What Drives the Value of Cryptocurrencies ?

A Time Series Analysis of Bitcoin

Master Thesis in Finance

November 2017

Author: Gerasimos Kavvadias

Supervisor: Dr. Bertrand Melenberg

Master Thesis in Finance

What Drives the Value of Cryptocurrencies ?
A Time Series Analysis of Bitcoin

November 2017

Tilburg University

Tilburg School of Economics and Management

M.Sc. in Finance

Student Information:

Gerasimos Kavvadias

Anr: 232228

Snr: 2004937

Supervisor Information:

Professor Dr. Bertrand Melenberg

Professor of Econometrics and Finance

Tilburg University

Second Reader Information:

Associate Professor Dr. Lieven Baele

Professor of Finance

Tilburg University

Abstract

This thesis is making use of time-series regression analysis that follows the changes in Bitcoin's prices, and based on three clusters of independent variables, derives possible value drivers. The sample consists of daily observations for both the dependent and the independent variables between May 1, 2014 and June 12, 2017. The analysis is conducted both in daily and in weekly frequency. After dealing with stationarity and cointegration, ordinary least square regressions are used to determine possible short-run dynamics. For the long-run dynamics, vector error correction models are used. Furthermore, a generalized autoregressive conditionally heteroskedastic process is used as a side project (only in weekly analysis), to further understand the behavior of Bitcoin. Short-run analysis in daily frequency shows that Hash Rate (computational power) and the VIX index (uncertainty in the market) have a negative relationship with the performance of Bitcoin. Moreover, weekly short-run analysis shows that total Transaction Volume and Trends (Google Trends as a measure for attendance) have positive effects on the price, while Gold Price and VIX index have negative ones. Differences between time frequencies are discussed. Then, regarding the long-run analysis, Gold Price and Trends seem to have a negative impact on the price of Bitcoin, while proxies for a world market portfolio (S&P500, MSCI-world) have a positive one. The speed of adjustment to the long run equilibrium is relatively low for Bitcoin price. Regarding the side project, Bitcoin seems to follow a leptokurtotic distribution, while variance does depend on past day's variance. Last but not least, propositions for further research in the future are discussed.

Table of Contents

1. Introduction

Cryptocurrencies (or digital currencies) are online digital assets with monetary characteristics created to be used as a medium of exchange. As of November 10, 2017, cryptocurrencies had a total market capitalization of approximately $205 billion and a daily trade volume of more than $10 billion (coinmarketcap.com). While there are more than 1000 actively traded digital currencies, the six largest in terms of market capitalization (Bitcoin, Ethereum, Bitcoin Cash, Ripple, Litecoin, Dash) represent 86.9% of the total market (November 10, 2017).

Bitcoin (BTC) is the original cryptocurrency and was initially introduced to the public by Satoshi Nakamoto (2008). It was launched in early 2009 as a decentralized "electronic payment system", "a peer-to-peer version of electronic cash" without intermediaries, constructed with a cryptographic system that enables it to protect the identity of its users and is able to steadily recreate itself according to predetermined supply. Bitcoin dominates the market of digital currencies, representing 53% of the total market capitalization (November 10, 2017).

Unfortunately, even if Bitcoin appears to be an extraordinary phenomenon, the research literature addressing it was limited. In the early stages of its existence, it was primarily researched from a computer-science perspective, with regard to its inner structure and the technical procedures relevant to its functioning, (Segendorf, 2014; Barber et al, 2012), and its legal point of view, discussing the legal problems that could possibly arise in the future, for the users and the system itself (Murphy et al., 2015; Brito et al, 2013). Finance and economics related papers were scarce in the first four to five years of its existence. It is interesting to see that, for example, Eric Schmidt, the CEO of Google in 2013, stated that he was unaware of its existence at that time (Wikileaks, 2013).

The situation drastically changed in the past three to four years, when a number of papers relating to the finance and economics of Bitcoin were published and attempted to understand the existence and behavior of its market (Yermark, 2015; Ciaian et al, 2016). The Bitcoin market is highly inefficient and extremely volatile (Gronwald, 2014).

Even if Bitcoin is used as a currency in more than 100 thousand online stores (Brandom, 2014, Franzen, 2014), the ongoing debate addresses whether Bitcoin is used as an investment asset more than as a payment system. Glaser et al. (2014) find evidence to support the case of Bitcoin as an investment asset. This thesis intends to treat Bitcoin more as a speculative investment asset and less as a currency.

This research aims to use time series regression analysis and, based on three clusters of independent variables, investigate whether those variables are value factors for Bitcoin's price and whether they have predictive power. Technological variables represent the possible internal determinants, financial variables represent the possible external determinants, and proxies for trends represent the general public attention. The sample used consists of daily observations from May 1, 2014 to June 12, 2017 - or 1139 daily observations or 163 weekly observations. Both short-run and long-run determinants are discussed. After addressing stationarity issues, we use ordinary least squares (OLS) for both daily and weekly analyses to establish short-run relationships. Then, after dealing with cointegration, vector error correction models (VECM) are used to determine possible long-run relationships. Finally, in a side project and in an attempt to further understand the mechanics of Bitcoin, we search for dynamic forms of heteroskedasticity and model the variance.

This chapter presents an introduction to the chosen research topic and a general display of the remaining chapters. The remainder of the research is comprised of seven more chapters. Chapter 2 discusses the current relevant state of literature. Chapter 3 covers the sample description, data collection and manipulation, and the summary statistics. In chapter 4, the main research question and relevant hypothesis are developed. Chapter 5 includes the methodology analysis. In chapter 6, results and evidence from the research - and based on the methodology used - are discussed. Chapter 7 presents a short conclusion and propositions for further research. Finally, appendices are presented in chapter 9.

2. Current State of Literature

In this part of the research, the current state of relevant literature is discussed and other relevant findings are briefly addressed.

2.1. Internal determinants

Firstly, DeLeo et al (2014) used OLS regressions based on weekly data to support a positive relationship between transaction volume and the performance of Bitcoin. On the contrary, Ciaian et al (2014, 2016) showed a negative long-run relationship between the number of transactions and the price of bitcoin. Furthermore, regarding internal determinants of bitcoin, Bouoiyour et al (2014, 2015) used ARDL analysis to support a positive correlation between hash rate and bitcoin’s price.

Then, Li et al (2016) researched the determinants of exchange rates in the case of bitcoin, and found that in the long-run analysis in the early stages of the market, the total number of bitcoins in circulation and the trading volume are possible determinants. They also found evidence for a long-run relation between mining difficulty and the exchange rate in the early market. Polasik et al (2015) also determined a relation between the traded volume and the performance of bitcoin.

Finally, Garcia et al (2014) use different possible determinants (socio-economic signals) and VAR methods to detect bubbles in the market of bitcoin.

2.2. External determinants

Even if Bitcoin is far more volatile and seems to have significant internal differences compared to gold, its movement is only dependent on the market rules of supply and demand. Similar to gold, bitcoin is not linked to any underlying asset. Dyhrberg (2016) demonstrated similarities between bitcoin and gold in their hedging capabilities and Yermack (2015) strengthened the belief of the use of bitcoin as an investment vehicle. On the contrary, Bouri e al (2017) showed no evidence of possible hedging and safe haven properties against other assets.

Ciaian et al (2016) used VAR to capture effects between possible determinants as the Dow Jones Index, the exchange rate (XBT/USD) and oil price and bitcoin's price. Then, Baek e al (2015) examined bitcoin's market volatility compared to the volatility of S&P500 index and showed its extreme volatility, while they concluded that bitcoin is only influenced by its internal drivers (demand and supply) and not by fundamental economic factors.

2.3. Public attention

Kristoufek (2013), a pioneer in pairing proxies for public attention and the performance of bitcoin, used weekly data and quantified through an error correction model the relationship of Google Trends and Wikipedia Views with bitcoin's price, resulting in a positive relationship. An interesting part of the research was the fact that it was crucial to distinguish between interest due to positive or negative events through those trends. Then, Kristoufek (2015) used daily analysis and a model with more possible determinants to show wavelet coherence between bitcoin's performance and the popularity measured by search engine queries. Even if the relationship was changing over time, it showed evidence of a co-movement. Subsequently, many had included search engine queries as measures for the public attendance or a certain momentum effect.

Furthermore, Buchholz et al (2012) used VAR regressions and showed a positive relationship between Google Trends and the number of transactions of bitcoin (as a proxy for the demand of bitcoin). Then, Georgoula e al (2015) introduced the sentiment ratio of Twitter users concerning bitcoin as a possible measure for attendance, showing a short-run positive effect between the two. Moreover, Polasik et al (2015) demonstrated a relation between attendance/interest through Google searches and bitcoin.

2.4. Other relevant literature

Another challenging part in the analysis of bitcoin's performance is to attempt to model dynamic forms of heteroskedasticity, and to ascertain the determinants of its variance. Glaser et al (2014) used ARCH and GARCH analysis to determine whether bitcoin behaves more like an investment asset or more like a currency. As mentioned before, they found evidence supporting the first case. Moreover, through their analysis, they found a positive and significant relation between the traded volume and the exchange rate, while they found a negative and significant relation between the traded volume and the price. Katsiampa (2017) then used different GARCH models to find the optimal conditional heteroskedasticity model based on goodness of fit terms. The paper also supports the leptokurtic behavior of bitcoin.

Likewise, Gronwald (2014) used GARCH analysis to argue that high volatility and extreme price movements characterize bitcoin, and subsequently show that the Bitcoin market is still immature and at an early stage. In addition, Urquhart (2016) suggested that even if the bitcoin market is not weakly efficient in the full sample researched (August 2010-July 2016), the inefficiency of the market is significantly strong. Finally, Dyhrberg (2016) used an asymmetric GARCH model that showed evidence in favor of the use of bitcoin as a risk management tool, or as a possible alternative investment for risk-averse investors.

Taking all of the above into account, we intend to use new possible determinants in different analyses and with the latest available data to contribute to the research of Bitcoin's performance and cryptocurrencies in general.

3. Data Collection and Manipulation

3.1. Introduction

This section provides an insight into the data used for this empirical analysis. The data used in this research covers the period between May 1, 2014 and June 12, 2017. In order to collect the data for the analysis, both public and non-public sources were used to gather daily observations for the dependent and independent variables. In the analysis, the daily Price of Bitcoin is used as the dependent variable and three clusters of variables represent the independent variables. The first cluster is relevant to the technical aspect of Bitcoin and contains daily prices on the Transaction Volume, the Hash Rate and Bitcoins in circulation (supply). The second is relevant to the financial aspect of Bitcoin and will contain daily prices of Gold, the MSCI World USD-denominated index and the S&P 500 index as proxies for the World Market Portfolio, and the VIX index. Finally, the third cluster contains indicators for a momentum effect and the general intention of traders by using Google Trends and Wikipedia page views.

Most of the relevant papers in the early existence of Bitcoin, and cryptocurrencies in general, used data from earlier points in its existence. Most of the times they used price data and transaction volumes from 2010 to 2014. That was due to the fact that from July 2010 (limited data in the early stages) to February 2014, Mt. Gox, the leading bitcoin exchange, was handling approximately 84% of the total bitcoin transactions. In February 2014, Mt. Gox filed for bankruptcy due to a scandal surrounding the theft of more than $400 million of bitcoin (Takemoto and Knight, 2014). Furthermore, most of the papers include one of the largest bubbles in bitcoin history: a Boom and Bust story that ended with more than 60% losses on bitcoin value and lasted from November 25, 2013 to April 7, 2014 (Cheung et al., 2015). This thesis aims to use more recent data for major aspects of bitcoin's "life". One of the main goals is to observe the outcome following those two events, and how bitcoin subsequently evolved through its determinants (explanatory variables).

A portion of the data gathered will come from publicly available online sources. An argument in favor of the validity of the data arises directly from the nature of bitcoin and the profile of its users. The Bitcoin market is a highly inefficient market and it seems to be relatively safe to assume that the behavior of its users is strongly inelastic and weakly tolerates fraud or intended mispricing (Mt. Gox, 2014). A second argument could arise from the fact that the majority of early papers were not diversified with regard to price and data collection, meaning that they focused on the same time span, as they were based on a single exchange market and the time in which it operated (Mt. Gox). Finally, we have to state that similar data, based on their source, are used heavily in the relevant literature.

For the full dataset, both daily and weekly analyses are used. The daily analysis is based on 1139 observations, while the weekly analysis is based on 163. On the weekly analysis, the closing prices on each Friday for each of the variables are used to create the dataset.

Every graph regarding the transformations of both the dependent and the independent variables is reported in Chapter 9 (Appendix, Charts A). In addition, a short description of the set of variables used in the research is reported in Chapter 9 (Appendix, Table A).

3.2. Dependent Variable

Bitcoin's Price

The market price of a single bitcoin in USD is used as the dependent variable. Both the dependent and the independent variables will be transformed, and the natural logarithms will be taken. This helps to overcome the problem of high skewness and kurtosis that some of the variables have. Additionally, by taking the natural logarithms, we are able to solve the issue of having many outliers in some of the variables. Finally, we have to state that this transformation will enable us to use the same metric system (through normalization), making the analysis of the relations more intuitive.

If we were analyzing an aspect of the stock market, it would be common to assume that returns are normally distributed. Even if it is convenient to assume that bitcoin is log-normally distributed, we have to remember that bitcoin and a common stock, for example, differ significantly. In our case, even if our observations are daily, we have to bear in mind that bitcoin is traded 24 hours per day, 7 days a week, meaning a roughly 3.4 times more trading time than a regular stock. A day in bitcoin's "life" is equal to approximately 3.4 days in trading time for a common stock; this widens the gap between the observation points. One last argument is that high returns are expected for a highly volatile and inefficient market such as bitcoin is.

