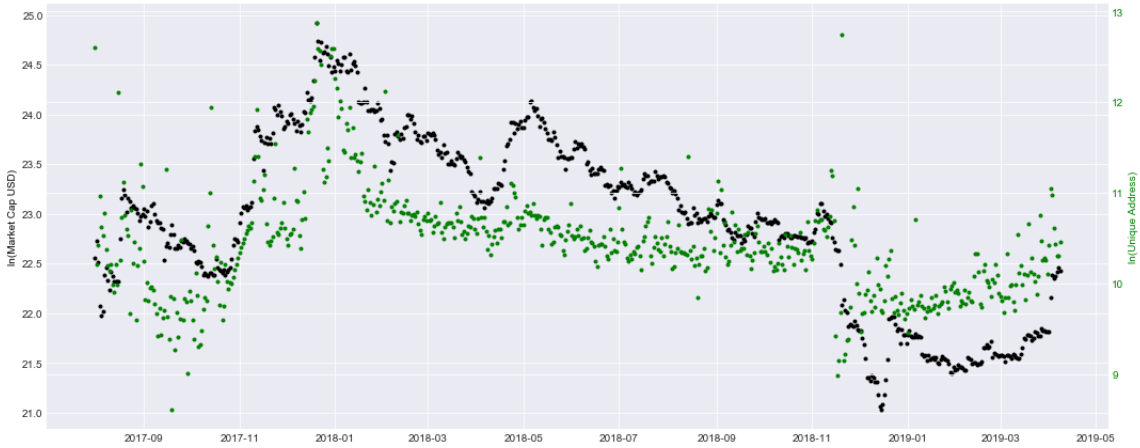

Figure 28 – Bitcoin Cash Market Cap and Daily Active Addresses Overlap

Source: Author(2019)

Figures 25, 26, 27 and 28 demonstrate the Market Cap values and overlaps the number of unique addresses connected to the network in the time line. Its visible the correlation between both, specially the peak near December 2017, present in all studied cryptocurrencies, except Bitcoin Cash, which started in 2018, but its still possible to see that the value followed the drop seen in all other 3 cryptocurrencies.

One statement made by Clearblocks article (CLEARBLOCKS, 2018) is that "In the case of Ethereum, ERC-20[1] transactions now represent upwards of one third of all on-chain transactions. Because ERC-20 transactions do not directly relate to trading of ETH, it's possible that we'll need to discount transactions in the future by some function to account for the rise in non-ETH transactions. Similarly, as DAPPS begin to launch and Ethereum sees increased non-speculative usage, further discounting may be needed". This basically means that smart contracts are now being processed in the network, but so far the correlation between price and addresses is still solid.

Kalichkin (KALICHKIN, 2018) showed in his article that "... Metcalfe's Law (Network Value n^2) probably overestimates network value, which is why it's logical to use it as an upper bound for valuation of Bitcoin network. At the same time we can use Odlyzko Law (Network Value n • log n) as a lower bound...". Based on that we fit a curve to both laws, following the formula bellow, also proposed in the same article cited:

$$ln(NVactual) = a + b * 30MA[ln(n^2)]$$

[1] ERC-20 is a technical standard used for smart contracts on the Ethereum blockchain for implementing tokens. ERC stands for Ethereum Request for Comment, and 20 is the number that was assigned to this request.

$$ln(NVactual) = a + b * 30MA[ln(n * ln(n)]$$

Where b is adjusted to the best fit to the Market Cap and Close Price while a is kept 0. A training and test set was defined using data from 2017 and 2018, testing with 2019 to prevent overfitting. After that, a is set empirically in order to define the best (narrowest) corridor that fits most of the observed variable (Close or Market Cap) within the corridor created, this way defining lower and upper bounds. The results can be observed in the figures 29, 30, 31 and 32.

The tables 2 and 3 display the model parameter results for Market Cap and Close Price. The figures display the results of the model for each currency.

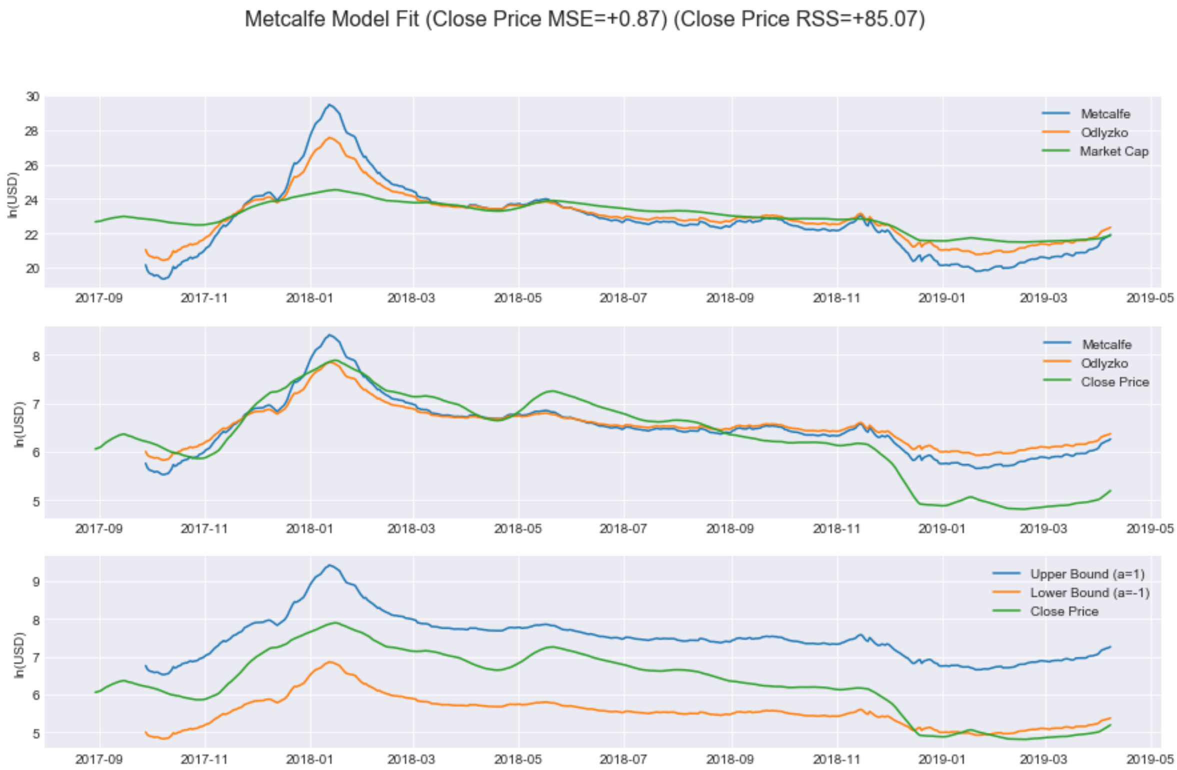

Figure 29 – Bitcoin Metcalfe Model

Source: Author(2019)

4.3.2 Other ways to evaluate Metcalfe model

After Ken Alabi (ALABI, 2017) proposed that digital blockchains appeared to be following Metcalfe‘s law, a number of studies derived from his views, mostly focused into investment advisory, but very relevant to this academic research. Its hard to track down the original source of these studies but the most relevant ones propose an analysis that, according to Powaga (POWAGA, 2017) is the PMR or Price-to-Metcalfe ratios. He states

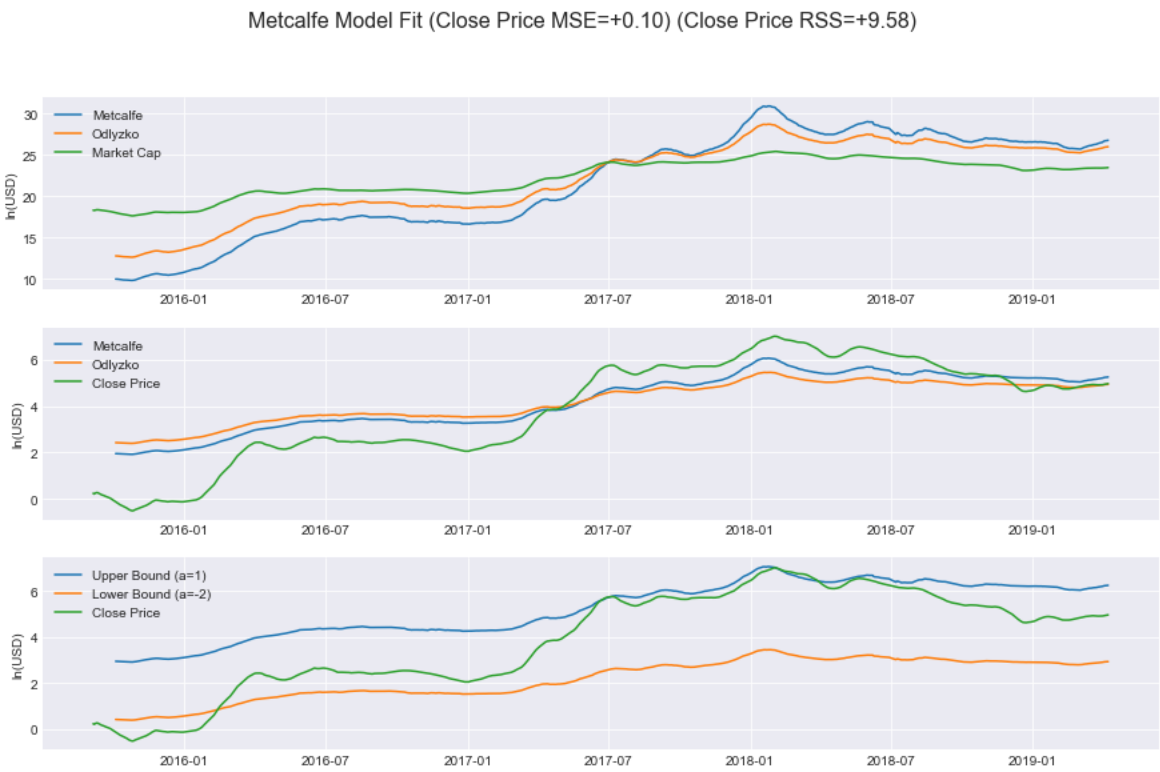

Figure 30 – Ether Metcalfe Model

Source: Author(2019)

Model for Market Cap				
	Metcalfe constant	Odlyzko constant	MSE	RSS
BTC	0.14431293	0.72375394	0.099762	9.776745
ETH	0.16649786	0.80670162	8.511013	834.079335
LTC	0.1928068	0.85337193	3.192534	312.868339
BCH	0.20752722	0.93332476	1.524510	149.402021

Table 2 – Best Parameters and Model Results for Market Cap

that it "(...) is somewhat analogous to price-to-book ratio in public equity analysis in that a higher ratio implies investors expect a given network to create more value from a given number of users(...)".