Another seemingly similar approach, instead of using the price of a single bitcoin and its corresponding return as the dependent variable, is to use the total market capitalization of bitcoin. We avoid using the total market capitalization of bitcoin because the structure of bitcoin itself has prevented/unabled many users that lost their digital wallets or the code to access them to trade the bitcoins at their disposal. That means that an immeasurable fraction of the total supply is not tradable, creating a bias in our data. The aim is to better understand the movement of the price linked to the independent variables.

The Price of bitcoin will be based on the Bitcoin Price Index (BPI) (coindesk.com), an Index of the exchange rate between bitcoin and USD. The BPI shows the average USD market price across major bitcoin exchanges and, more accurately, the weighted average of USD trading prices of Bitstamp, Bitfinex, Coinbase, itBit, OKCoin, and ultimately represents more than 39% of the total bitcoin exchange volume (bitcoincharts.com), in an effort to estimate the “fair” value of bitcoin.

From the total sample of 1139 daily observations, 7 where missing or destroyed, representing 0.6% of the sample. We used linear interpolation to accommodate the problem.

3.3. Independent Variables

3.3.1. Hash Rate

The Hash Rate is an estimation of the daily number of tera-hashes per second (computational power), and it is an indicator for the total processing power used in the market. It is an absolute measure of the total computational power that is used in the market in a single day during the mining process. For security reasons and in an effort to keep bitcoin consistent and unaltered, the network has to conduct continuous mathematical operations. One hash rate or 1 Th/s (tera hashes per second) means that the system makes one (1) trillion calculations per second. The network itself creates new bitcoins (supply) with a predetermined function. Every new group of transactions in the network (block) is followed by a cryptographic hash of each of the previous created blocks extending all the way back to bitcoin’s birth, using the SHA-256 hashing algorithm. This exact algorithm is the link between the previous blocks and the future ones. Each time the network wants to create new bitcoin, it is reproducing the same method, but each time it needs more computational power and more time due to the higher difficulty. It takes roughly 2016 blocks (or approximately 14 days) to adjust the difficulty level based on the network performance and with respect to keeping the average time between new blocks at approximately ten minutes.

Blockchain, a free database for the bitcoin market, is our basis for data related to the technological aspects of bitcoin and, subsequently, Hash Rate (blockchain.info). We collect daily data for the same time interval mentioned above. Blockchain allows us to collect daily data for the last two years of a variable relevant to bitcoin, and from then on we have access to data every second day. It uses the same approach to every variable that it counts. Thus, even if we have accurately gathered data every day for the past two years, we still need to find a suitable way of interpolating the rest of the “every other day” data in the beginning of the dataset. Based on the visual representation of the hash rate with respect to time and the core meaning of the hash rate itself, meaning that it needs to be trended and time-dependent to ensure the security of the network, we conclude that linear interpolation would be an inaccurate representation of the series. Using the cipolate (2002) command in STATA (installed from SSC) to perform cubic interpolation, we manage to use a cubic polynomial interpolation method that is not discontinuous, has the tendency to be constrained to join smoothly, and subsequently appears to better fit the hash rate series. The algorithm used fits interpolated values on a cubic curve to two data points before and two data points after each observation, filling the respected variable. Having data in the range of our sample made us to not use an extrapolation method in creating data from beyond the range.

3.3.2. Transaction Volume

The second technical aspect of bitcoin is the total Transaction Volume of bitcoin per day. This variable will represent the number of daily confirmed bitcoin transactions. To further understand the origin of this variable and the way that it is measured we need to state that downloading and installing relevant mining software will provide any user access to information regarding the transaction volumes. Therefore, a user with a common bitcoin wallet is able, through the mining software, to download the full blockchain, which is able to reflect every bitcoin transaction that ever took place.

Blockchain (blockchain.info) will provide the number of daily transactions based on the method described above, having the same benefits and drawbacks discussed regarding the previous variable. Daily observations for the time interval between May 1, 2014 and June 12, 2017 will be collected. For the same reasons as mentioned with respect to the previous variable, we decided to use a cubic polynomial interpolation method (cipolate) to address the problem of the missing observations at the early stages of the time interval. The parameter choice and use of the algorithm are the same as for the Hash Rate case. In this way, almost 17% of data are interpolated for the first year of the data.

One argument against the use of that specific variable as an explanatory variable in our model could arise from the fact that as every transaction that takes place is counted, transactions that have the same person as buyer and seller from different wallets (bitcoin trades) are also counted in the total. It is reasonable to assume that this is a minority in the total transaction volume. In an effort to further weaken this argument, we can assume that those transactions aim to restructure or reorganize someone's exposure to bitcoin.

3.3.3. Supply of bitcoins (bitcoins in circulation)

Bitcoins in circulation is the third technological aspect, representing the sum of all bitcoins that have already been mined (counted in millions) and will serve as a proxy for the supply of bitcoin. Bitcoin has an absolute predetermined limit of approximately 21 million units that it would be reached close to 2140. The graph of total supply of bitcoins on time demonstrates an interesting aspect of that variable. When eventually a new block is decoded, a miner is rewarded with new bitcoins (and transaction fees). The structure of the regeneration of bitcoin rewards the same amount of new bitcoins to the miners that manage to solve each block. This sequence continues for about four years and then, due to its structure, miners are consequently rewarded with half of the bitcoins they used to. In July 9, 2016, we observe this exact phenomenon, with the slope of the line changing radically.

Once more, Blockchain (blockchain.info) will be the database we use to download the daily observation with the same advantages and disadvantages as we had until now. The only drawback is that the text data (.csv, comma delimited data) are now poorly updated and completely absent in some of the cases for more than 40% of the time that we research. In order to download the data in the same way as for Hash Rate and Transaction Volume, we absorb the javascript code from the graph that contains the recommended data in .json form (JavaScript Object Notation, as a lightweight data-interchange format). Then we use the free online converter convertcsv.com/json-to-csv.htm to convert the .json file to a standard .csv file. Finally, we generate a counter, to determine whether observations are in line with the date and the same as the observations from the damaged original .csv file. Then we continue the standard process of manipulating the variable.

In this instance, we use the command pchipolate (2012) in STATA (installed from SSC) to perform a piecewise cubic Hermite interpolation. The default algorithm based on Moler (2004) is used. We base that choice on the time-trended series of supply of bitcoins and the fact that we want it to be accommodated with an interpolation method that will create an interpolant that is shape-preserving and cannot overshoot locally. This time our goal is to preserve the shape of the series and create a smooth continuous function. Technically, the piecewise cubics will afford us the flexibility to smoothly continue the process without creating any new variability. The extrapolation method will not be used here either.

3.3.4. Gold Price

The daily natural logarithm of the price of gold will be the first economic relevant variable, and will be used due to the nature of bitcoin. Bitcoin is far more volatile than gold but its movement only depends on the market rules of supply and demand (similar to gold), based on the fact that bitcoin is not linked to any underlying asset. Dyhrberg (2016) showed similarities between bitcoin and gold in their hedging capabilities and Yermack (2014) strengthened the belief of the use of bitcoin as an investment vehicle.

Datastream will provide us with daily gold prices for the requested period. The gold price will be calculated as the average between the London Gold Fixing Companies, from the London Bullion Market Association (LBMA) (Gold Bullion LBM U$/Troy Ounce), and the Handy & Harman reported daily price (Gold, Handy & Harman Base $/Troy Oz). Gold price, as bitcoin price, will be USD denominated and will measure $ per troy ounce.

3.3.5. World Market Portfolio

As proxies for a well-diversified global portfolio, the MSCI World USD-denominated price index and the S&P 500 price index will be used separately. The prices of those price indices will be used as our second economic-relevant variable.

Datastream will provide us with daily data for both price indices that will represent our World Market Portfolio (MSCI WORLD U$ - PRICE INDEX, S&P 500 COMPOSITE - PRICE INDEX).

3.3.6. VIX index

The Chicago Board Options Exchange’s (CBOE) volatility index (also known as the VIX index) will be our third finance-related independent variable. The VIX index measures the implied volatility of the S&P 500 index options and will represent market risk or, more accurately, a measure for uncertainty in our model. The VIX index if often called the “fear index”.

To download the VIX index for the requested time period, both the official CBOE’s site (http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data) and quandl.com were used to verify the validity of the data gathered.

3.3.7. Assumptions Following the Second Cluster

The problem with the analysis between the dependent variable and the cluster of economic related variables is the way in which they are traded to the public. As mentioned earlier, bitcoin is traded all day, every day, and 365 days a year. On the other hand gold, the two proxies for the world market portfolio (MSCI World index and S&P 500 index), and the VIX index are traded with the usual frequency: every weekday except weekends and some holidays to approximately 250 days a year, as a simple equity such as, for example, a stock, is traded. This creates an inconsistency in our data between dependent and independent variables. However, we will attempt to determine the correlation and causal relationship using both daily and weekly analysis. In the case of weekly analysis, we will not face any problem relevant to when each variable is traded.

In the case of daily analysis, we will use two different approaches in an attempt to address the problem. In the first approach, we will assume that throughout the weekends traders believe that nothing has changed relevant to each of our variables, and will continue to operate as though Friday’s closing price is still the ‘real’ market price until Monday, and the beginning of a new trading day. Even if this approach is highly unlikely to be accurate, we need to mention that it is common practice when addressing similar ‘holiday’ issues in datasets and relevant literature. We are going to use reasoning close to the near neighbor interpolation method, and we will assume that Friday’s closing price stands for Saturday and Sunday. Variables that follow this assumption will be followed by the suffix _nn.

In the second approach, we assume that our traders are sophisticated individuals with financial knowledge, and that they anticipate Monday’s closing price, meaning that we assume that for them the price tends linearly to approach Monday’s price. To show that tendency, we use linear interpolation to express that move from Friday’s price to Monday’s one. Variables that follow this assumption will be followed by the suffix _linear.

3.3.8. Measures of Trends

The third cluster of independent variables will attempt to capture the momentum effect and will indicate extra attention to the bitcoin market. In order to do so, we will use two different measures of trends separately, views on Google Trends and views on Wikipedia. In both of the variables, the word used as the search term was “Bitcoin”.

Google Trends shows us the normalized number of search queries in the Google search engine for a given location at a given time (trends.google.com). Google Trends can be downloaded in a daily form of up to 90 days of data and that is why we will use weekly data based on Google Trends, and further use that variable as our measure for trends in the weekly frequency analysis. Thus, weekly data from May 1, 2014 to June 12, 2017 based on worldwide observations (location) will be used. It is important to mention that Google Trends are not an absolute measure but a relative measure of attendance. In Google Trends, each data point is divided by the total searches of the geography and time range they represent, to compare relative popularity. The resulting numbers are then scaled on a range of 0 to 100, with the value of 100 being the peak popularity for the period requested and region chosen, and 0 being the lowest point in the query. Thus, a point with the value of 10 is 10% as popular as the peak in the specified period.

Thereafter, we use the number of daily searches in Wikipedia (search term: Bitcoin) as our second measure of trend. Wikipedia series are the actual number of daily queries with a specified search term on a given day. They could be an absolute measure of popularity. We use daily data from July 1, 2015 to June 12, 2017 (tools.wmflabs.org/pageviews) and the official tool from Wikipedia pageview analysis.

The initial intention is to use different measure of trends to the daily and weekly frequency analysis.

3.4. Summary Statistics

In this section, we take the first step in the statistical analysis of our sample. Tables I and II present summary statistics for each of our variables, before and after the natural logarithm transformation respectively. In both tables, each variable's number of observation, the mean as a central tendency measure, standard deviation as a dispersion measure, skewness and kurtosis as measures for the shape of each distribution, and the minimum and maximum values of each variable to provide an extra indication of the range, are reported.

Table I

Summary Statistics (Original)

This table reports descriptive statistics for the dependent and independent variables mentioned in the Data Collection and Manipulation section. Price represents the price of a single bitcoin in USD. Traded volume is the number of daily confirmed bitcoin transactions. Hash Rate is the daily number of tera-hashes per second (computational power) expressed in millions. Supply is the number of bitcoins in circulation expressed in millions. Gold Price is a "fair" gold price. S&P 500 and MSCI world are price index for the WMP. VIX index is the CBOE's volatility index. Wikipedia Views is the daily measured proxy for trend. Google trend is the weekly measured proxy for trend. The data are over the period May 1st, 2014 to June 12th, 2017. *N* denotes the number of observations and Std Dev the standard deviation.

Variables	*N*	Mean	Std Dev	Skewness	Kurtosis	min	max
Price	1,139	564.4	411.1	2.737	12.78	177.3	3,019
Traded Volume	1,139	170,083	82,671	0.315	1.882	47,730	369,098
Hash Rate (m)	1,139	1.182	1.203	1.423	4.363	0.0497	5.688
Supply (m)	1,139	14.79	1.111	-0.225	1.724	12.72	16.39
Gold Price	813	1,219	71.86	-0.275	2.434	1,051	1,367
S&P500	813	2,095	135.2	0.677	2.983	1,829	2,439
MSCI WORLD	813	1,718	80.52	-0.0301	3.599	1,469	1,935
VIX index	785	15.10	4.013	1.738	7.086	9.750	40.74
Wikipedia Views (daily)	713	12,311	9,793	4.767	30.48	5,760	91,099
Google Trends (weekly)	163	16.39	11.70	4.180	23.95	9	100

As expected and presented in Table I, price is highly volatile, with a standard deviation close to the mean observation and a significant gap between the minimum and the maximum observation. Following Cont's (2001) stylized facts and statistical issues (mainly for log-returns), it appears that bitcoin's price, log-price and log-returns share some relevant features with financial variables. The positive skewness (more than 2 but less than 5) and the positive kurtosis create a heavy tail in bitcoin's price distribution and that is a common stylized fact for a financial asset. Furthermore, from the range of the observation, we can safely state that we have some extreme values that may result from the heavy-tailed distribution. Additionally, the two variables that are used as attention-trend indicators seem to have high variability. On the other hand, the four financial related variables have more common attributions with the technological related ones than expected. Nevertheless, the financial variables appear to more accurately reflect the stylized facts of Cont (2001).

Table II

Summary Statistics (Natural Logarithm)

This table reports descriptive statistics for the dependent and independent variables mentioned in the Data Collection and Manipulation section. Every variable is transformed in natural logarithm form (_ln). Both the "naïve" and the "sophisticated" assumptions are represented with the suffixes _nn and _linear respectively. Those assumptions are relevant only to the daily analysis and not the weekly analysis. The data are over the period May 1st, 2014 to June 12th, 2017. *N* denotes the number of observations and Std Dev the standard deviation.

Variables	*N*	Mean	Std Dev	Skewness	Kurtosis	min	max
ln_price	1,139	6.159	0.559	0.708	3.254	5.178	8.013
ln_TV	1,139	11.91	0.532	-0.224	1.733	10.77	12.82
ln_HR	1,139	13.44	1.088	-0.0276	2.076	10.81	15.55
ln_S	1,139	16.51	0.0760	-0.303	1.786	16.36	16.61
ln_GP	813	7.104	0.0596	-0.397	2.523	6.957	7.221
ln_GP_nn	1,139	7.104	0.0597	-0.390	2.512	6.957	7.221
ln_GP_linear	1,139	7.104	0.0595	-0.392	2.511	6.957	7.221
ln_SP500	813	7.645	0.0634	0.530	2.839	7.512	7.799
ln_SP500_nn	1,139	7.645	0.0634	0.543	2.841	7.512	7.799
ln_SP500_linear	1,139	7.645	0.0635	0.539	2.838	7.512	7.799
ln_MSCI_WORLD	813	7.448	0.0470	-0.215	3.699	7.292	7.568
ln_MSCI_WORLD_nn	1,139	7.448	0.0470	-0.196	3.677	7.292	7.568
ln_MSCI_WORLD_linear	1,139	7.448	0.0471	-0.199	3.669	7.292	7.568
ln_VIX	785	2.685	0.234	0.988	3.778	2.277	3.707
ln_VIX_nn	1,139	2.683	0.237	0.965	3.593	2.277	3.707
ln_VIX_linear	1,139	2.685	0.235	0.995	3.740	2.277	3.707
ln_WikipediaViews	713	9.282	0.442	2.013	8.252	8.659	11.42
ln_GoogleTrends_weekly	163	2.676	0.422	2.017	7.770	2.197	4.605

In the second table, where we use the natural logarithm for each of our variables and both of our assumptions about the "naïve" (_nn) and "sophisticated" (_linear) traders, we can easily observe a somewhat different result. Indeed, all series seem to have been corrected as far as the shape of their distribution is concerned (skewness and kurtosis). The differences between the original variables and those from the two different assumptions initially do not seem to be significant. Finally, as expected, part of the variability caused by outliers, especially in the dependent, financial, and attendance variables, is accommodated.

4. Main Research Question and Hypotheses Development

The main idea of this thesis is to build up an econometric model and to use time series regression analysis that describes the changes in Bitcoin's prices, and based on three clusters of independent variables, investigate whether those variables are value factors for Bitcoin and whether they have predictive power.

The goal of the thesis will be to utilize the current relevant research and extend its results to make a positive step in understanding the behavioral and financial mechanics of bitcoin and the way users trade it. The first cluster will contain variables relevant to the technical aspects of bitcoin (internal effects), the second will contain finance related variables (external effects), and the third cluster will include to the model those variables that will be used as the momentum-attendance factor in the analysis of bitcoin (trends). The idea is to include relevant variables that research until now has shown that influence the movement of bitcoin and, including possible relevant new ones, estimate whether our model gains predictive power using the latest data. The goal is to use different econometric techniques and through them understand the mechanics of bitcoin. Both short-run and long-run dynamics will be researched in different time frequencies (daily and weekly).

Hypotheses that possibly determine the relationship between the dependent and the independent variables:

First cluster of independent variables:

A) Transaction Volume

We expect that as traded bitcoin volume will rise due to the increase of the participants and the general attention that bitcoin's community has, from not only its most loyal users but also from the basis of the trading world, prices will follow.

B) Hash Rate

The expectation here is that when the computational power (processing power) needed from the market raises, bitcoin prices tend to follow a different way. We intend to use Hash Rate as a measure of difficulty in the system, as a systemic drawback in the process. To avoid misunderstanding, Hash Rate is vital for different aspects of the system, as for example someone could support that it is a measure for security.

C) Supply (bitcoins in circulation)

We intend to use the total number of bitcoins already mined and in circulation as a proxy for the supply of bitcoin. The system itself creates its own steady, finite and predetermined inflation and we expect that this fact will create some kind of complications in the system (inflation effect).

Second cluster of independent variables:

D) Gold Price

It seems that bitcoin is significantly more volatile than gold, but its movement only depends on market rules of supply and demand, based on the fact that bitcoin, just like gold, is not linked to any underlying asset. Even if they are notable differences between the two, we expect to see evidence of a complementary relationship.

E) World Market Portfolio

We expect that our proxies for a well-diversified global portfolio (S&P500 index and MSCI-world index) will further support the general perception and the use of bitcoin as an investment asset, and not exclusively as a currency. The goal is to show that bitcoin is a substitute for the ordinary investment choices that the proxies will represent. Subsequently, we expect to see that whether there is a downswing in the market, bitcoin thrives.

F) VIX Index

VIX Index will represent the uncertainty in the market; it will be the sense of fear in our model. We expect that when the general "real" economy is on turmoil, bitcoin will have the tendency to follow a completely different way.

Third cluster of independent variables:

G) Measures of Trends (Wikipedia Views and Google Trends)

We expect to find evidence that the closer the bitcoin is getting to the public, the more its price will reflect the acceptance from a larger number of participants. As it walks the road from anonymity to "trend" in the daily news, it will actively and positively absorb this kind of attendance.

All of the above will be the fundamental questions in this research. That does not mean that this thesis will not try to understand and explain bitcoin behavior with different methods and in different circumstances. We will investigate all of the above in different time frequencies to see whether differences are occurring. Short-run and long-run dynamics might support or change part of the hypotheses. Even a result that states no relation with our dependent variable will have a lot to offer to the research. Finally, we will endeavor to understand the high volatile nature of bitcoin.

5. Analysis-Methodology

5.1. Introduction

This chapter presents the methodology used and the results produced to answer the main question and each of the sub-hypotheses in this research. The goal is to use quantitative methods that will help us to further understand the behavior of bitcoin. Based on the nature of our data, a multivariate time series regression analysis will be used. The starting point in this analysis is the following model:

$$ln\_Price_t = a_0 + \beta_{TV} ln\_TV_t + \beta_{HR} ln\_HR_t + \beta_S ln\_S_t$$
$$+ \gamma_{GP} \mathrm{ln\_GP_t} + \gamma_{WMP} \mathrm{ln\_WMP_t} + \gamma_{VIX} \mathrm{ln\_VIX_t}$$
$$+\delta_{Trend} ln\_Trend_t + u_t \qquad (1)$$

Here, Price is the dependent variable, and represents the price of a single bitcoin in a given day, in USD. The independent variable TV represents the total transaction volume of bitcoin per day, HR is the daily number of tera-hashes per second (computational power), S is the total supply of bitcoins (bitcoins in circulation), GP is a “fair” gold price, WMP is the world market portfolio and is represented by the S&P 500 price index and the MSCI world USD-denominated price index separately, VIX is the CBOE’s volatility index (VIX index), and, finally, Trend is an independent variable that will capture the momentum effect or the extra attention to the bitcoin market, and will be represented by Google trends in the weekly analysis and Wikipedia views in the daily analysis. Both a daily (1139 observations) and a weekly (163 observations) analysis will be conducted. The dependent and the independent variables will be used in their natural logarithm form based on the arguments stated in the Data Collection and Manipulation section. The goal is to find indications regarding both short-run and long-run dynamics between the variables.

Different letters are used in the coefficients to simply underline the different clusters of independent variables, the technical aspect (β), the financial aspect (γ) and the momentum-attendance effect (δ).

5.2. Stationarity-Weakly Dependence

A time series is called covariance stationary when its probability distribution is stable over time, it has a constant mean, a constant variance, and the covariance of two different points in the series depend only on the step between them. On the other hand, a weakly dependent time series demonstrates correlation between observations across time that is “not too strong”. Even if we know that it is possible to proceed with non-stationary series, in order to use standard time series regression analysis in this research, we need to know whether a variable is stationary or not. If we have non-stationary series we cannot do hypothesis testing (in our case) due to the fact that t-statistic will not have approximately a standard normal distribution, and standard errors will converge to zero as a result of time, possibly creating a spurious regression. This will result in inaccurate relations. The statement above is regarding the first part and the short-run dynamics (standard OLS regressions), because on the second part (long-run dynamics), non-stationary series will be used for the cointegration test and the VECMs.

To avoid that problem we need to test for stationarity. The non-stationary series are said to be integrated of order k (or I(k)) when taking k times their differences will transform them to stationary or integrated of order zero (or I(0)) series. Most of the non-stationary finance-related variables tend to be I(1) series that need the first difference to be used in a regression analysis. A non-stationary series could be a unit root process, which means that it has a stochastic trend, or a trend-stationary process, which means that it follows a deterministic trend. The goal is to distinguish between the two categories and transform the variables accordingly. In the case of unit roots, differences will transform our series to stationary. For trend-stationary series, de-trending will be used to erase the deterministic trend and make the series stationary. It is also important to state that because at least some of our variables are finance-related, we expect time, as natural ordering in the data, to create trends.

Knowing that stationarity tests are not the most powerful tests, and sometimes it is possible to produce contradictory results, we decide to use more than one to further validate the results. The most commonly used test is the augmented Dickey-Fuller (ADF, 1979) that has the null hypothesis of a unit root presence and the alternative of stationarity. An alteration of the ADF test, the Phillips-Perron (PP, 1988) test, will be the second stationarity test, where again the null hypothesis is the existence of unit root process and the alternative is that the variable is stationary. The PP test uses Newey-West standard errors to accommodate serial correlation and heteroskedasticity in the errors. Finally yet importantly, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS, 1992) test will be used. The KPSS test tests the null hypothesis that a time series is trend-stationary (has a deterministic trend) against the alternative of non-stationarity and unit root.

All variables both in the daily and the weekly analysis will be investigated. Therefore, stationarity tests will be made not only for the dependent and independent variable but also to every variation of them that is going to be used in the regression models. Both the “naïve” (_nn) and the “sophisticated” (_linear) approaches on the cluster of economic related variables will be checked.

5.3. OLS Regression Models

Having transformed to stationary process every variable, whether dependent or independent, we will use Ordinary Least Squares (OLS) regression method to estimate the parameters of the model and derive possible statistical and economic significance of the coefficients, and subsequently answer part of the research questions. To check the validity of the results and further correct the models we conduct a series of diagnostic tests, which will both be performed for daily and for weekly data.

Firstly, we check whether errors (u_t) are serially correlated. Finding that errors in two different periods are correlated will lead to inefficiency (biased standard errors) and inconsistency of the estimated coefficients (in some cases). The Breusch-Godfrey test (1978) with a null hypothesis of no serial correlation in the residuals and the alternative of autocorrelation will be used. In addition, the Durbin-Watson test (1950, 1951) will be used to decide whether our model suffers from autocorrelation or not. The Durbin-Watson test has the same null and alternative hypothesis as the Breusch-Godfrey test. Finally, to graphically represent the situation we will use the Bartlett's periodogram-based test, whereas the null is the white-noise process of uncorrelated variables, avoiding serial correlation. Nevertheless, even if all unit root processes show evidence of serial correlation, not all serial correlated series will have unit root processes. While it is unlikely that a model with many first-differences demonstrates serial correlation, in case we do, we will proceed by estimating serial-correlation-consistent standard errors (Newey-West standard errors) and make an effort to model serial correlation (possibly by including lagged dependent variables in the model).

Then, as we solve the problem of serial correlation, we proceed by testing the homoskedasticity assumption, meaning that we will use the Breusch-Pagan test (1979) to check whether our residuals are homoskedastic or heteroskedastic (having constant and finite variance for any explanatory variable at the same time period or not). In the case of evidence in favor of heteroskedasticity, we expect biased standard errors, and subsequently t-statistics, and we will report both the original regressions and regressions using heteroskedasticity robust standard errors. The last diagnostic test regarding the u_t will be the Shapiro-Wilk test on the residuals of the regression we estimated. This tests the assumption in OLS regressions that residuals should be normally distributed. Last but not least, we have to mention that the correlation matrix of the variables that are going to be used in the regression analysis will be examined to avoid overlooking the possibility of multicollinearity between two or more independent variables.