Clearblocks article states that (CLEARBLOCKS, 2018) "the idea here is to study the relationship between the price (or value) of a cryptoasset and its fundamentals (as suggested by network usage). This type of ratio analysis is gaining popularity with Network Value-to-Transactions (NVT) being the most widely studied."

Network value to transactions (NVT) formula is defined as follows:

$$NVT = \frac{DailyMarketCap.(USD)}{N - DayMovingAverageTransactionVolume(USD)}$$

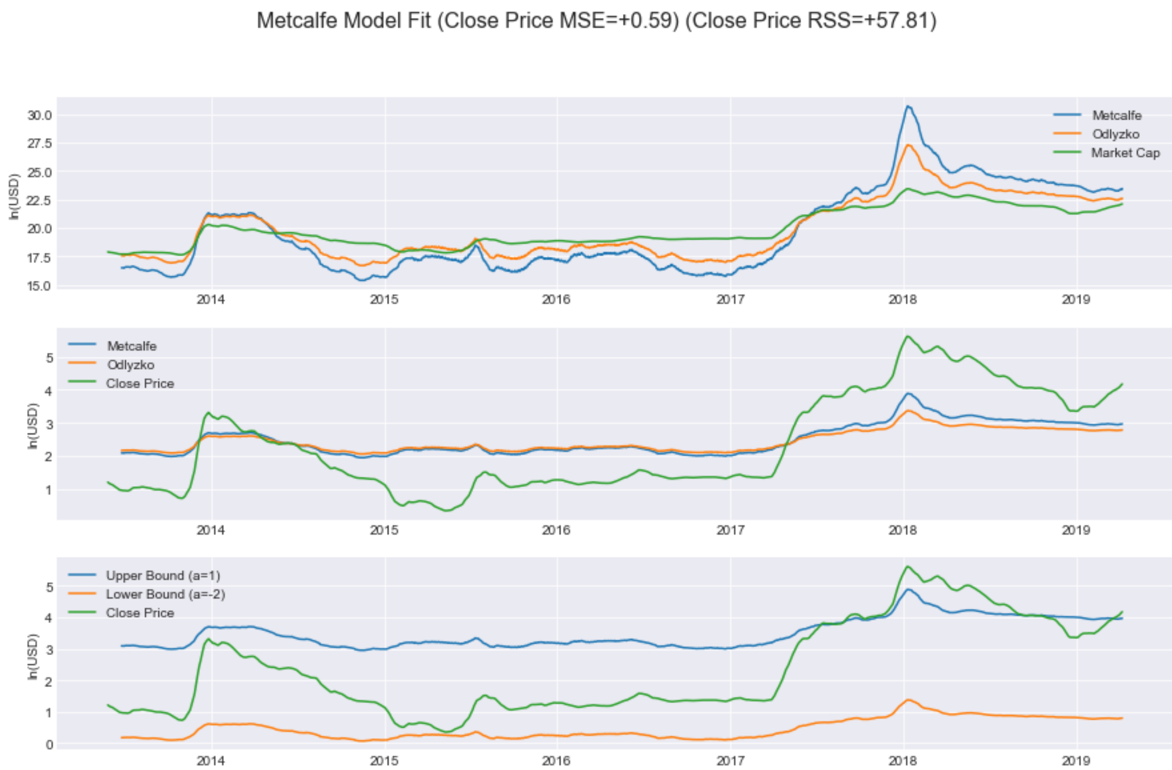

Figure 31 – Litecoin Metcalfe Model

Source: Author(2019)

Model for Close Price				
	Metcalfe constant	Odlyzko constant	MSE	RSS
BTC	0.04244276	0.21217174	0.708925	69.474708
ETH	0.03271134	0.15347398	0.097770	9.581546
LTC	0.02443636	0.10530999	0.589880	57.808332
BCH	0.0592456	0.26605885	0.868059	85.069848

Table 3 – Best Parameters and Model Results for Close Price

Where N-days was observed in several works mostly to be set as 30-days and 90-days, in some cases splitted 14 days backward facing and 14 days forward facing moving average and so on.

One of the key aspects to be considered here are the widely accepted challenges to calculate such ratio: obtaining reliable data. While the close price is widely available and can be easily obtained, exchanges like coinmetrics.io state that daily transaction volumes, although provided, are somewhat difficult to estimate due to off-chain transactions, and can suffer distortions of more than one third of its value.

Trying to face these challenges, Clearblocks proposed called Price-to-Metcalfe ratio approach, which was defined by the formula bellow:

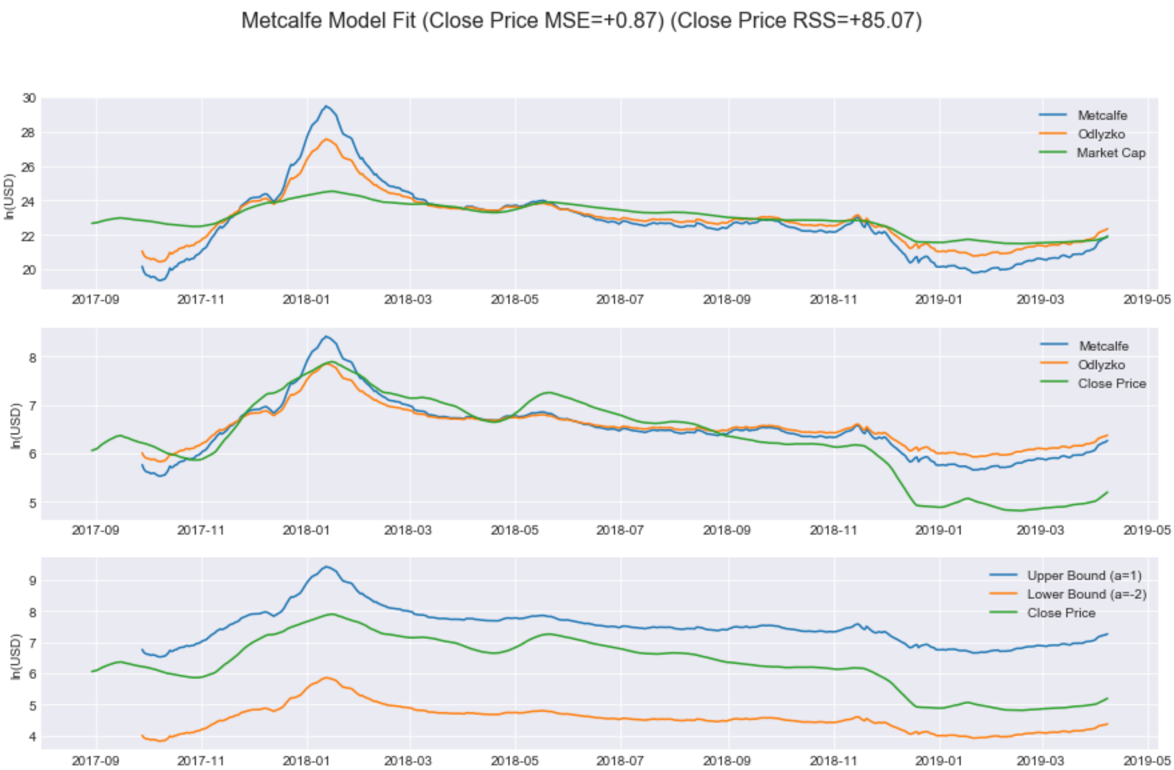

Figure 32 – Bitcoin Cash Metcalfe Model

Source: Author(2019)

$$PMR = ln(\frac{DailyNetworkValue(USD)}{30 - dayMAofMetcalfe'sFormula(orOdlyzko's)})$$

Latelly, the already mentioned before Kalichkin (KALICHKIN, 2018) proposed an evolution to the PMR formula by composing a single metric using 2 of the laws presented (Metcalfe's and Odlyzko's) as upper and lower bounds, and having the network value obtained by its average. This is acceptable because its already proven that the first over valuate the network, while the later under valuate it.

$$Upper\ Bound = C * 30MA(ln(n^2))$$

$$Lower\ Bound = C * 30MA(ln(n * ln(n))$$

$$NV = \frac{C1 * 30MA(ln(n^2)) + C2 * 30MA(ln(n * ln(n))}{2}$$

With this formulas, Kalichkin stated that he had "robust upper and lower bounds" and from there derived the ratio Network to Metcalfe (NVM) as per the formula bellow:

$$NVM = ln(NV\ actual) - ln(NV\ metcalfe) = ln(\frac{NV\ actual}{NV\ metcalfe}$$

For the purpose of this work, instead of setting the parameters of the formula empirically, we seek the best fit by running a non linear regression to the Market Cap value, which in all 4 cases have a near perfect correlation with the Close price. We then used the fitted formulas to calculate the average and obtain NV. Results are demonstrated in the figures 33, 34, 35 and 36.