5.4. Dynamic Forms of Heteroskedasticity – A GARCH (1,1) model

Having obtained evidence of heteroskedasticity from the previous section, we are interested in its dynamic forms. While this section does not intend to answer one of the main research questions of the thesis, it is crucial for us to understand, for example, what determines the variance of our model and subsequently the variance of bitcoin prices. This section could be perceived as a side project but simultaneously a way to shed light on another, relevant to its function, part of bitcoin's behavior. To further investigate we will model the variance of the residuals (u_t^2). In order to do so, a generalized autoregressive conditionally heteroskedastic (GARCH) process will be used. We will use the GARCH (1,1) model with one ARCH and one GARCH term based on Bollerslev (1986). This analysis will only take place for the weekly data and not for the daily ones. As for the OLS analysis, S&P 500 and MSCI world indexes will be researched separately (as proxies for the WMP).

Before estimating the GARCH (1,1) model two conditions had to be checked. The first is whether we observe clustering volatility in the residuals. Graphically, we can receive a clear indication whether that is the case. If periods of high volatility are followed by periods of high volatility, while periods of low volatility seem to be followed by periods of low volatility for a certain amount of time, then the error term is conditionally heteroskedastic and can be represented by a GARCH model. The second step is to conduct the LM test for autoregressive conditional heteroskedasticity (ARCH), which supports a null hypothesis of no ARCH effects and an alternative of ARCH disturbance.

The GARCH (1,1) model that we are going to use models the variance of a regression model's disturbances as a linear function to both a lagged value of the squared regression disturbances and a lagged value of the conditional variance. The model starts primarily with the conditional mean regression:

$$y_t = \beta_0 + \beta_1 Z_{1t} + \cdots + \beta_\kappa Z_{kt} + u_t \qquad \text{(mean equation)} \qquad (2)$$

which is estimated normally with OLS being a BLUE estimator and $u_t \sim N(0, \sigma_t^2)$. Moreover, the conditional variance of the GARCH (1,1) model will be the following:

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + a_2 \sigma_{t-1}^2 \qquad \text{(variance equation)} \qquad (3)$$

Where: u_{t-1}^2 is the ARCH term, a_1 is the ARCH parameter, σ_{t-1}^2 is the GARCH term and a_2 is the GARCH parameter. The ARCH term is the squared residual, derived from the mean equation, and it shows the impact that information at t-1 has to the variance of the dependent variable at t. The GARCH is the variance of the dependent variable in the mean equation, and it reveals whether the variance at t does depend on the variance at t-1 (on past variance). No exogenous variable will be used in the variance equation. The ARCH and GARCH terms will represent the internal shock within our dependent variable (price of bitcoins).

Finally, we need to mention that, as stated above, the error term follows a Gaussian distribution (normal) (Engle, 1982). Many questioned that assumption. That is why we are also going to use other distributions, with fatter tails than the Gaussian, in an attempt to make our results more robust. Both the student's (t) distribution and the generalized error distribution (GED) will be used to ascertain whether results stay consistent. The t distribution has the tendency to converge to the Gaussian when the degrees of freedom go to infinity. We are going to use seven (7) degrees of freedom based on the number of different independent variables in the mean model. The GED distribution with a shape parameter of two (2) tends to converge to the Gaussian distribution, which is why we are going to use a shape parameter of less than two (1.5) to keep GED as a distribution with fatter tails.

5.5. Cointegration and Long Run Equilibrium

As discussed above, we will use differences to transform non-stationary variables to I(0) series in order to use OLS regression analysis. Unfortunately, the economic interpretation of the relationship between two differences is not the same as between two levels and we will not be able to determine a certain long run relationship between them. Furthermore, if we want to use our model to forecast time series we need to have a levels relationship. If the series involved in our model are cointegrated, then it is possible to find long run levels based relationships even if the series are not stationary.

Two or more time series are termed cointegrated when they share a stochastic trend. More accurately, we say that two or more series are cointegrated of order CI(d,b), if all the variables in the set are of order I(d) and a linear combination of all of them exists and it is of order I(d-b). In our models it is more likely to find the basic CI(1,1) situation, suggesting that all of the time series used are I(1) and the linear combination between them will result in an I(0) stationary series. Our goal is to investigate whether if y_t, X_t are I(1), but there is one or more B that make y_t-BX_t an I(0) process (with $B \neq 0$). In our case y_t represents the dependent variable (the price of bitcoin) and BX_t represents a vector of the independent variables $[\beta_1 x_{1t} + \ldots + \beta_k x_{kt}]$. We will use the variables that were of the same integration order in our previous OLS analysis, for either the daily or the weekly frequency, to determine whether they are cointegrated. If one or more cointegration parameters (B) exist, then we will use a Vector Error Correction Model (VECM) to estimate both the short and long run relations.

The Akaike's information criterion (AIC), the Schwarz's Bayesian information criterion (SBIC), and the Hannan and Quinn information criterion (HQIC) will be used to optimally select the lag-order for the VECM models. In our case, in which it is possible that more than two variables are cointegrated, cointegrated parameters could be more than one. If that is the case, we will use the Johansen's method (1995) to determine the existence and then the number of cointegrating vectors. Both the trace and the eigenvalue (maximum eigenvalue) statistic methods will be used to add extra validity to the results. Both methods are based on the Johansen's maximum likelihood estimator with the same null hypothesis and different alternative ones.

Based on the fact that both the Johansen's method and the VECM contain the same system of equations we will use the same lag selection for both. The VECM will help us determine the long run relationships and essentially it is a restricted form of a Vector Autoregression (VAR) model that contains a possible number of cointegrating relationships. We will only discuss the model that will have as a dependent variable the price of bitcoin (target model).

A series of diagnostic tests will be done to test the target model. At first, a Lagrange multiplier (LM) test will be conducted, to test for autocorrelation in the residuals of the model. The null hypothesis in the LM test will be the absence of autocorrelation. Consequently, the Jarque-Bera statistic (which tests skewness and kurtosis jointly) will be used to test whether the disturbances in the model are normally distributed. The null hypothesis will be that indeed the errors in the VECM are normally distributed. Ultimately, the eigenvalue stability condition in the VECM models will be checked. The stability test will provide evidence on whether the cointegrating equations are indeed stationary (as assumed by the VECM model) and whether the number of cointegrating equations, previously estimated, is misspecified. The graph of the eigenvalues of the companion matrix will be used to graphically complement the results.

6. Evidence of the Study – Results

This part of the thesis intends to illustrate, to examine, and to explain the results found. Through those results, we will try to give sufficient answers to the main and sub-questions, while suggesting new questions for further research and addressing unexpected results. The structure of this section will follow the structure previously discussed in the Analysis-Methodology part, starting with the specification of the model and dealing with stationarity, moving on to the OLS regressions for both daily and weekly data, discussing results relevant to the GARCH model and finally searching for cointegration and discussing possible long run dynamics. In order to conduct the quantitative analysis we primarily used the statistical package STATA (14.1), in addition to Matlab (part of the stationarity tests) and Python (3.6.1) (translating data after the download).

In the first part we discuss stationarity tests and the different time frequencies used to conduct this research. Consequently, we will demonstrate the models used for the OLS regression analysis and explain their format.

6.1. Stationarity Tests

As mentioned previously, the ADF, the PP and the KPSS stationarity tests will be used to determine whether a series is stationary. We aim to have a unanimous result for each variable (series). If that is not possible, further arguments will be made to explain each choice. To avoid misjudging a trend-stationary series for a unit root process, the graph of each variable and the linear and quadratic relationship with time through regressions will be examined. Thus, we conclude in whether we include a trend or a drift in our ADF and PP tests.

Firstly, it must be noted that stationarity tests were conducted for both the daily and the weekly form of each variable, dependent or independents, and for each of the assumptions in the daily analysis (“naïve” and “sophisticated” approaches). Results show that variables at both daily and weekly frequencies and regardless the assumptions that follow them, tend to behave in the same way and they demontsrate only insignificant differences that are not able to change the significance of a stationarity test. Only mild deviations were observed.

In the case of the second cluster of independent variables, the Gold Price and the two proxies for the World Market Portfolio (S&P500 index and MSCI-world index) exhibit unit root processes when being in levels, while they are stationary when testing their first difference. Finance related variables, similar to commodity prices and stock indexes, usually are non-stationary, as the natural order of time affects them. In addition they both have significant linear relations with time, as demonstrated in the ADF and PP tests. On the other hand, the VIX index is a stationary process for the time we are researching and regarding the ADF and PP tests. However, even if the KPSS test shows that the VIX index is marginally non-stationary, we will use the variable at its levels form, as two out of three tests proposed and based on the fact that it was marginally non-stationary for the last test.

The dependent variable, the price of a single bitcoin in the market, showed once more characteristics relevant to a finance variable, indicating that it is a unit root process in levels, but becomes stationary when transformed into first differences. Subsequently, regarding the first cluster of variables, we took a slightly different path. Both the Transaction Volume of bitcoins and the Hash Rate appear to have a clearly significant linear trend. The KPSS test, which has as null hypotheses the existence of a deterministic trend (trend-stationarity), showed that the Transaction Volume was marginally not trend-stationary and the Hash Rate was trend-stationary at a significance level that was lower than the requested. Moreover, the ADF and PP tests without a trend showed clear evidence of non-stationarity, while with the inclusion of a statistical significant trend the two variables are presented as stationary. This led us instead of first difference the series, to try to de-trend them. Accordingly, we regress the series linearly on time, predicting their residuals, which were the variables themselves but de-trended. The de-trended versions of the Traded Volume and the Hash Rate were now stationary, based on the new ADF and PP tests.

Finally, we should pay attention on the way Supply (total bitcoin in circulation) is used in this research. As already discussed in the Data Collection and Manipulation section, Supply had a radical shift in its shape as shown in the graph (June 9, 2016). As expected, there was evidence of a unit root process in each test. The problem was that this exact shift was also present in the first differenced variable. It was not a problem of discontinuity, and it made the first difference of Supply also non-stationary. We decided not to use partial regressions for immediately before and immediately after the turning point to accommodate the problem. In the case of “dividing” the variable, we saw that even if the second part was stationary, the first part was not stationary at the requested significance level. Rather, we took the second difference of the variable that it seems to be stationary in each of the tests.

Moving on, we found that Wikipedia Views, which is a variable used only in the daily analysis, is stationary, while Google Trends, which is a variable used only for weekly analysis, showed evidence of non-stationarity and had to be transformed into first difference form to be included in the OLS models. As a result, daily analysis will feature a proxy for trends in levels, while weekly in first difference.

Table III summarizes the previous results regarding stationarity tests.

Table III

Stationarity Tests

This table presents the three stationary tests used to determine the nature of each variable. The table provides *p*-values for the augmented Dickey–Fuller (ADF), the Phillips–Perron (PP) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. Results were similar both in daily and weekly analysis, with only minor differences. The table displays the results from the stationarity tests in the weekly analysis. Variables related to the second cluster of variables and the assumptions about the “naïve” and “sophisticated” traders in the daily analysis had the same results as the variables without the assumptions. Only very mild deviations were observed.

Variables	ADF *p*-values	PP *p*-values	KPSS *p*-values
ln_price	>0.1	>0.1	<0.01
D1.ln_price	<0.01	<0.01	>0.1
ln_TV	>0.1 (no trend incl.)	>0.1 (no trend incl.)	<0.01 (marginally)
ln_TV_de-trended	<0.01	<0.01	<0.01 (marginally)
ln_HR	>0.1 (no trend incl.)	>0.1 (no trend incl.)	<0.05
ln_HR_de-trended	<0.01	<0.01	<0.05
ln_S	>0.1	>0.1	<0.01
D1.ln_S	>0.05	>0.1	<0.01
D2.ln_S	<0.01	<0.01	>0.1
ln_GP	>0.1	>0.1	<0.01
D1.ln_GP	<0.01	<0.01	>0.1
ln_SP500	>0.1	>0.1	<0.01
D1.ln_SP500	<0.01	<0.01	>0.1
ln_MSCI_WORLD	>0.1	>0.1	<0.01
D1.ln_MSCI_WORLD	<0.01	<0.01	>0.1
ln_VIX	<0.01	<0.01	<0.01 (marginally)
ln_WikipediaViews_daily	<0.01	<0.01	>0.1
ln_GoogleTrends_weekly	>0.1	>0.1	<0.01
D1.ln_GoogleTrends_weekly	<0.01	<0.01	>0.1

6.2. Specifying the Models

This section presents the models that are going to be estimated in both the daily and the weekly analysis. The transformation of each variable is based on the results of the stationarity tests and aims to eliminate the possibility of a spurious regression through unit root processes or non-stationary series in general. As the fundamental regression we use the equation (1) presented in the Analysis-Methodology section.

In the daily analysis, the following model will be used:

$$\Delta \ln Price_t = \alpha_0 + \sum \begin{bmatrix} \beta_{TV} lnTV\_d_t \\ \beta_{HR} lnHR\_d_t \\ \beta_S \Delta\Delta lnS_t \end{bmatrix} + \sum \begin{bmatrix} \gamma_{GP} \Delta lnGP\_x_t \\ \gamma_{WMP} \Delta lnWMP\_x_t \\ \gamma_{VIX} lnVIX\_x_t \end{bmatrix} + \delta_{WikiViews} lnWikiViews_t + u_t \tag{4}$$

In equation (4), the first difference of the natural logarithm of Price is the dependent variable. Traded Volume and Hash Rate are de-trended and the stationary second difference of Supply is used as mentioned before. The Gold Price and the proxies for the World Market Portfolio are used in their first difference forms, while the VIX index is used in levels. Variables in the second matrix will be followed by the suffixes "_nn" and "_linear" separately, to represent the assumptions following the second cluster. In the daily analysis, a model including trends (WikiViews) and one without trends will be used.

In the weekly analysis, the following model will be used:

$$\Delta \ln Price_t = \alpha_0 + \sum \begin{bmatrix} \beta_{TV} lnTV\_d_t \\ \beta_{HR} lnHR\_d_t \\ \beta_S \Delta\Delta lnS_t \end{bmatrix} + \sum \begin{bmatrix} \gamma_{GP} \Delta lnGP_t \\ \gamma_{WMP} \Delta lnWMP_t \\ \gamma_{VIX} lnVIX_t \end{bmatrix} + \delta_{GoogleTrends} \Delta lnGoogleTrends_t + u_t \tag{5}$$

In equation (5), there are no assumptions relevant to the second matrix. Trends (GoogleTrends) will be used in first differences.

We should mention that the correlation matrices for both weekly and daily regressions showed almost no evidence of high correlations between the independent variables. This led us to refrain from further investigating a possible multicollinearity issue. The only exception was the correlation between S&P500 index and MSCI-world index, with 95.27% and 90.64% correlation in weekly and daily observations respectively. Even that small difference may be enough to observe differences between them. To formalize the argument, we used variance inflation factors for the independent variables (VIFs) to quantify the severity of a possible multicollinearity issue in each of the OLS regressions reported. Based on a rule of thumb that suggests that if the VIF for a certain variable is over 10 (ten) then multicollinearity is high, we ascertained that VIFs were between 1.01 and 1.42 for the independent variables and were from 1.06 to 1.19 for the combined models. Results suggest that the impact of collinearity in the models is not significant to greatly increase the variance of estimated coefficients.

The results of the correlations between the variables are reported in the Correlation Matrix at Chapter 9 (Appendix, Table B).

6.3. Daily OLS Estimations

A series of diagnostic tests were conducted to check the validity of each model and subsequently correct each model, if needed. Starting on, based on Durbin-Watson (Durbin's alternative test) and Breusch-Godfrey tests, there is evidence of autocorrelation in the errors (u_t) for equation (4) when trends are not included (in any case) and no serial correlation when trends are included (in any case). Bartlett's cumulative periodogram of white noise demonstrated graphically that u_t, in cases without trend, was not a white-noise process. To eliminate serial correlation, we choose to model the problem rather than use serial-correlation-consistent standard errors (Newey-West). Even if Newey-West standard errors will help us with inefficiency (biased standard errors), a possible inconsistency of the estimated coefficients will not be corrected. Modeling serial correlation will include lagged dependent variables to the model, specifying the appropriate dynamic model. We used the autocorrelation function (ACF) of the residuals with 365 lags (daily data, rule of thumb) and elected to include the second lagged first difference and the sixth lagged first difference of the dependent variable in the model. We expect that modeling serial correlation including lagged forms of the dependent variable in the regression might render results more difficult to interpret.

No sign of heteroskedasticity was detected when we excluded trends from equation (4), meaning that the variance of the error term is constant and finite for any value of explanatory variables during the same period. On the other hand, we used heteroskedasticity robust standard errors for the models with trends that showed the absence of homoskedasticity, and corrected standard errors accordingly. Finally, the Shapiro-Wilk test indicated that residuals are not normally distributed for equation (4), putting the quality of the model in question. Financial data are usually not normally distributed even if we need to assume that they are. In our case, high frequency data (daily observations) tend to contain additional patterns causing the problem. The sample used is large enough, which reduces our concerns regarding the outcomes had it been a smaller sample.

Results on the daily OLS analysis are reported on Table IV and Table C (Appendix, Chapter 9).

Table IV

OLS Daily (MSCI WORLD)

This table presents OLS regression estimates of Bitcoin Price on three clusters of independent variables. Both the dependent and the independent variables are transformed in the way discussed in section (5.2.). Suffices "_nn" and "_linear" represent the "naïve" and "sophisticated" assumptions created in section (3.3.7.). Each model uses the MSCI-world index as proxy for the world market portfolio (WMP). Models (1) and (3) exclude the third cluster of independent variables (trends) while models (2) and (4) include Wikipedia Views as a measure for trends. Models (2) and (4) present models with heteroskedasticity-consistent standard errors. Both the second (L2) and the sixth (L6) lagged values of the dependent are used for models (1) and (3) to model serial correlation. The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 1139 daily observations. t-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Variables	MSCI_nn	MSCI+trends_nn	MSCI_linear	MSCI+trends_linear
Traded Volume	0.00797	-0.000074	0.00794	-0.000163
	(1.14)	(-0.01)	(1.14)	(-0.02)
Hash Rate	-0.00928*	0.00117	-0.00924*	0.00125
	(-1.94)	(0.16)	(-1.93)	(0.17)
D2.Supply	-18.10	-22.56	-18.24	-21.84
	(-0.39)	(-0.45)	(-0.39)	(-0.44)
D1.Gold Price_nn	0.208	0.214		
	(1.59)	(1.34)		
D1.MSCI WORLD_nn	-0.197	-0.183		
	(-1.21)	(-0.81)		
VIX index_nn	-0.00760*	-0.00528		
	(-1.75)	(-0.99)		
L2.D1.Price	-0.0780***		-0.0793***	
	(-2.61)		(-2.65)	
L6.D1.Price	0.0829***		0.0829***	
	(2.77)		(2.77)	
Wikipedia Views_daily		0.00151		0.00153
		(0.41)		(0.41)
D1.Gold Price_linear			0.201	0.224
			(1.45)	(1.28)
D1.MSCI WORLD_linear			-0.178	-0.194
			(-1.01)	(-0.85)
VIX index_linear			-0.00768*	-0.00506
			(-1.76)	(-0.92)
Constant	0.0221*	0.00359	0.0223*	0.00282
	(1.88)	(0.09)	(1.89)	(0.07)
Observations	1,132	713	1,132	713
Adj. R-squared	0.022	0.008	0.022	0.008
Robust Standard Errors	NO	YES	NO	YES
Wiki-Views	NO	YES	NO	YES

6.3.1. The Hash Rate Effect

The daily OLS regressions (1) and (3) for both world market portfolio proxies show a negatively correlated and statistically significant relationship between the Hash Rate and the Price of bitcoin (90% significance level). As mentioned, Hash Rate is the speed at which a computer completes an operation in the Bitcoin system. Thus, a higher Hash Rate favors the miner as it increases his opportunity to find the solution to the next block. To avoid any possible misunderstanding, it needs to be clarified that the Hash Rate is not only the amount of computational power that users allocate to the system, but it is also a measure of security. The higher the Hash Rate, the more difficult it is for single users to take advantage of the system, a situation that may result in alteration or fraud. In addition, some could argue that the Hash Rate also indicates the amount of investment that the bitcoin community allocates to the system, and subsequently is a proxy for the trust in the process of mining and bitcoin in general. We state the above to explain that even if we found a negative correlation in the short-run, it is possible to witness a positive one in the long-run analysis. To support the case for the short-run dynamics, we can argue that even if the above were true, mining only becomes more difficult. There was a time, mainly in the initial stages of the system, that significant mining was possible with a common CPU of a personal computer. From then on, we went up to GPU home-based mining systems, and to specialized mining companies with custom-made hardware for mining. Even if the number of participants in the mining process is increasing, the intensification of difficulty will eventually render mining inefficient for some miners, or be a barrier to their participation. Vastly different hardware can be used in the mining process but only the one with the better computational power (Hash Rate) has a better chance of being rewarded. This may result in the loss of active participants in the bitcoin community and influence the price accordingly.

6.3.2. The VIX Index Effect

Regressions (1) and (3) for the MSCI-world index and regression (3) for the S&P500 index show a negatively correlated and statistically significant relationship between the VIX index and the bitcoin Price (90% significance level), which is in line with our expectations. Regression (1) of the S&P500 marginally fails to show the same result. We are again dealing with short-run dynamics on a daily basis. The result suggests that as the uncertainty in the market rises, bitcoin is used as an alternative asset and is expected to gain from that kind of attendance.

6.3.3. Other Relevant Results

Modeling serial correlation and eliminating the problem in regressions (1) and (3) for both cases, resulted in a particular price momentum effect. A change in the price two days prior to today will result in a negatively correlated and statistically significant relationship with the price today, while a change in the price six days prior to today will result in a positively correlated and statistically significant relationship with the price today (both at the 99% significance level). A possible interpretation results from the values of the coefficients of the second and the sixth lagged first difference of the dependent value. The different signs in the coefficients support the high variability in bitcoin price. In addition, based on the coefficients, it appears that those deviations exhibit a mean reversion with an upward trend for bitcoin's price.

Regarding the goodness of fit, both the R^2 and the adjusted R^2 for regressions (1) and (3) for both cases were slightly over 2% (2.2% for MSCI-world, 2.1% for S&P500), indicating that only approximately 2% of the sample variation in the dependent variable is explained by the independent variables. On the contrary, the corresponding measures for regressions (2) and (4) that include trends (Wikipedia Views) are below 1%. In general, as expected, noise in the daily analysis could question the validity of the model.

The inclusion of Wikipedia Views in the model tilted all aspects. Regressions (2) and (4) showed no sign of significant short-run determinants for the dependent variable. One of the reasons is that data were available only from July 1, 2015 and on (creation of the database). Even though past studies showed that Wikipedia Views were significant, they used a different database, which may account for this difference. Past research used the old pageview analysis tool that Wikipedia offered, a tool that sometimes was unavailable to the user and has not been updated since January 2016 (stats.grok.se). Furthermore, in the case of both the old and the new tools, there have been some rare cases in which the data are purposely manipulated.

Last but not least, the differences between the "naïve" (_nn) and the "sophisticated" (_linear) assumptions in the results must be addressed. Unfortunately, the differences between the two are insignificant, showing that further research is needed to determine how bitcoin's price is affected by financial variables on weekends in a daily analysis. Based on the goodness of fit (R^2), the F-test (for joint significance of the coefficients of the independent variables) and the t-statistics of the statistically significant variables in each model, the "sophisticated" (_linear) approach seems to marginally be more plausible.

6.4. Weekly OLS Estimations

As far as the diagnostic tests are concerned, no serial correlation in the u_t of either the S&P500 or the MSCI-world cases in the weekly analysis for equation (5) was found. Then, once more, the assumption relevant to the normality of the distribution of the residuals did not hold. The same arguments used in the daily analysis, will once more be used to explain the problem.

The interesting part of the diagnostic tests in the weekly OLS analysis emerged when we tested for the assumption of homoskedasticity. The Breusch-Pagan test showed strong evidence of heteroskedasticity in the residuals, and to accommodate the biased standard errors we report both the original and a regression with robust standard errors. The existence of heteroskedasticity piqued my curiosity regarding possible dynamic forms of heteroskedasticity. In the next chapter, and as a useful side-project, we examine an effort to model heteroskedasticity and learn about the way in which the variance of bitcoin behaves.

Results of daily OLS analysis are based on Table V.

Table V

OLS Weekly

This table presents OLS regression estimates of Bitcoin Price on three clusters of independent variables. Both the dependent and the independent variables are transformed in the way discussed in section (5.2.). Models (1) and (3) use the S&P500 and the MSCI-world indexes as proxies for the world market portfolio (WMP) respectively. Models (2) and (4) present the same models with heteroskedasticity-consistent standard errors. The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 163 weekly observations. t-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Variables	S&P500	S&P500	MSCI-world	MSCI-world
Traded Volume	0.104**	0.104*	0.105**	0.105**
	(2.02)	(1.97)	(2.04)	(2.01)
Hash Rate	-0.047	-0.047	-0.048	-0.048
	(-1.54)	(-1.51)	(-1.55)	(-1.51)
D2.Supply	-3.986	-3.986	-5.567	-5.567
	(-0.08)	(-0.07)	(-0.10)	(-0.10)
D1.Gold Price	-0.554*	-0.554*	-0.545*	-0.545
	(-1.71)	(-1.66)	(-1.70)	(-1.645)
D1.S&P500	-0.230	-0.230		
	(-0.54)	(-0.57)		
D1.MSCI WORLD			-0.284	-0.284
			(-0.65)	(-0.70)
VIX index	-0.065**	-0.065**	-0.067**	-0.067**
	(-2.16)	(-2.26)	(-2.21)	(-2.27)
D1.Google Trends (weekly)	0.074*	0.074	0.074*	0.074
	(1.78)	(1.59)	(1.79)	(1.60)
Constant	0.185**	0.185**	0.190**	0.190**
	(2.29)	(2.39)	(2.33)	(2.41)
Observations	161	161	161	161
Adj. R-squared	0.095	0.095	0.096	0.096
Robust Standard Errors	NO	YES	NO	YES

6.4.1. The Transaction Volume Effect

Regressions (1), (2), (3), and (4) for the weekly OLS analysis show a positively correlated and statistically significant relationship between Transaction Volume and bitcoin's Price (95% significance level). That means that as the number of transactions per day rises, the dependent variable will follow. This is hardly surprising, since Transaction Volume could be perceived as a proxy for the general attention that the system has. In addition, arguing in terms of equilibrium, someone could support this variable as a proxy for demand in the Bitcoin community.

6.4.2. Gold Price and VIX Index

Regressions (1), (2), and (3) show a negatively correlated and statistically significant relationship between Gold Price and the dependent variable (90% significance level). Regression (4) marginally fails to do the same (t-stat of 1.645). This is not a counter intuitive result, even if we expected a positive relation. The result implies that Gold and Bitcoin are treated as substitute investments. A decrease in the Gold Price would eventually result in an increase in the demand for Bitcoin. This also illustrates how bitcoin is treated by possible investors as an alternative asset.