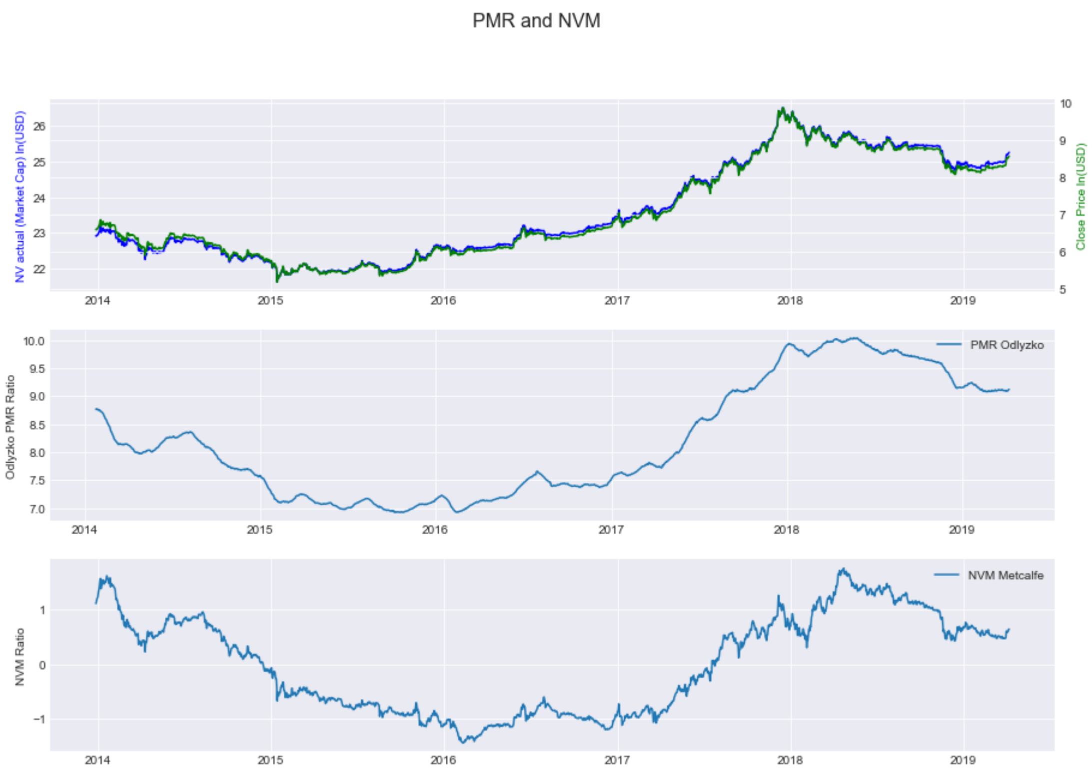

Figure 33 – Bitcoin, Price-To-Metcalfe, NetworkValue-to-Metcalfe

Source: Author(2019)

From the results obtained, this work seems to confirm what was stated by Kalichkin (KALICHKIN, 2018) in his work back in may 2018. NVM, although with known limitations, seems to be showing that there is still an over valuation on BTC price in the market at the moment. This can be observed by looking at the NVM normalized chart close to the upper limit of 1, which means close to the formula that over valuates the network. This behavior can also be observed for LTC, and that seems to make sense once you observe that both cryptoassets seem to be having a very similar behaviour over time.

ETH and BCH on the other side are showing undervaluation signs, which indicates a good moment to buy, what shall drive the price up. This movement can already be glimpsed in the last bit of the network price apparently.

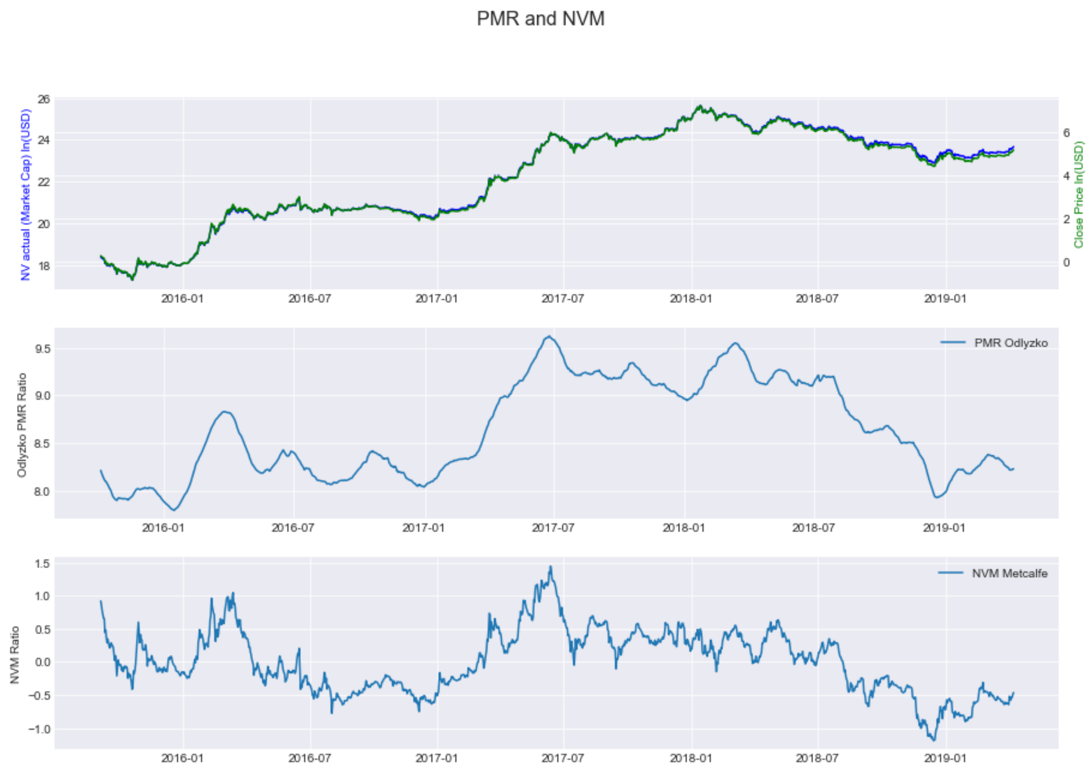

Figure 34 – Ether, Price-To-Metcalfe, NetworkValue-to-Metcalfe

Source: Author(2019)

In all cases, one of the most notorious moment that is correctly predicted by the ratios is the 2017 Q4 explosion of cryptoassets market followed by a correction in 2018 Q1/Q2.

Ether seems to be showing a slightly different behaviors, but this could be explained by the fact that the Ethereum blockchain started to process smart contracts. The lack of similar blockchains to compare leaves us with the possibility of only guessing in this case. Only time will tell!

4.3.3 Can we forecast the NV?

That seem to be a hard question to answer because the oldest blockchain evaluated is too young to tell how the growth of the network will happen.

Metcalfe (METCALFE, 2013) in his work proposed that the growth of users in a network as a function of the time happens in a sigmoid way, following what he called "Netoid" function. As per he states: "The sigmoid models a population's growth from 0 percent at time minus infinity to 100 percent at time plus infinity. The sigmoid adoption rate peaks at time 0.0 with a population fraction of 50 percent(...)The netoid has the same S-curve shape as the sigmoid. Its slope (the adoption rate) is proportional to the

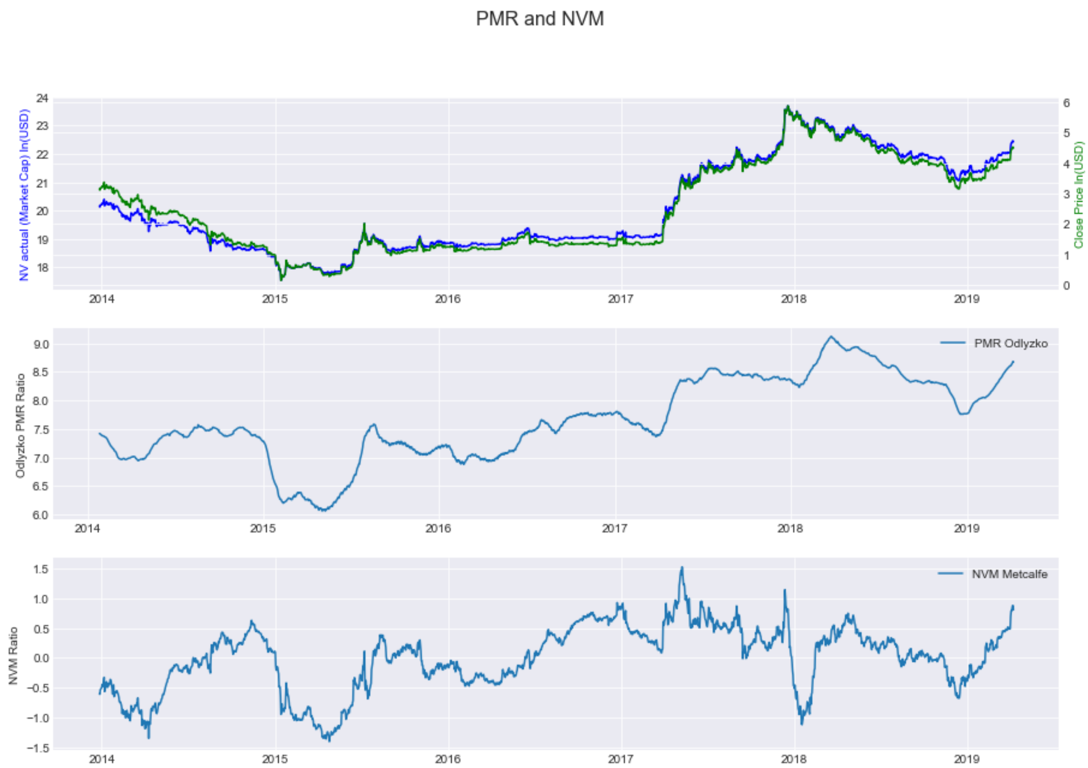

Figure 35 – Litecoin, Price-To-Metcalfe, NetworkValue-to-Metcalfe

Source: Author(2019)

product of the fraction of the population already adopted times the fraction awaiting adoption. It peaks when adoption is 50 percent. The adoption rate is driven by the number of adoptions so far and limited by the number of those awaiting adoption. The netoid offers three parameters: h, the point in time at which the growth rate is maximum, when the population is half the peak; v, the virality or speed with which adoption occurs; and p, the peak value, which the netoid approaches asymptotically. In short, the netoid can model when and how fast adoption will occur, and how large it will get".

$$Netoid(t) = \frac{p}{1 + e^{-v*(t-h)}}$$

Alabi (ALABI, 2017) then proposes a rationale in his work for the Netoid deriving it to a exponential equation. According to him "none of the blockchains we examined has yet reached a point of stagnating value, so we are only able to model the initial growth portion of the network. In addition, there no basis to expect that the parameter v will also govern the deceleration rate of the network. Consequently, the growth models examined here will focus on the initial growth section(...)". This statement remains true for this current work. These are the formulas proposed by Alabi:

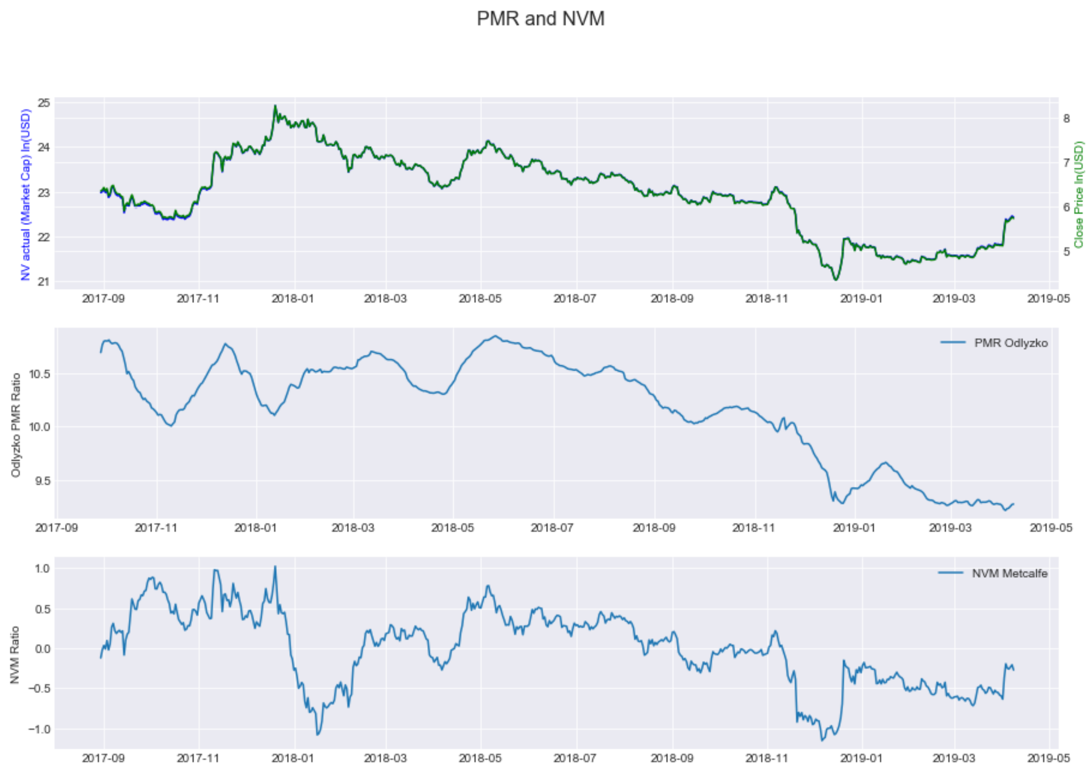

Figure 36 – Bitcoin Cash, Price-To-Metcalfe, NetworkValue-to-Metcalfe

Source: Author(2019)

$$\frac{\partial n}{\partial t} = vN$$

$$N(t) = N_0 e^{vt}$$

In this work, its possible to observe in the figures 25, 26 and 27, (for BTC, ETH and LTC respectively) that the growth of Daily Unique Address is close to constant (with some noise obviously) in the logarithm scale, which means it shall fit to the proposed formulas, but still, won‘t be possible to derive any conclusions due to the low maturity of the networks. Any attempt to fit and project the growth would be based on guesses and speculations, which is not the purpose of this work. The only currency that show a slight difference in the chart 28 is Bitcoin Cash, which we attribute to the fact that this is the youngest of the currencies evaluated and if you observe the beginning of every cryptoasset, what you normally see is a long period of a near constant adoption until it "gains traction" and you start to see a growth in the number of addresses.

All currencies also show a distortion followed by correction by the end of 2017 and beginning of 2018.

4.4 Sentiment Analysis on Metcalfe

During the course of this work we asked ourselves the question if sentiment analysis could improve whatever was modeled by Metcalfe's law, bringing the result closer to the actual result. This could be useful by having an adjustment factor derived from the live analysis of news in social media using their wide available APIs to do so.

The main inspiration came from an article by Kaminski (KAMINSKI, 2014) where he evaluated how twitter signals would enable now-casting of the BTC market. In his findings, he states that "a lagged correlation analysis showed that the sum of emotional sentiments and especially negative signals positively correlate with the intraday trading volume within the last 48 hours (...) The Granger causality analysis shows that there is no statistical significance for Twitter signals as a predictor of Bitcoin with regard to the close price, intraday spread or intraday return. (...) the microblogging platform Twitter may be interpreted as a virtual trading floor that emotionally reflects Bitcoin's market movement."

The pilot for this study was done using data from Reddit.com, with an intention of moving to analyze Twitter data in the future. Unfortunately during the course of this development the SSD Drive storing and processing the Twitter data set for over 4 months stopped working out of thin air (together with the whole laptop) and was not possible to recover that data specifically.

Due to that, and in order to avoid loosing this study, the author of this dissertation decided to failover to the Reddit.data, improve it as much as possible, and dive deep into the analysis obtaining the results that are presented in this section. The suggestion to use twitter stays as a further work suggestion.

Differently from Kaminski, in this dissertation it was used a gold standard lexicon to extract the sentiment from the news title, instead of using query keywords, and the objective is to observe if there is correlation between the sentiment detected per day and the current close price or lagged version of it, up until 72 hours. This will be conducted in 24 hours windows since the data we have have daily granularity.

Correlations were evaluated against all 4 currencies, and we lately try to find a potential impact of the sentiment result over that day result, creating a correction factor that, for the purpose of this dissertation, shall be set empirically.

4.4.1 Using Vader

We extracted the posts from the social network using the official API, added that current extraction to the data set used in the pilot, achieving a total of 47065 posts. Since Reddit allow cross posting in multiple communities from the subset we are evaluating, after de duplication the number of posts to work with went to 38119 spread across approximately

76 days (30 in 2018 and 30 in 2019).

We took the titles of the posts and applied the process of sentiment analysis using the VADER (HUTTO; GILBERT, 2014) lexicon, obtaining an output similar to the ones bellow:

```
{'neg': 0.216, 'neu': 0.784, 'pos': 0.0, 'compound': -0.296}
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
{'neg': 0.0, 'neu': 0.5, 'pos': 0.5, 'compound': 0.4588}
```

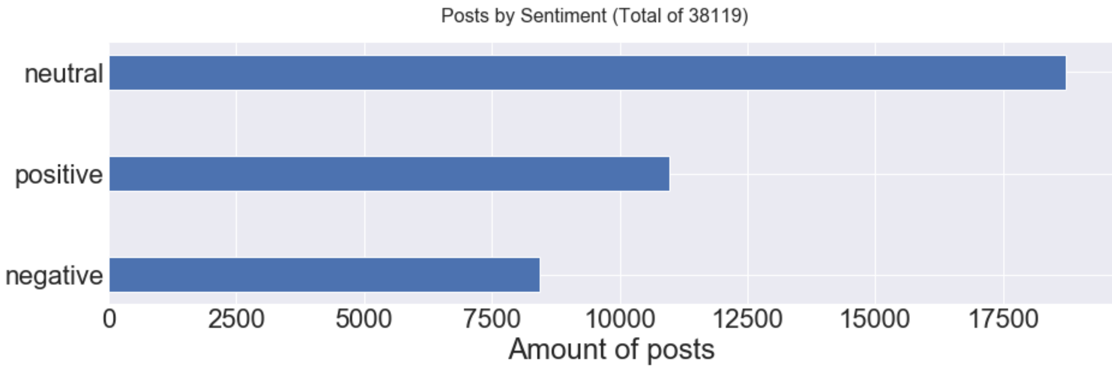

Figure 37 – Sentiment Data Volumes

Source: Author(2019)

The compound is the metric we use to define if a day is positive or negative. It goes from -1 to 1 being -1 negative, gradually increasing to 0 (neutral) and to 1 (positive). Given the distribution of posts collected, exhibited in figures 37 and 38 here we then proposed the following approaches to determine if a day was positive or negative:

- Count the number of positive, neutral and negative posts in a certain date and use those as metrics.
- Average the compound within that day as a new metric.

With those metrics we investigated further if there is correlation of this data with the Close price and lagged versions of it so after that we can apply a certain amount of correction to the formula prediction and see if the Metcalfe model improves by any means, compared to the predicted variable.

4.4.2 Price Correlation Analysis

Correlation analysis seem to confirm the findings by Kaminski (KAMINSKI, 2014) since we found, although weak, a lagged correlation between the Close price and the counts

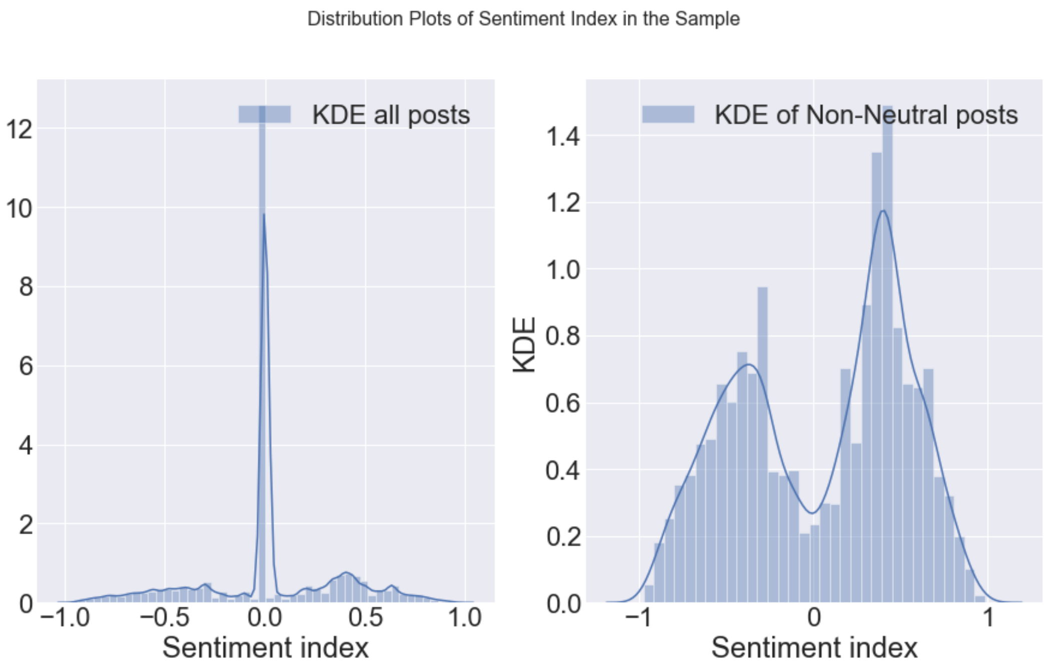

Figure 38 – Sentiment Data Distribution

Source: Author(2019)