Moreover, regressions (1), (2), (3), and (4) report a negatively correlated and statistically significant relationship between the VIX index and the Price of bitcoin (95% significance level), which is in line with the results found in the daily analysis and the original expectations. The difference in the weekly data is that the impact of the relation is greater and the determinant is more statistically significant. More specifically, increased market uncertainty induces investors to search for alternative investments, as Bitcoin is.

6.4.3. The Attendance Effect

The last short-run effect that possibly influences the Price of bitcoin could be seen as an attendance effect. Regressions (1) and (3) exhibit a positive correlated and statistically significant relation between Google Trends and the dependent variable (90% significance level). It seems that trends could be seen as a measure of attendance that creates new participants in the bitcoin community, and subsequently could partly predict the price performance of bitcoin. On the contrary, regressions (2) and (4) that confront heteroskedasticity with robust standard errors indicate an insignificant impact of trends and attendance on the bitcoin's price. In the section regarding the long-run equilibrium, we will observe possible differences between short-run dynamics and long-run dynamics of trends on the dependent variable.

6.4.4. Other Relevant Results

Results fail to establish a certain significant correlation between the Supply of bitcoin (total bitcoin in circulation) and the Price. This might be due to the fact that supply is predetermined and finite, which makes its impact in short-run weekly analysis insignificant and presents bitcoin as free of inflation. A second, technical reason that supply fails as a determinant of the price, may arise from the way that the variable was manipulated (second differences).

With regard to the weekly analysis, R^2 and the adjusted R^2 were higher than the daily analysis, at approximately 10% (9.5% for the S&P500 and 9.6% for the MSCI-world). Partly, this was the result of the lesser noise of weekly data compared to the daily data and partly because, as expected, weekly data are fewer than daily data. Furthermore, it seems that Hash Rate was insignificant in the weekly analysis, suggesting that there is a possibility that noisy daily data resulted in a significant Hash Rate or that there are some kind of differences between the time frequencies in the short-run analysis.

6.5. Differences Between Time Frequencies

Based on the fact that most common estimation issues are controlled and solved through the diagnostic tests, and assuming that noise in the daily data does not change the significance of the factors, we can argue that it is possible that bitcoin has different determinants in different time frequencies (short-run dynamics). This is hardly surprising, since it is possible that technical aspects of bitcoin (Hash Rate) could influence the price in a more direct mode than financial aspects of bitcoin and trends (Gold Price, VIX index, Google Trends). The latter would need more time to make an impact on price. The models about daily and weekly short-run dynamics suggest evidence in favor of that case.

6.6. Propositions on Modeling Bitcoin's Heteroskedasticity

This part of the thesis discusses dynamic forms of heteroskedasticity, as mentioned and explained previously in the section about Methodology. Even if this part is a side project, due to the fact that it does not actively answer the main or some of the sub-questions of the research, it is useful in order to understand different aspects of Bitcoin behavior.

Estimates of the model are reported in Table VI and in Table D (Appendix, Chapter 9), while graphs depicting residuals of the weekly analysis are presented in Charts B (Appendix, Chapter 9).

As previously reported, regressions (1) and (3) in the weekly analysis were suffering from heteroskedasticity. Using the same variables as in equation (5) and creating a

GARCH(1,1) model based on the mean equation (2) and the variance equation (3), we will endeavor to establish a possible framework to tackle any relevant issue in future research.

Firstly, we checked for the existence of clustering volatility and the ARCH effect as the two preconditions in order to estimate the GARCH(1,1) model. For the first instance, we plot residuals through time for both weekly analysis models to observe a possible clustering volatility effect in the residuals. In financial variables, sometimes periods of high volatility are followed by periods of high volatility, while periods of low volatility tend to be followed by periods of low volatility. This effect would suggest that residuals in our case (or the error term) are conditionally heteroskedastic. In our graphs there is no clear evidence of a clustering volatility effect. Furthermore, the ARCH effect was marginally significant for the S&P500 case, while it was insignificant for the MSCI-world case. Either way, we decided to run the GARCH(1,1) model based on the fact that even the absence of results are interesting results in themselves.

As noted earlier, Bitcoin's price and the first difference of Bitcoin's ln-price (which is a good approximation for returns on Bitcoin) tend to follow some of the stylized facts that financial variables tend to follow. It seems that the dependent variable does not follow a Gaussian (normal) distribution, or more accurately, normal distribution might not be the best assumption to explain this distribution. On the contrary, we saw evidence (based on skewness and kurtosis) of leptokurtosis, of the tendency to have distribution that exhibit fat tail and excess peakedness at the mean. Therefore, we selected not only a model with the normal, but also models with the student's (t) distribution and the generalized error distribution (GED). The specifications of the models were discussed previously in the section about Methodology. All of the above were mentioned to explain that indeed, the dependent variable has some common stylized facts with financial variables, which may be one of the reasons why Bitcoin is perceived as an alternative financial asset.

As far as the mean equation is concerned, estimations and the significance of the independent variables for each of the models are the same as with the weekly OLS analysis. Transaction Volume and Google Trends both exhibit a positively correlated and statistically significant relationship with the dependent variable, while Gold Price display a negatively correlated and statistically significant relationship with the dependent one. The only difference is that, in this case, the VIX index is indifferent from zero.

The variance equation for each model shows an insignificant ARCH factor and a positively correlated and significant GARCH factor. In a different setting this could probably mean that the variance of the dependent variable does depend on past variance, or that the volatility at t-1 can influence the volatility today (at t). Another interesting result showed that using an F-test we have clear evidence of a jointly significant impact of both the ARCH and the GARCH terms. Last but not least, we test the hypotheses of having ARCH and GARCH terms to sum to unity (F-test) that indicates the existence of a restricted version of the GARCH model, an integrated GARCH (IGARCH). This indicates a possible exploding variance forecast.

The residuals for each model are tested for both the presence of serial correlation and to prove whether they are normally distributed. No evidence of serial correlation was found. On the contrary, the Shapiro-Wilk test demonstrated clear evidence that residuals were not normally distributed. Further diagnostic tests definitely need to be conducted to further test the validity of each model.

Table VI

GARCH[1.1] (MSCI world)

This table presents GARCH[1.1] regression estimates of Bitcoin Price on three clusters of independent variables. Both the dependent and the independent variables are transformed in the way discussed in section (5.2.). Each model has a mean and a variance equation as discussed in the section (5.4.). Both one lagged ARCH and one lagged GARCH term is reported for each model in the variance equations. Model (1) is based on the Gaussian (normal) distribution, model (2) is based on the student's t distribution with seven (7) degrees of freedom and model (3) is based on the generalized error distribution and with a shape parameter of (1.5). Each model uses the MSCI-world index as proxy for the world market portfolio (WMP). The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 163 weekly observations. z-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Variables	Mean (1) Equation	Variance (1) Equation	Mean (2) Equation	Variance (2) Equation	Mean (3) Equation	Variance (3) Equation
Traded Volume	0.124**		0.107**		0.116**	
	(2.11)		(2.32)		(2.41)	
Hash Rate	-0.052		-0.036		-0.039	
	(-1.63)		(-1.37)		(-1.41)	
D2.Supply	-26.509		0.827		2.619	
	(-0.45)		(0.02)		(0.05)	
D1.Gold Price	-0.583*		-0.551*		-0.554*	
	(-1.74)		(-1.92)		(-1.88)	
D1.MSCI world	-0.209		-0.112		-0.167	
	(-0.38)		(-0.27)		(-0.39)	
VIX index	-0.052		-0.036		-0.039	
	(-1.39)		(-1.21)		(-1.27)	
D1.Google Trends (weekly)	0.082**		0.083***		0.078**	
	(2.26)		(2.62)		(2.34)	
L.ARCH		0.103		0.112		0.100
		(1.19)		(1.14)		(1.06)
L.GARCH		0.721***		0.691***		0.711***
		(3.16)		(2.67)		(2.65)
Constant	0.147	0.001	0.103	0.001	0.111	0.001
	(1.49)	(1.00)	(1.32)	(0.96)	(1.36)	(0.88)
Observations	161	161	161	161	161	161
Gaussian[normal]	YES	YES	NO	NO	NO	NO
Student's t	NO	NO	YES	YES	NO	NO
Generalized error distribution	NO	NO	NO	NO	YES	YES

6.7. Evidence for Long-Run Dynamics

In this part of the research we discuss results based on the methodology introduced in chapter 5.5. Starting on, we should mention that Johansen's cointegration test resulted in no signs for cointegration for the daily analysis. Therefore, vector error correction models will only be made for the weekly analysis that showed positive signs of one cointegration rank. The use of the variables is discussed in chapter 5.5., and we mainly want to see variables relative to the second and the third cluster of independent variables of the weekly analysis. Based on AIC, SBIC, and HQIC, one lag was the optimal selection for both the cointegration tests and the VECMs. We had to be cautious with the lag selection because too many lags could result in increasing the error in the forecasts, while too few could exclude relevant information. In addition, both the trace statistic and the max statistic suggested the existence of one cointegrating rank for both the S&P500 and the MSCI-world indexes. Table E reports the results of the Johansen's cointegration test for the two proxies of the world market portfolio (Appendix, Chapter 9).

Moving on to the discussion about the results regarding the VECMs, results for both the short-run and the long-run dynamics are reported in Table VII. The short-run dynamics demonstrated a positive correlated and statistically significant relationship between Google Trends (lagged-difference) and the dependent variable. This is hardly surprising based on the fact that this result is in line with our previous analysis on the weekly OLS regressions. Moreover, there is a short-run relationship that is positive and statistically significant between the two proxies for the world market portfolio and the price of bitcoin. This is a counter intuitive result. We expected to see a negative relationship that could explain a substitute relation between the stock market and an alternative asset such as bitcoin. On the contrary, we find evidence of a complementary relationship. The impact of MSCI-world index seems to be more statistically significant than the impact of S&P500 index (99% to 95% significance level).

One of the most interesting results is the negative and statistically significant error correction term. The error correction term represents the speed of adjustment/convergence to the long run equilibrium. Thus, a negative sign in our models indicates that the models revert to a specific long run equilibrium. More accurately, approximately 5% of the gap between the price of bitcoin at period t-1 and the equilibrium price tends to be reversed back at period t (4.34% for S&P500 model, 5.18% for MSCI-world model). The approximately 5% adjustment could be perceived as relatively low, but based on the way in which bitcoin reaches all-time highs almost every week for the last many months, it makes it seem only rational. Another result that we should not overlook is the positive and statistically significant constant term in the short-run dynamics. A positive constant term in a differenced equation corresponds to a positive trend term in the levels. The positive sign of this trend could partly explain the explosion in bitcoin's performance.

With regard to the long-run cointegrating relationships we observe that, as also suggested in the short-run weekly analysis, Gold Price has a negative and statistically relationship with the performance of Bitcoin. On the other hand, the proxies for the world market portfolio once more indicated a long-run positive and significant relationship with bitcoin. It seems that in the long-run an increase in the stock market could propel participants in the Bitcoin community to involve further and improve the performance of the price. Finally, we should mention that, contrary to our expectations and the signs from the short-run analysis, trends tend to have a negative and statistically significant relationship with bitcoin.

The VECMs were used to establish a long-run relationship between the variables (especially financial and trends) and to show a predictive power (forecasting ability) of the independent variables to bitcoin's behavior. That allegation is supported by the fact that the target models discussed (models that have as dependent variable the price of bitcoin) had the most significant R^2, suggesting that independent variables are possibly causing bitcoins fluctuations and not the opposite (possible sign against reverse causality issue).

Table VII reports both short-run and long-run dynamics discussed in the VECMs.

Last but not least, we performed a series of post-estimation diagnostic tests to check different aspects of the models. Firstly, we used a Lagrange multiplier (LM) test to test for autocorrelation in the residuals of the models and ultimately were unable to reject the null hypothesis of no autocorrelation in the residuals (both 2 and 4 lag orders were used), suggesting that there is no evidence of model misspecification. Then, we rejected the null hypothesis of the Jarque-Bera statistic (which jointly tests skewness and kurtosis), suggesting that disturbances in the VECMs were not distributed normally. Finally, we concluded diagnostics by testing the eigenvalue stability condition in the models. The results suggested that cointegrating equations were stationary and the number of cointegrating equations was correctly specified. Results on the stability test can be found on Charts C (Appendix, Chapter 9).

Table VII
VECM

These tables present the VECM models regression estimates of Bitcoin Price on cointegrated independent variables. The first table shows the short-run dynamics (lagged-differenced independent variables) and the adjustment to the long-run equilibrium (error correction term), while the second shows the long-run dynamics. Both the dependent and the independent variables are transformed and follow the regression discussed in section (5.5.). Only the target equations are reported (equations that have as the dependent variable the first difference of bitcoin's (ln)Price). Based on the information criteria AIC, SBIC and HQIC the models use the first lagged value for each of the cointegrated independent variables. The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 163 weekly observations. z-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Variables	Target equation (S&P500 case)	Target equation (MSCI-world case)
Error correction term	-0.0434**	-0.0518***
	(-2.33)	(-2.66)
LD.Price	-0.0041	-0.0054
	(-0.05)	(-0.07)
LD.Gold Price	-0.2903	-0.3817
	(-0.91)	(-1.20)
LD.S&P500	0.9373**	
	(2.49)	
LD.Google Trends (weekly)	0.0998**	0.0970**
	(2.25)	(2.23)
LD.MSCI WORLD		0.9744***
		(2.59)
Constant	0.0104*	0.0112*
	(1.67)	(1.82)
Observations	161	161
R-squared	0.1497	0.1609

Estimates of the parameters in the cointegrating equations.

Variables	Coefficient	z-statistic	Coefficient	z-statistic
Gold Price	-2.331**	-2.47	-2.894***	-3.13
S&P500	4.768***	3.57	-	-
MSCI WORLD	-	-	4.428***	3.23
Google Trends (weekly)	-2.070***	-9.51	-1.783***	-10.84
Constant	-20.487	-	-13.778	-

6.7.1. World Market Portfolio Effect

The use of two different proxies for the WPM variable was not made to only include a new relevant variable to the literature. The main goal was to search for evidence that could support the validity of arguments in favor of a certain relationship. As we saw earlier, both the goodness of fit of the VECM model (R^2) and the individual z-statistics of each coefficient in the models showed that MSCI-world index might be a better choice than S&P500 index, to explain Bitcoin's behavior. Furthermore, we have seen that regardless of the significance of the coefficients of the world market portfolio proxies, models with MSCI-world (OLS or GARCH) seem to be more accurate for the same reasons. All the above suggest that it might be a diversification effect in our case. To explain further, the way that a more diversified price index of common stocks (MSCI-world index, 1,652 "world" stocks) relates to Bitcoin, than a less diversified price index (S&P500 index, 500 U.S. stocks), shows that indeed the way that "real" economy behaves is a determinant for Bitcoin. In addition, participants in the Bitcoin community could possibly also invest in other more ordinary asset classes such as stocks. A future study could incorporate different stock indexes to test that argument, or even test the performance of well-diversified portfolios that include more asset classes with the performance of Bitcoin.

6.7.2. Correlation or Causality ?

One of the greatest challenges of this research was to establish causal effects for the possible determinants of the performance of Bitcoin's price. We are fully aware that the relationships described above are based on correlations, estimated from the quantitative methods used. The main purpose was to use those correlations and interpret the results in an intuitive way both by using the relevant theoretical background and by trying to add to the research on the subject. Future research will demonstrate whether we manage to establish causal relationships. Being pragmatic, we have to admit that problems such as endogeneity due to one or more omitted relevant variables or reverse causality in the models are possible.

7. Conclusion - Suggestions for Further Research

This final part of the thesis summarizes the most important conclusions and offers some suggestions for relevant research in the future. This part is concluded with a final personal statement.

The main research question and the relevant sub-hypotheses were sufficiently addressed in the methodology used. We used time-series regression analysis to build an empirical model that consists of possible Bitcoin value drivers. Three different clusters of determinants were proposed; the technological, the financial, and the attendance through trends. By consulting existing literature and further extending the process using new variables, different data, and an alternative modeling approach, we manage to estimate possible price factors both in the short-run and the long-run. We first dealt with stationarity and constructed OLS regression models for a daily and a weekly analysis. We saw evidence of a negative short-run impact on the dependent for both the VIX index and the Hash Rate in the daily analysis. Moreover, Transaction Volume and Google Trends (based on Google queries) demonstrated evidence of a positive short-run impact on bitcoin's price, while Gold Price and VIX index showed a negative short-run impact. Since the phenomenon is relatively new and the price performance has been more than significantly impressive for the past months and, generally, since the inception of the cryptocurrency, we decided to also search for long-run determinants. After finding the existence of cointegration for the models used in the weekly analysis, we estimated a VECM to test for possible long-run relationships. For the set of cointegrated variables and discussing only the target model that has as dependent variable the Bitcoin Price, we saw that Gold Price and Google Trends exhibit a negative long-run relationship with the price, while the proxies for the World Market Portfolio (S&P500 and MSCI-world) seemed to have a positive impact on the price of bitcoin. Additionally, we saw that the speed of adjustment to the long-run equilibrium for the price of bitcoin was relatively low, at approximately 5%. A possible diversification effect was introduced. Finally, a GARCH model was proposed to model volatility in the weekly data and shed light on some interesting aspects of bitcoin such as, for example, the common ground that bitcoin shares with financial variables (leptokurtic distribution).

The next part of this section focuses on suggestions for future research. The available literature on bitcoin is still limited in the areas of finance and economics in comparison to other areas. The propositions focus on possible new determinants of the price performance and some ideas regarding the modeling of the problem in the future. Finally, we must state that each proposition was researched to the extent that it would be feasible and relevant to the academic research.

A possible determinant of bitcoin's price could be the price of other "powerhouses" in the cryptocurrency community. Thus, the price of other major players in the market could influence bitcoin to some extent. The prices of Ethereum, Ripple, Litecoin, Dash, Monero, NEM and Bitcoin Cash could be used to determine whether they influence Bitcoin. Cumulatively, they represent approximately 32% of the total market capitalization for cryptocurrencies, while Bitcoin dominates the market with approximately 53% (data based on November 10, 2017). A preferable approach would be to include a weighted average price based on their market capitalization as an additional independent variable to determine part of bitcoin's behavior. However, it would be somewhat challenging to avoid any reverse causality issue regarding whether the rest of participant affect or get affected by Bitcoin.

In the case of financial variables that have a significant relationship with Bitcoin, dummy variables that will represent a certain number of the highest and the lowest return days could be used, in an effort to produce more informative results. The "best" and "worst" dummies could also be implemented in our own analysis for the Price of Gold, the proxies of the World Market Portfolio, or the VIX index.

In addition, instead of Gold Price or S&P500 and the MSCI-world stock price index, different proxies could be used. It would be interesting to see whether a certain commodity, for example, correlates more efficiently than others do. In such a case, commodity price indexes or different stock indexes to support the diversification effect should be used.

With regard to the models used to approach the situation, it might be that an unrestricted vector autoregression model (VAR) in levels or differences (depending on the nature of the variables used) could be a plausible proposal to replace OLS regressions in the first part of a similar research. Furthermore, a Granger's causality test, between the independent and the dependent variable, to accompany the VAR model could possibly result in some kind of predictive causality in the model. The test will not prove causal effects, but it could offer predictive power to the results. Moreover, it would be interesting to conduct multiple short-run analyses, but instead of using ordinary least squares (OLS), to use generalized least squares (GLS) to avoid assumptions on the normality of the residuals in the models.

An interesting idea would be to use event study analysis to see how the hard fork in the blockchain of Bitcoin on August 1, 2017, which resulted in the creation of Bitcoin Cash, influences Bitcoin. Thus, we can observe the difference before and immediately after the hard fork. This was one of the most important moments in Bitcoin history. In July 20, 2017, bitcoin miners voted to pass the Bitcoin Improvement Proposal 91 (BIP). The main part of the proposal was to activate the Segregated Witness (SegWit). For some time, the Bitcoin community has argued that the system suffers a certain scalability problem. This would be an effort to solve that blockchain limitation problem that increases bitcoin's transaction fees while reducing transaction speed. Hard fork is a term used to explain a situation in which a certain blockchain splits into two separate chains based on different governing and regulating rules for the system. An event study would probably research on how this period influences the performance of bitcoin. However, some limitations in the data make this a rather challenging task.

8. Acknowledgements

I would like to thank my supervisor professor dr. Bertrand Melenberg for his guidance, his advice and most of all for the liberty he gave to expand every idea to as far as I wanted.

Moreover, I would like to thank my friends and family for their support and that weird belief they have on me. Over and above, I would like to thank Eirini for constantly inspiring me and always standing by me. Lastly, I would like to thank Makis for teaching me the virtue of curiosity.

9.Appendix

Table A

Description of Variables

This table describes the set of variables used in this research.

Variables	Abbreviations	Description	Source
Price	ln_price	price of a single bitcoin in USD in Bitcoin Price Index (BPI) (weighted average across major exchanges)	coindesk.com
Traded Volume	ln_TV	total daily unique transaction volume in bitcoin system	blockchain.info
Hash Rate	ln_HR	average daily number of tera-hashes per second (computational power)	blockchain.info
Supply	ln_S	daily total bitcoins in circulation	blockchain.info
Gold Price	ln_GP, ln_GP_nn, ln_GP_linear	"fair" gold price as an average from major exchanges	datastream
S&P 500 index	ln_SP500, ln_SP500_nn, ln_SP500_linear	Standard & Poor's 500 stock market price index	datastream
MSCI world index	ln_MSCI_WORLD, ln_MSCI_WORLD_nn, ln_MSCI_WORLD_linear	MSCI world USD-denominated price index	datastream
VIX index	ln_VIX, ln_VIX_nn, ln_VIX_linear	CBOE's volatility index	cboe.com, quandl.com
Wikipedia Views (daily)	ln_WikipediaViews	number of daily searches in Wikipedia as an actual number	tools.wmflabs.org/ pageviews
Google Trends (weekly)	ln_GoogleTrends	normalized number of weekly search queries in Google (scaled from 0 to 100)	trends.google.com

Table B

Correlation Matrix

This table provides correlations between all the variables used for the weekly analysis. Variables are transformed into their stationary forms and the data used for the weekly correlation matrix consists of 163 weekly observations.

	D1.Price	TV	HR	D2.S	D1.GP	D1.SP500	D1.MSCI	VIX	D1.Google
D1.Price	1.0000								
Traded Volume	0.0627	1.0000							
Hash Rate	-0.1330	0.0374	1.0000						
D2.Supply	-0.0097	-0.0489	0.0155	1.0000					
D1.Gold Price	-0.1423	0.1850	0.1170	-0.0338	1.0000				
D1.S&P500	0.0670	0.0346	0.0010	-0.0760	-0.2088	1.0000			
D1.MSCI world	0.0564	0.0309	-0.0043	-0.1073	-0.1488	0.9527	1.0000		
VIX index	-0.1630	0.3113	-0.0162	0.0174	0.1887	-0.4116	-0.4245	1.0000	
D1.Google Trends	0.1426	-0.0990	-0.0479	0.0094	0.0425	-0.0127	0.0055	-0.1021	1.0000

Table C

OLS Daily (S&P500)

This table presents OLS regression estimates of Bitcoin Price on three clusters of independent variables. Both the dependent and the independent variables are transformed in the way discussed in section (5.2.). Suffices "_nn" and "_linear" represent the "naïve" and "sophisticated" assumptions created in section (3.3.7.). Each model uses the S&P500 index as proxy for the world market portfolio (WMP). Models (1) and (3) exclude the third cluster of independent variables (trends) while models (2) and (4) include Wikipedia Views as a measure for trends. Models (2) and (4) present models with heteroskedasticity-consistent standard errors. Both the second (L2) and the sixth (L6) lagged values of the dependent are used for models (1) and (3) to model serial correlation. The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 1139 daily observations. t-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Variables	(1) S&P500_nn	(2) S&P500+trends_nn	(3) S&P500_linear	(4) S&P500+trends_linear
Traded Volume	0.00749	-0.000628	0.00758	-0.000580
	(1.07)	(-0.08)	(1.09)	(-0.07)
Hash Rate	-0.00930*	0.00105	-0.00927*	0.00115
	(-1.95)	(0.14)	(-1.94)	(0.16)
D2.Supply	-17.71	-22.13	-18.66	-22.28
	(-0.38)	(-0.44)	(-0.40)	(-0.45)
D1.Gold Price_nn	0.213	0.227		
	(1.61)	(1.41)		
D1.S&P500_nn	-0.0896	-0.0500		
	(-0.62)	(-0.26)		
VIX index_nn	-0.00700	-0.00443		
	(-1.62)	(-0.83)		
L2.D1.Price	-0.0788***		-0.0798***	
	(-2.63)		(-2.67)	
L6.D1.Price	0.0824***		0.0824***	
	(2.75)		(2.76)	
Wikipedia Views_daily		0.00173		0.00169
		(0.47)		(0.46)
D1.Gold Price_linear			0.200	0.227
			(1.44)	(1.29)
D1.S&P500_linear			-0.102	-0.101
			(-0.66)	(-0.52)
VIX index_linear			-0.00726*	-0.00445
			(-1.68)	(-0.81)
Constant	0.0204*	-0.000796	0.0212*	-0.000375
	(1.75)	(-0.02)	(1.81)	(-0.01)
Observations	1,132	713	1,132	713
Adj. R-squared	0.021	0.007	0.021	0.007
Robust Standard Errors	NO	YES	NO	YES
Wiki-Views	NO	YES	NO	YES

Table D
GARCH[1.1] (S&P500)

This table presents GARCH[1.1] regression estimates of Bitcoin Price on three clusters of independent variables. Both the dependent and the independent variables are transformed in the way discussed in section (5.2.). Each model has a mean and a variance equation as discussed in the section (5.4.). Both one lagged ARCH and one lagged GARCH term is reported for each model in the variance equations. Model (1) is based on the Gaussian (normal) distribution, model (2) is based on the student's t distribution with seven (7) degrees of freedom and model (3) is based on the generalized error distribution and with a shape parameter of (1.5). Each model uses the S&P500 index as proxy for the world market portfolio (WMP). The sample used, covers the period between May 1, 2014 and June 12, 2017 and consists of 163 weekly observations. z-statistics are reported in brackets under each coefficient. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Variables	Mean (1) Equation	Variance (1) Equation	Mean (2) Equation	Variance (2) Equation	Mean (3) Equation	Variance (3) Equation
Traded Volume	0.122**		0.106**		0.113**	
	(2.08)		(2.28)		(2.34)	
Hash Rate	-0.052		-0.036		-0.038	
	(-1.63)		(-1.36)		(-1.36)	
D2.Supply	-24.921		1.779		3.828	
	(-0.42)		(0.03)		(0.07)	
D1.Gold Price	-0.590*		-0.553*		-0.563*	
	(-1.72)		(-1.90)		(-1.88)	
D1.S&P500	-0.123		-0.071		-0.097	
	(-0.23)		(-0.17)		(-0.23)	
VIX index	-0.049		-0.034		-0.036	
	(-1.26)		(-1.13)		(-1.16)	
D1.Google Trends (weekly)	0.081**		0.083***		0.079**	
	(2.26)		(2.63)		(2.36)	
L.ARCH		0.105		0.113		0.101
		(1.18)		(1.14)		(1.06)
L.GARCH		0.719***		0.692***		0.710***
		(3.16)		(2.68)		(2.66)
Constant	0.138	0.001	0.100	0.001	0.104	0.001
	(1.36)	(1.01)	(1.24)	(0.97)	(1.25)	(0.90)
Observations	161	161	161	161	161	161
Gaussian[normal]	YES	YES	NO	NO	NO	NO
Student's t	NO	NO	YES	YES	NO	NO
Generalized error distribution	NO	NO	NO	NO	YES	YES

Table E

Johansen's Cointegration Test

This table provides the results for both the trace statistic and the max statistic for the Johansen's cointegration test used for the weekly analysis. Similar test has been made for the daily analysis and resulted in no signs for cointegration. On the other hand both the model including the S&P500 and the MSCI world indexes showed cointegration in weekly analysis. Also the information criteria (SBIC, HQIC) showed the same number of cointegration equations in the weekly data (1), while they confirm the no-cointegration result in daily data.