Pearson Correlation

COUNTS		Close Price	Close d-1	Close d-2	Close d-3
BTC	Positive	0,174141923	0,188242396	0,198357822	0,184458462
BTC	Negative	0,160119731	0,178926549	0,208970178	0,20797458
BTC	Neutral	0,212618028	0,231583282	0,243099958	0,224808145
ETH	Positive	0,146535117	0,156162371	0,165520778	0,169111273
ETH	Negative	0,167445449	0,181371867	0,20244833	0,210733154
ETH	Neutral	0,189966805	0,20099994	0,212651807	0,213140484
LTC	Positive	0,174956025	0,185035421	0,190888396	0,194303594
LTC	Negative	0,173303435	0,185801386	0,213471935	0,225798363
LTC	Neutral	0,214467706	0,224830675	0,238321908	0,237567437
BCH	Positive	0,191716352	0,198677363	0,203657522	0,200628376
BCH	Negative	0,198993431	0,206506375	0,227396035	0,229039366
BCH	Neutral	0,225974736	0,237490667	0,244591817	0,23497646

Figure 39 – Correlation Analysis - Sentiment & Close Price

Source: Author(2019)

of posts classified with each sentiment. As already stated, although the correlation is week, by the figure 39 we can observe that it improves while we compare it with the lagged versions of the price, peaking on D-2.

The numbers obtained from this Pearson correlation made us not to go further in this direction since would be really hard to prove the relation between both events in any of the cases

4.4.3 Using average compound metric

To proceed with this test we propose a variation to the Metcalfe formula where we introduce an error factor per day, affecting a percentile of the total result positively or negatively depending on the compound metric being greater than 0 or not.

$$Metcalfe = N^2$$

$$Metcalfe_{Sentiment\,adjusted} = N^2 + (a * N^2 * Compound)$$

Where "a" is a number between 0 and 1 representing the percentile of impact over that specific time period the compound metric will have some impact on. In other words the prediction of that day can be boosted or pushed down by "a%" of the value predicted multiplied by the compound obtained for that day. For this first evaluation "a" will be no more than a guess, since there are no study that defines how much of a value can be impacted by the news.

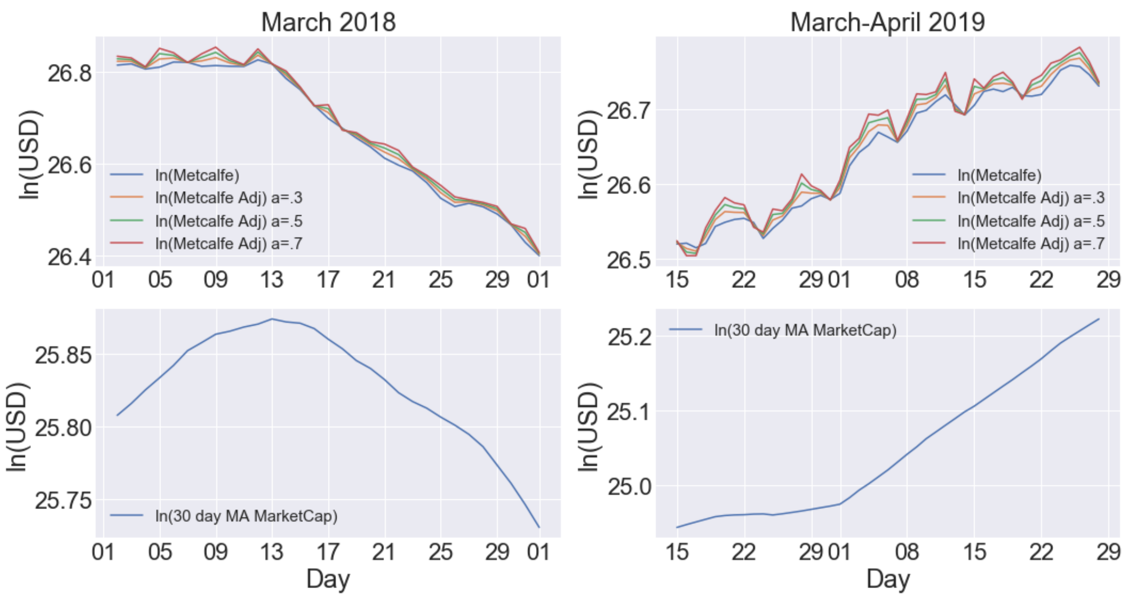

Figure 40 – Metcalfe Adjusted - Bitcoin Compound Correction

Source: Author(2019)

Plots in the figures 40, 41, 42 and 43 demonstrate how the adjustment is changing the original prediction, but when we go down to the statistical test, evaluation demonstrates that in fact the correction is not showing and improvement, but it is making the prediction worse. Following numbers demonstrate that:

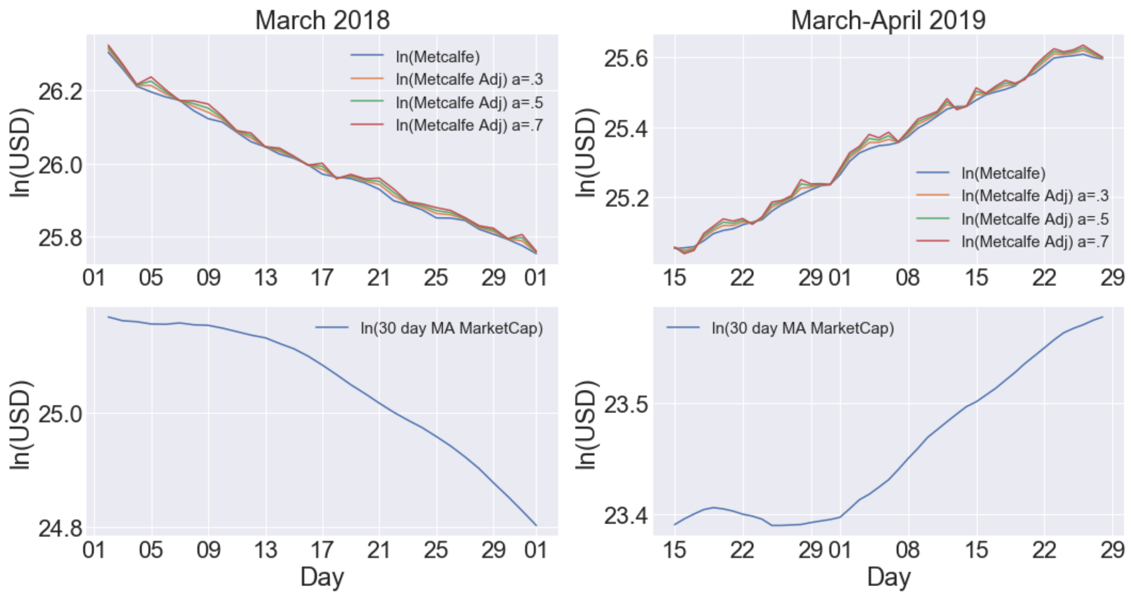

Figure 41 – Metcalfe Adjusted - Ether Compound Correction

Source: Author(2019)

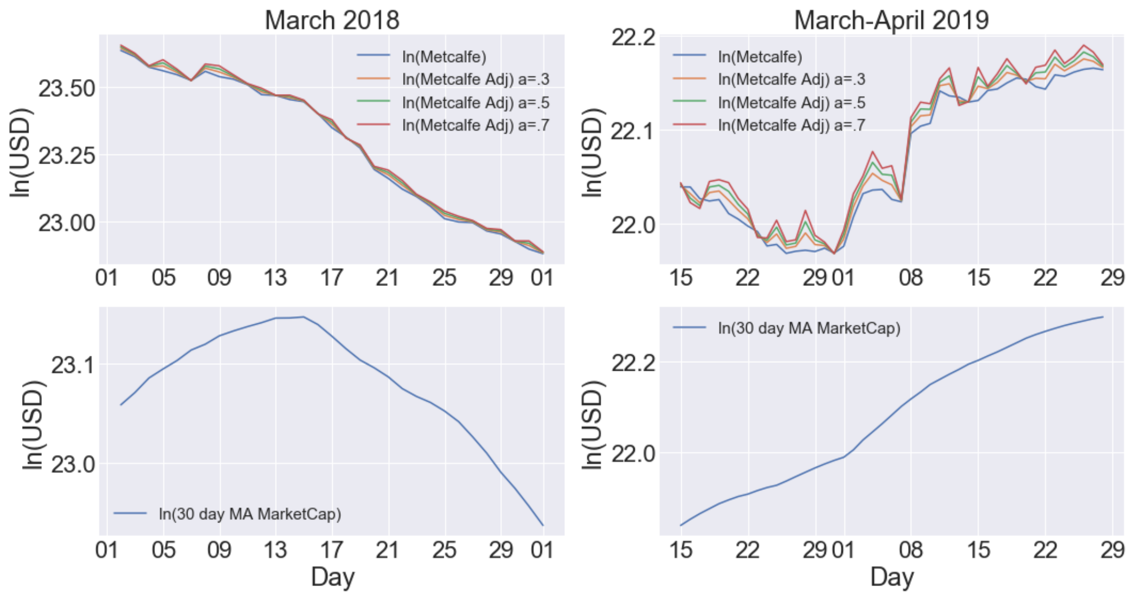

Figure 42 – Metcalfe Adjusted - Litecoin Compound Correction

Source: Author(2019)

Again here, the values are too bad and seem to get worse when we try to boost the impact of the sentiment analysis. This adds up to what we already saw in the previous section, leading us to believe that in fact this hypothesis shall be rejected.

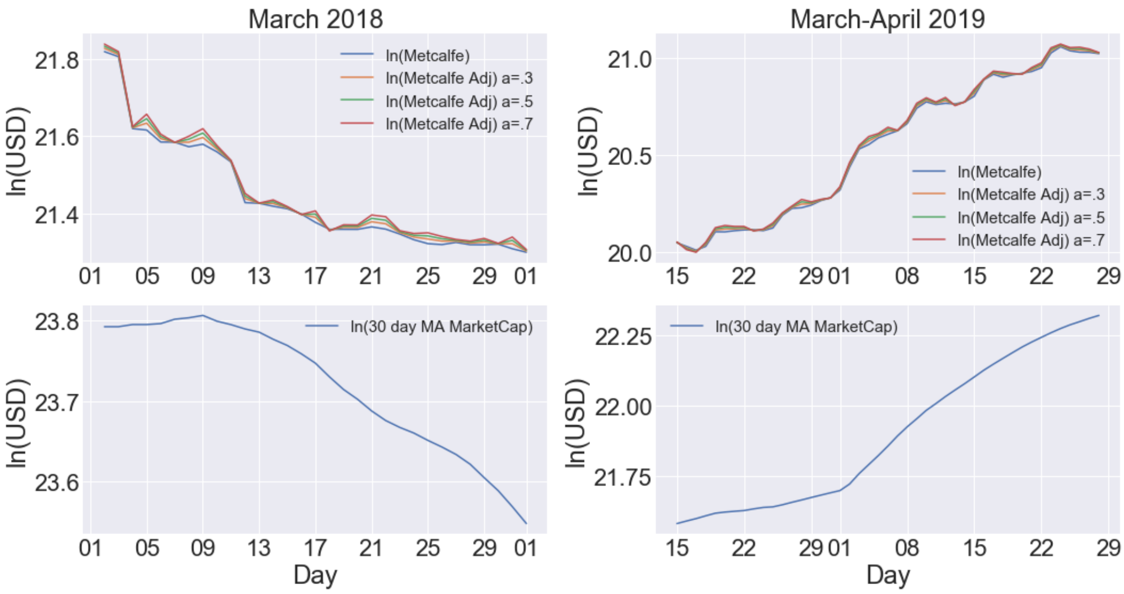

Figure 43 – Metcalfe Adjusted - Bitcoin Cash Compound Correction

Source: Author(2019)

Mean Squared Error			
		March/2018	**March-April/2019**
BITCOIN	**Metcalfe (a=0)**	0.7294802444489514	2.547972789855084
	Metcalfe Adjusted (a=.3)	0.7410995193039241	2.5698031078447596
	Metcalfe Adjusted (a=.5)	0.7488561791014701	2.5843026725019014
	Metcalfe Adjusted (a=.7)	0.75662046234057	2.5987592121363305
ETHER	**Metcalfe (a=0)**	0.902906067884878	3.5641220702631706
	Metcalfe Adjusted (a=.3)	0.9159234905026153	3.590038745996229
	Metcalfe Adjusted (a=.5)	0.9246044983271539	3.607237504445088
	Metcalfe Adjusted (a=.7)	0.933287040319335	3.624373561874783
LITECOIN	**Metcalfe (a=0)**	0.08728678616728155	0.008063756065468688
	Metcalfe Adjusted (a=.3)	0.0903008374775301	0.007785131986114746
	Metcalfe Adjusted (a=.5)	0.09237119754132235	0.007687464856538133
	Metcalfe Adjusted (a=.7)	0.09448883433666848	0.007658372925086023
BITCOIN CASH	**Metcalfe (a=0)**	5.1856668970061985	1.8420377668910979
	Metcalfe Adjusted (a=.3)	5.154980047594118	1.8241523658047416
	Metcalfe Adjusted (a=.5)	5.13478518842589	1.8124328190828167
	Metcalfe Adjusted (a=.7)	5.114796141050637	1.800872977563471

Table 4 – MSE Market Cap to Predicted Values

5 Conclusion

5.1 Hypothesis 1 - Metcalfe can fit to other cryptocurrencies?

As our first hypothesis we aimed to answer if, considering recent experiments with Metcalfe's law modeling Bitcoin price, can the same law be used to model other cryptocurrencies?"

This dissertation delivers evidence that support this hypothesis and offers more than that, by demonstrating unconsidered applications of the law, like as a bubble predictor.

Daily active addresses are the corner stone of this model and, although we still have uncertainties around a market that is still giving its first steps, everything is pointing towards the fact that more and more we shall be able to understand the growth of the network and use it in our favour to model the asset and predict the best moment to buy and sell.

The exchanges are moving forward and recently we observed some articles published by them providing more reliable data for volume estimation. With this we might soon see the need to revisit the models of this dissertation to understand what will be the impact when the so called "off-chain" transactions are included in the calculations.

5.2 Hypothesis 2 - Is Metcalfe somehow comparable to ARIMA? How?

Our second research question was: Is Metcalfe's law able to model cryptocurrencies prices better than known time series models (given that blockchain is a network in essence)?

And based on the results presented, the author believe that yes, Metcalfe law and its variations performed better than ARIMA. Also Metcalfe models offered a bigger amount variation possibilities to play around and tweak to find better results.

Say that a model is better than other is mathematically correct but also relative, since both performed well in the tests executed. ARIMA is a pure statistical method of forecasting, widely used, easy to understand and explain, while Metcalfe is closer to a qualitative method, that assess a second variable to forecast its objective. We could easily, for instance, use ARIMA to forecast the predictions of Metcalfe formula.

The reasons behind Metcalfe tests are still hard to explain, and the cryptocurrency phenomena is far to young to tell if this correlation will persist. Cryptocurrencies are not 100% supported by the fundamentals and we can't say if they will ever be. The fact is that, in the worse of the cases, investors should follow the components of Metcalfe (DAA in our case) closely as an attempt to monitor the health of the market before jumping in or out.

5.3 Hypothesis 3 - Can sentiment analysis in social media improve Metcalfe?

Final question this dissertation tried to answer was: Considering that network models are based on the number of nodes, connections and users, can social media network sentiment analysis provide a better view on the size of the network and improve the way the price is modeled?

And in this case we provided evidence that we can reject this research question by considering that the social media reflects the market instead of the opposite.

Even though we faced a step back here with the loss of some data in the course of the development of this work, the data we had provided enough insights that the correlation of the two subjects (social media and cryptocurrencies) is too small to derive some cause from one over the other.

Qualitative methods such as market research need to have a more direct focus in a specific group and subject to enable us to prove some sort of cause-effect situation. To do that with social media we would have to monitor and process all the data, not only a subset.

5.4 Next Steps

A few areas can be derived here as opportunities for further work. One being to split the topics here discussed in different lines of development, a bit more independent and enable the delivery of more objective articles, focused in a single aspect of the problem.

For the time statistical series models:

- Explore further the best ways to find parameters and test windows, maybe exploring RNN Neural Networks to do so.
- With the time going by, opportunities to evaluate trend projection shall appear in the horizon.

For network models:

- Identify differences across data sources and find the most mature data source to deliver the best results.

- Explore further the movement of daily active users and try to fit the netoid or some other variation.

5.5 Disclaimer

Following the suggestion kindly given by a fellow auditor of this research and market specialist we decided to include this disclaimer!

None of the statements in the article should be considered investment advice and the authors of this dissertation are not investors of any cryptocurrencies currently.

This work fits an academic purpose in trying to explain the behaviour of cryptocurrencies.

Due to the various risks and uncertainties, actual performance of the assets may differ materially from that reflected or contemplated in forward-looking statements.

Bibliography

ALABI, Ken. Digital blockchain networks appear to be following metcalfe's law. *Electronic Commerce Research and Applications*, v. 24, p. 23 – 29, 2017. ISSN 1567-4223. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1567422317300480>. Citado 6 vezes nas páginas 16, 31, 36, 37, 54, and 60.

ANKALKOTI, Prashant; SANTHOSH, S G. A Relative Study on Bitcoin Mining. *Imperial Journal of Interdisciplinary Research (IJIR)*, v. 3, n. 5, p. 1757–1761, 2017. Disponível em: <http://www.imperialjournals.com/index.php/IJIR/article/view/5024/4834>. Citado 2 vezes nas páginas 9 and 24.

AZARI, Amin. Bitcoin price prediction: An arima approach. 10 2018. Citado 4 vezes nas páginas 37, 42, 46, and 49.

BOX, George.E.P.; JENKINS, Gwilym M. *Time Series Analysis: Forecasting and Control*. [S.l.]: Holden-Day, 1976. Citado 2 vezes nas páginas 35 and 42.

BRISCOE, Bob; ODLYZKO, Andrew; TILLY, Benjamin. Metcalfe's law is wrong. *IEEE Spectrum: Technology, Engineering, and Science News*, IEEE Spectrum, Jul 2006. Disponível em: <https://spectrum.ieee.org/computing/networks/metcalfes-law-is-wrong>. Citado 2 vezes nas páginas 22 and 35.

CLEARBLOCKS. *Valuing Bitcoin and Ethereum with Metcalfe's Law*. Medium, 2018. Disponível em: <https://medium.com/@clearblocks/valuing-bitcoin-and-ethereum-with-metcalfes-law-aaa743f469f6>. Citado 3 vezes nas páginas 35, 53, and 55.

De Gooijer, Jan G.; HYNDMAN, Rob J. 25 Years of Time Series Forecasting. *International Journal of Forecasting*, v. 22, n. 3, p. 443–473, 2006. ISSN 01692070. Citado 3 vezes nas páginas 9, 19, and 37.

HUTTO, CJ; GILBERT, Eric. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*. [S.l.: s.n.], 2014. Citado 4 vezes nas páginas 18, 26, 32, and 63.

KALICHKIN, Dmitry. *Rethinking Metcalfe's Law applications to cryptoasset valuation*. Cryptolab Capital, 2018. Disponível em: <https://medium.com/cryptolab/network-value-to-metcalfe-nvm-ratio-fd59ca3add76>. Citado 4 vezes nas páginas 51, 53, 57, and 58.

KAMINSKI, Jermain. Nowcasting the Bitcoin Market with Twitter Signals. jun 2014. Disponível em: <http://arxiv.org/abs/1406.7577>. Citado 4 vezes nas páginas 17, 32, 62, and 63.

KHATWANI, Sudhir. *Top Five Blockchain 3.0 To Watch Out For In 2018*. 2018. Https://coinsutra.com/3rd-generation-blockchain/. Citado 2 vezes nas páginas 9 and 25.

KROLL, Joshua a; DAVEY, Ian C; FELTEN, Edward W. The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries. *The Twelfth Workshop on the Economics of Information Security (WEIS 2013)*, n. Weis, p. 1–21, 2013. ISSN 15437221. Citado na página 25.

LIU, Bing. Sentiment analysis and subjectivity. *Handbook of natural language processing*, v. 2, p. 627–666, 2010. Citado na página 26.

LIU, Bing. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012. Citado na página 26.

METCALFE, Bob. Metcalfe's law after 40 years of ethernet. *Computer*, v. 46, n. 12, p. 26–31, 2013. ISSN 00189162. Citado 8 vezes nas páginas 15, 20, 21, 31, 35, 36, 51, and 59.

NAKAMOTO, Satoshi. Bitcoin: A Peer-to-Peer Electronic Cash System. *Www.Bitcoin.Org*, p. 9, 2008. ISSN 09254560. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Citado 2 vezes nas páginas 23 and 24.

NOFSINGER, John R. Social mood and financial economics. *The Journal of Behavioral Finance*, Taylor & Francis, v. 6, n. 3, p. 144–160, 2005. Citado na página 17.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. v. 2, n. 1, 2008. Citado na página 26.

PETERSON, Timothy F. Metcalfe ' s Law as a Model for Bitcoin ' s Value Metcalfe ' s Law as a Model for Bitcoin ' s Value. p. 1–23, 2017. Citado 4 vezes nas páginas 16, 21, 25, and 37.

POWAGA, Stephen. *Bitcoin and Metcalfe's Law*. ETF Momentum Investing, 2017. Disponível em: <https://www.advisorperspectives.com/commentaries/2017/11/28/bitcoin-and-metcalfes-law>. Citado na página 54.

SWAN, Melanie. *Blockchain: Blueprint for a new economy*. [S.l.]: " O'Reilly Media, Inc.", 2015. Citado na página 23.

TLEGENOVA, Daniya. Forecasting Exchange Rates Using Time Series Analysis: The sample of the currency of Kazakhstan. *SIAM Journal on Financial Mathematics*, v. 6, n. 1, p. 467–486, 2015. Disponível em: <https://arxiv.org/ftp/arxiv/papers/1508/1508.07534.pdfhttp://epubs.siam.org/doi/10.1137/140967635>. Citado na página 37.

Van Vliet, Ben. An alternative model of Metcalfe's Law for valuing Bitcoin. *Economics Letters*, Elsevier B.V., v. 165, p. 70–72, 2018. ISSN 01651765. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0165176518300557>. Citado na página 37.

WHEATLEY, Spencer; SORNETTE, Didier; HUBER, Tobias; REPPEN, Max; GANTNER, Robert N. Are bitcoin bubbles predictable? combining a generalized metcalfe's law and the lppls model. *SSRN Electronic Journal*, 03 2018. Citado na página 52.

WOO, Willy. *Is Bitcoin In A Bubble? Check The NVT Ratio*. Forbes Magazine, 2017. Disponível em: <https://www.forbes.com/sites/wwoo/2017/09/29/is-bitcoin-in-a-bubble-check-the-nvt-ratio/#4852c0166a23>. Citado na página 51.

ZHANG, Xing Zhou; LIU, Jing Jie; XU, Zhi Wei. Tencent and Facebook Data Validate Metcalfe's Law. *Journal of Computer Science and Technology*, v. 30, n. 2, p. 246–251, 2015. ISSN 10009000. Citado 3 vezes nas páginas 21, 22, and 51.

A Correlation Tables of Cryptocurrencies and Metcalfe Law

		Metcalfe (N**2)			Odlyzko (N log(N))		
		Original	30 day MA	90 day MA	Original	30 day MA	90 day MA
Close Price	2014	-0,358353812	-0,446258088	-0,445045681	-0,354939165	-0,4451218	-0,46452756
	2015	0,820961965	0,759521971	-0,183761793	0,788691522	0,72005538	-0,232574068
	2016	0,584957444	0,744494891	0,794142938	0,568241012	0,741973566	0,778492625
	2017	0,807512396	0,724215608	0,592423008	0,762388646	0,71402792	0,587809427
	2018	0,68975965	0,755325635	0,466952482	0,653841321	0,743924109	0,506327448
	2019	0,494188923	0,859940839	-0,953615761	0,467384018	0,858194249	-0,954911678
	All Time	0,735984132	0,777415674	0,789842251	0,69665698	0,737817555	0,761948527
Market Cap	2014	-0,340100619	-0,472271727	-0,35393665	-0,337209923	-0,47123785	-0,375310938
	2015	0,850190412	0,836714146	0,07597696	0,823010894	0,804573701	0,026668741
	2016	0,589837862	0,749828199	0,802704328	0,573373644	0,747917085	0,787225324
	2017	0,807578355	0,724008064	0,590136241	0,762406338	0,71377652	0,585568504
	2018	0,687903217	0,750794267	0,464944163	0,652401005	0,739162325	0,504259242
	2019	0,501773464	0,844833223	-0,955521136	0,475455648	0,842586372	-0,956802287
	All Time	0,733234763	0,775375962	0,790848269	0,697088774	0,738627512	0,764763345
Est. Volume	2014	0,130103004	-0,06775653	-0,87855471	0,15603894	-0,073787182	-0,879022443
	2015	0,617151272	0,732233338	0,585563056	0,590374797	0,71043979	0,557508888
	2016	0,231152061	0,270832996	0,516686885	0,23480321	0,279475862	0,513610692
	2017	0,772573832	0,758201842	0,546439007	0,723719791	0,743928368	0,544609112
	2018	0,816426109	0,827668107	0,461617964	0,775938952	0,814838732	0,502482537
	2019	0,644131779	0,620377476	0,505910026	0,621872399	0,613710813	0,507689296
	All Time	0,684482888	0,741637113	0,743903853	0,640422747	0,692976289	0,717404188

Figure 44 – Bitcoin - Metcalfe Formula Correlation with Price, Market Cap and Est. Volume

Source: Author(2019)

		Metcalfe (N**2)			Odlyzko (N log(N))		
		Original	30 day MA	90 day MA	Original	30 day MA	90 day MA
Close	2016	0,533926097	0,640081767	0,543808918	0,605895426	0,703695133	0,655468481
	2017	0,807849936	0,868882552	0,764925624	0,941245838	0,900354374	0,840703525
	2018	0,742903103	0,713259474	-0,094645634	0,800898595	0,743092145	-0,17164726
	2019	0,569599671	-0,747821119	0,835286043	0,575301397	-0,74725985	0,825916414
	All Time	0,836590311	0,830138331	0,629691686	0,890239939	0,863474641	0,697970911
MarketCap	2016	0,585561175	0,711686672	0,602464895	0,662531989	0,770280828	0,708430681
	2017	0,815223919	0,875770983	0,775296889	0,945697143	0,904147501	0,849380628
	2018	0,739212935	0,713132922	-0,077745994	0,798673267	0,743875171	-0,153090373
	2019	0,57488817	-0,775702233	0,811007261	0,579892101	-0,774842876	0,801078123
	All Time	0,838579884	0,83720643	0,647353061	0,896807574	0,873931559	0,715538788
Volume	2016	0,259966528	-0,099485801	-0,226854924	0,216023548	-0,024317874	-0,070141372
	2017	0,782122	0,696236191	0,554151493	0,833187991	0,730857138	0,64804126
	2018	0,832064965	0,463242515	-0,406981206	0,795712618	0,458921106	-0,468789162
	2019	0,629445926	-0,958997408	-0,884446349	0,603386229	-0,956124927	-0,89049586
	All Time	0,700777176	0,646544631	0,628711247	0,77455231	0,769397664	0,765369892

Figure 45 – Ether - Metcalfe Formula Correlation with Price, Market Cap and Est. Volume

Source: Author(2019)

		Metcalfe (N**2)			Odlyzko (N log(N))		
		Original	30 day MA	90 day MA	Original	30 day MA	90 day MA
Close Price	2014	0,410294647	0,910276065	0,599588919	0,732742812	0,944954346	0,631543219
	2015	0,030829717	0,232021083	0,651249945	0,012047845	0,206745619	0,645296856
	2016	-0,155065939	0,1136016	0,612230955	-0,148410604	0,075807106	0,591315382
	2017	0,855745493	0,85384094	0,90084759	0,951810795	0,829109594	0,930260442
	2018	0,541932365	0,551656802	0,284026743	0,701217332	0,641921076	0,245420106
	2019	0,412153554	-0,710266204	-0,464656878	0,428223307	-0,705841834	-0,477040317
	All Time	0,680654979	0,674793184	0,692844279	0,89815916	0,848354729	0,789915484
Market Cap	2014	0,402541645	0,890120122	0,636940716	0,717630477	0,932708658	0,669583331
	2015	0,017055224	0,238254315	0,744085594	-0,001743991	0,219862465	0,737227541
	2016	-0,246792371	-0,072979468	0,576251077	-0,254338354	-0,113071467	0,55857004
	2017	0,8576678	0,85340216	0,906828127	0,952672439	0,826009671	0,933649099
	2018	0,535201602	0,547036613	0,301113647	0,694091475	0,638155378	0,264318883
	2019	0,41120049	-0,711119954	-0,521350537	0,42693602	-0,706647394	-0,533270216
	All Time	0,671802233	0,672972098	0,703686482	0,895544546	0,851461742	0,802086912
Est. Volume	2014	0,242734735	0,797928334	0,259147493	0,394973664	0,73485707	0,282107405
	2015	0,234999731	0,303366206	0,591064686	0,235793495	0,283880722	0,579428293
	2016	0,005312032	0,276422985	0,319362671	0,021896025	0,238593831	0,28701905
	2017	0,706747729	0,650233154	0,71758746	0,761559942	0,646716947	0,774251979
	2018	0,750607159	0,406446072	-0,184465541	0,772685887	0,45957263	-0,24967869
	2019	0,458366195	-0,728440501	-0,950118513	0,499881109	-0,723375169	-0,954242964
	All Time	0,519258881	0,422845385	0,433798537	0,64207607	0,598331948	0,590635739

Figure 46 – Litecoin - Metcalfe Formula Correlation with Price, Market Cap and Est. Volume

Source: Author(2019)

		Metcalfe (N**2)			Odlyzko (N log(N))		
		Original	30 day MA	90 day MA	Original	30 day MA	90 day MA
Close	2018	0,026038803	0,660547023	0,150825639	0,267025168	0,718134046	0,140290423
	2019	0,565291876	0,309526283	-0,057684262	0,584447716	0,364777203	-0,041808728
	All Time	0,07896432	0,653987838	0,322927194	0,371469734	0,751270975	0,333324838
MarketCap	2018	0,025076848	0,655597291	0,152143384	0,265799236	0,712871012	0,142028421
	2019	0,56712277	0,288071951	-0,06077821	0,586981134	0,343003384	-0,044940111
	All Time	0,078426576	0,649115593	0,324998859	0,370965217	0,746808988	0,335892701
Volume	2018	0,028355833	0,473603831	-0,137562503	0,218619111	0,496360616	-0,168468471
	2019	0,663098314	0,151389182	-0,603671307	0,654712439	0,1757533	-0,603879386
	All Time	0,047516528	0,502886096	0,022294023	0,245896468	0,55174389	0,009865132

Figure 47 – Bitcoin Cash - Metcalfe Formula Correlation with Price, Market Cap and Est. Volume

Source: Author(2019)

B ARIMA Model Output

B.1 Bitcoin

ARIMA Model Results

Dep. Variable:	D.Close	No. Observations:	820
Model:	ARIMA(2, 1, 8)	Log Likelihood	1370.624
Method:	css-mle	S.D. of innovations	0.045
Date:	Mon, 29 Apr 2019	AIC	-2719.249
Time:	23:59:15	BIC	-2667.446
Sample:	01-09-2017	HQIC	-2699.372
	- 04-08-2019		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1.D.Close	0.9150	0.118	7.750	0.000	0.684	1.146
ar.L2.D.Close	0.0079	0.037	0.212	0.832	-0.065	0.081
ma.L1.D.Close	-0.9046	0.111	-8.135	0.000	-1.123	-0.687
ma.L2.D.Close	0.0133	nan	nan	nan	nan	nan
ma.L3.D.Close	0.0026	nan	nan	nan	nan	nan
ma.L4.D.Close	-0.0238	0.016	-1.477	0.140	-0.055	0.008
ma.L5.D.Close	0.0002	nan	nan	nan	nan	nan
ma.L6.D.Close	0.0123	nan	nan	nan	nan	nan
ma.L7.D.Close	-1.0008	nan	nan	nan	nan	nan
ma.L8.D.Close	0.9036	0.110	8.187	0.000	0.687	1.120

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0827	+0.0000j	1.0827	0.0000
AR.2	-116.3792	+0.0000j	116.3792	0.5000
MA.1	-0.9013	-0.4332j	1.0000	-0.4287
MA.2	-0.9013	+0.4332j	1.0000	0.4287
MA.3	-0.2179	-0.9760j	1.0000	-0.2850
MA.4	-0.2179	+0.9760j	1.0000	0.2850
MA.5	0.6199	-0.7846j	1.0000	-0.1436

```
MA.6            0.6199          +0.7846j            1.0000            0.1436
MA.7            1.0043          -0.0000j            1.0043           -0.0000
MA.8            1.1020          -0.0000j            1.1020           -0.0000
-----------------------------------------------------------------------------
```

B.2 Ether

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                D.Close   No. Observations:                  820
Model:                 ARIMA(2, 1, 9)   Log Likelihood                1102.909
Method:                       css-mle   S.D. of innovations              0.062
Date:                Mon, 29 Apr 2019   AIC                          -2181.818
Time:                        23:59:03   BIC                          -2125.306
Sample:                    01-09-2017   HQIC                         -2160.134
                         - 04-08-2019
=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
ar.L1.D.Close     1.6495      0.210      7.870      0.000       1.239       2.060
ar.L2.D.Close    -0.8362      0.204     -4.109      0.000      -1.235      -0.437
ma.L1.D.Close    -1.6392      0.203     -8.087      0.000      -2.036      -1.242
ma.L2.D.Close     0.8458      0.191      4.433      0.000       0.472       1.220
ma.L3.D.Close     0.0073      0.032      0.229      0.819      -0.055       0.069
ma.L4.D.Close    -0.0070      0.027     -0.259      0.796      -0.060       0.046
ma.L5.D.Close     0.0113      0.029      0.385      0.700      -0.046       0.069
ma.L6.D.Close    -0.0032      0.032     -0.101      0.920      -0.066       0.059
ma.L7.D.Close    -0.9982      0.028    -35.587      0.000      -1.053      -0.943
ma.L8.D.Close     1.6397      0.206      7.972      0.000       1.237       2.043
ma.L9.D.Close    -0.8421      0.187     -4.495      0.000      -1.209      -0.475
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            0.9863          -0.4723j            1.0936           -0.0711
AR.2            0.9863          +0.4723j            1.0936            0.0711
MA.1           -0.9016          -0.4330j            1.0002           -0.4287
MA.2           -0.9016          +0.4330j            1.0002            0.4287
```

MA.3	-0.2217	-0.9752j	1.0001	-0.2856
MA.4	-0.2217	+0.9752j	1.0001	0.2856
MA.5	0.6244	-0.7811j	1.0000	-0.1427
MA.6	0.6244	+0.7811j	1.0000	0.1427
MA.7	0.9675	-0.4890j	1.0840	-0.0745
MA.8	0.9675	+0.4890j	1.0840	0.0745
MA.9	1.0102	-0.0000j	1.0102	-0.0000

B.3 Litecoin

ARIMA Model Results

Dep. Variable:	D.Close	No. Observations:	820
Model:	ARIMA(2, 1, 8)	Log Likelihood	1027.005
Method:	css-mle	S.D. of innovations	0.068
Date:	Mon, 29 Apr 2019	AIC	-2032.011
Time:	23:38:11	BIC	-1980.208
Sample:	01-09-2017	HQIC	-2012.134
	- 04-08-2019		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1.D.Close	-0.7784	0.173	-4.506	0.000	-1.117	-0.440
ar.L2.D.Close	0.0328	0.037	0.894	0.371	-0.039	0.105
ma.L1.D.Close	0.7927	0.170	4.675	0.000	0.460	1.125
ma.L2.D.Close	0.0028	0.023	0.121	0.904	-0.042	0.048
ma.L3.D.Close	0.0185	0.019	0.985	0.325	-0.018	0.055
ma.L4.D.Close	0.0146	0.025	0.587	0.558	-0.034	0.063
ma.L5.D.Close	0.0197	0.019	1.036	0.301	-0.018	0.057
ma.L6.D.Close	0.0291	0.023	1.286	0.199	-0.015	0.073
ma.L7.D.Close	-0.9732	0.022	-43.998	0.000	-1.017	-0.930
ma.L8.D.Close	-0.7857	0.167	-4.702	0.000	-1.113	-0.458

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-1.2218	+0.0000j	1.2218	0.5000
AR.2	24.9627	+0.0000j	24.9627	0.0000

MA.1	1.0096	-0.0000j	1.0096	-0.0000
MA.2	0.6260	-0.7799j	1.0001	-0.1424
MA.3	0.6260	+0.7799j	1.0001	0.1424
MA.4	-0.2194	-0.9759j	1.0003	-0.2852
MA.5	-0.2194	+0.9759j	1.0003	0.2852
MA.6	-0.9034	-0.4336j	1.0021	-0.4288
MA.7	-0.9034	+0.4336j	1.0021	0.4288
MA.8	-1.2547	-0.0000j	1.2547	-0.5000

B.4 Bitcoin Cash

ARIMA Model Results

Dep. Variable:	D.Close	No. Observations:	617
Model:	ARIMA(2, 1, 8)	Log Likelihood	606.403
Method:	css-mle	S.D. of innovations	0.088
Date:	Mon, 29 Apr 2019	AIC	-1190.806
Time:	23:38:18	BIC	-1142.132
Sample:	07-31-2017	HQIC	-1171.882
	- 04-08-2019		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1.D.Close	0.4487	0.349	1.286	0.199	-0.235	1.133
ar.L2.D.Close	-0.0974	0.057	-1.723	0.085	-0.208	0.013
ma.L1.D.Close	-0.3061	0.348	-0.880	0.379	-0.988	0.376
ma.L2.D.Close	0.0004	nan	nan	nan	nan	nan
ma.L3.D.Close	0.0061	nan	nan	nan	nan	nan
ma.L4.D.Close	-0.0086	nan	nan	nan	nan	nan
ma.L5.D.Close	0.0007	nan	nan	nan	nan	nan
ma.L6.D.Close	-0.0026	0.015	-0.170	0.865	-0.032	0.027
ma.L7.D.Close	-0.9990	nan	nan	nan	nan	nan
ma.L8.D.Close	0.3090	0.346	0.892	0.373	-0.370	0.988

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.3026	-2.2272j	3.2035	-0.1224

AR.2	2.3026	+2.2272j	3.2035	0.1224
MA.1	-0.9018	-0.4322j	1.0000	-0.4289
MA.2	-0.9018	+0.4322j	1.0000	0.4289
MA.3	-0.2217	-0.9751j	1.0000	-0.2856
MA.4	-0.2217	+0.9751j	1.0000	0.2856
MA.5	0.6221	-0.7830j	1.0000	-0.1432
MA.6	0.6221	+0.7830j	1.0000	0.1432
MA.7	1.0000	-0.0000j	1.0000	-0.0000
MA.8	3.2364	-0.0000j	3.2364	-0.0000