The results below are for the model that includes MSCI world index.

Cointegration Rank	Trace Statistic	5% critical value	Max Statistic	5% critical value
0	71.3031	47.21	42.7299	27.07
1	28.5731*	29.68	13.6904*	20.97
2	14.8828	15.41	10.4465	14.07
3	4.4362	3.76	4.4362	3.76
4	-	-	-	-

The results below are for the model that includes S&P500 index.

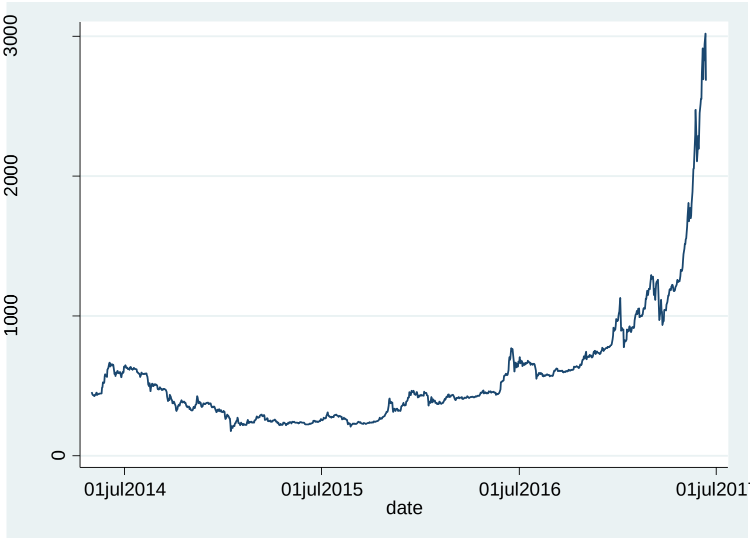
Cointegration Rank	Trace Statistic	5% critical value	Max Statistic	5% critical value
0	74.3034	47.21	45.8419	27.07
1	28.4615*	29.68	14.6035*	20.97
2	13.8581	15.41	9.7066	14.07
3	4.1515	3.76	4.1515	3.76
4	-	-	-	-

Charts A

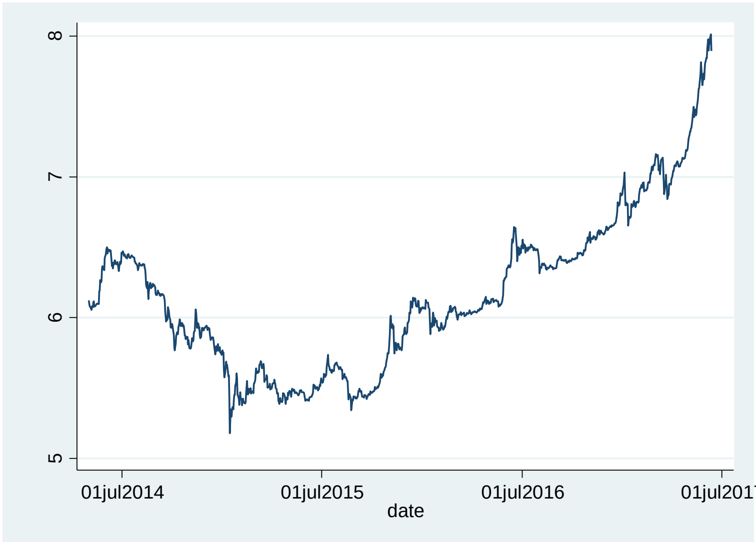
Graphs and relevant Transformations of the Variables used in research.

Graphs of the Dependent variable:

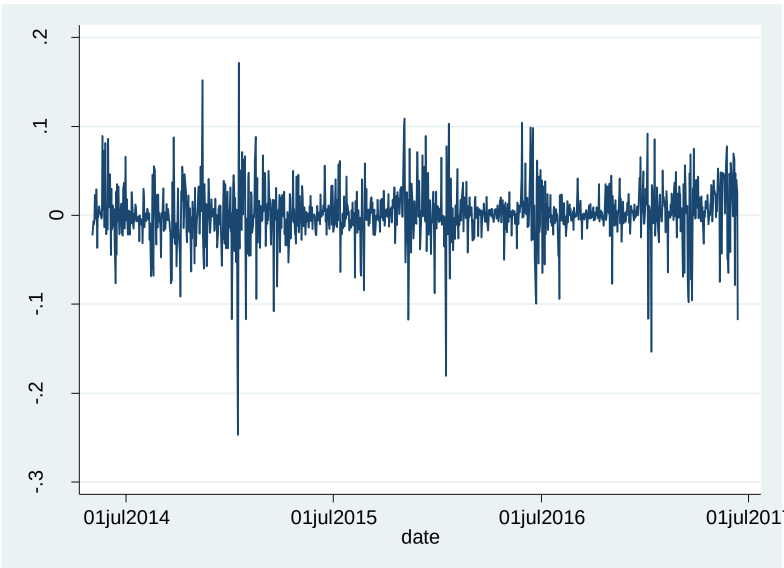
Price of Bitcoin (Price)



ln_price (natural logarithm of Price)

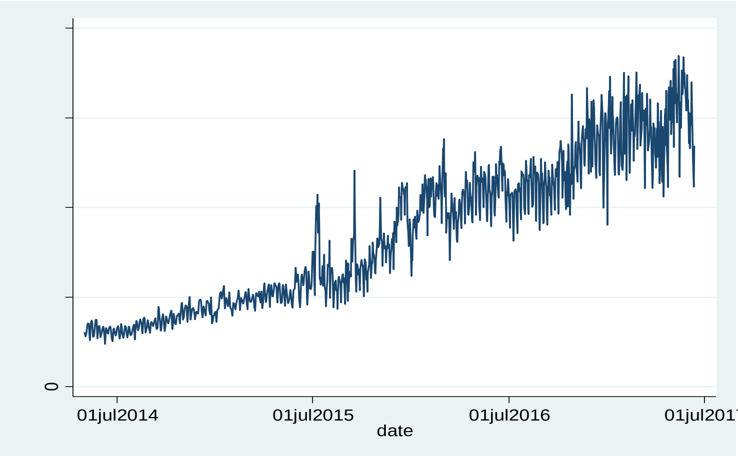


D1.ln_price (first difference of ln_price)

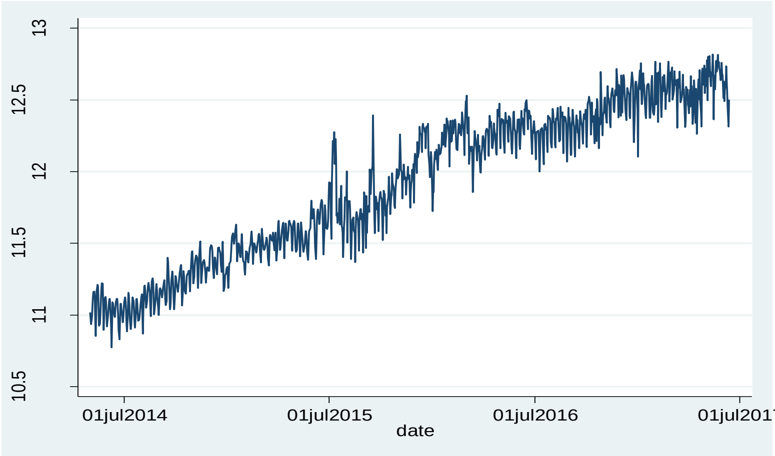


Graphs of the Independent variable:

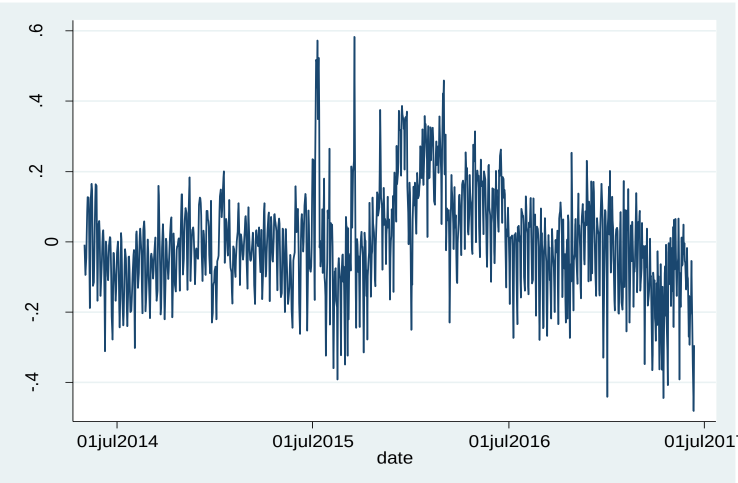
Transaction Volume (TV)



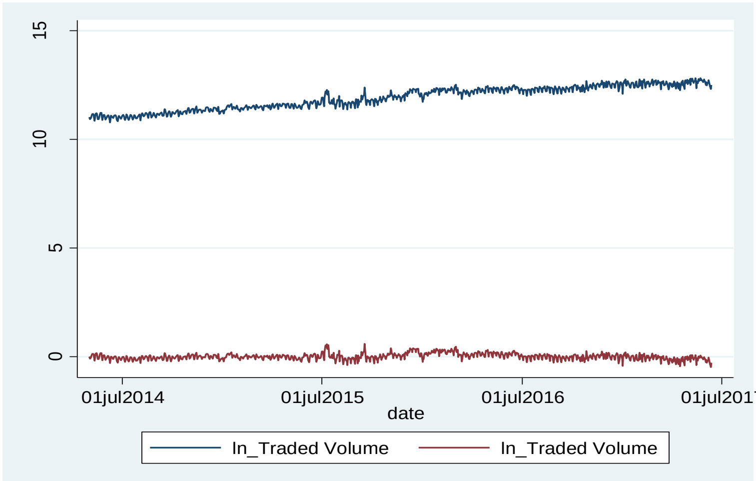
ln_TV (natural logarithm of TV)



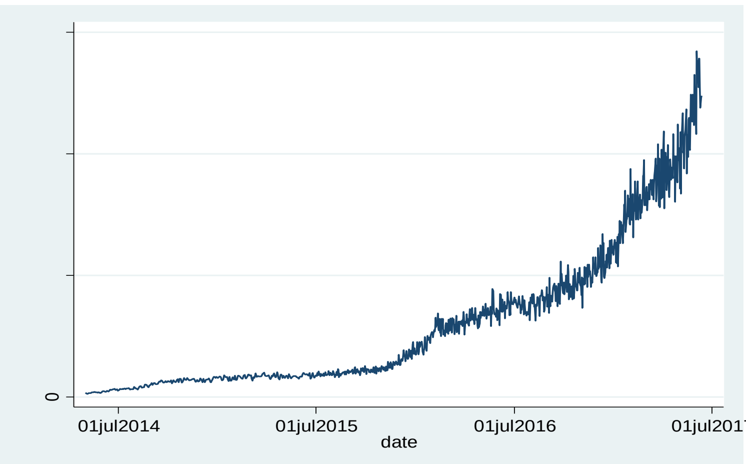
ln_TV_r (de-trended ln_TV)



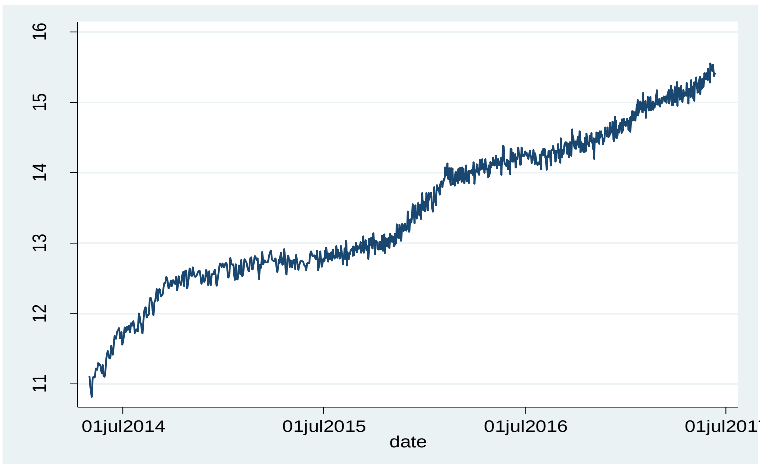
Both ln_TV and ln_TV_r



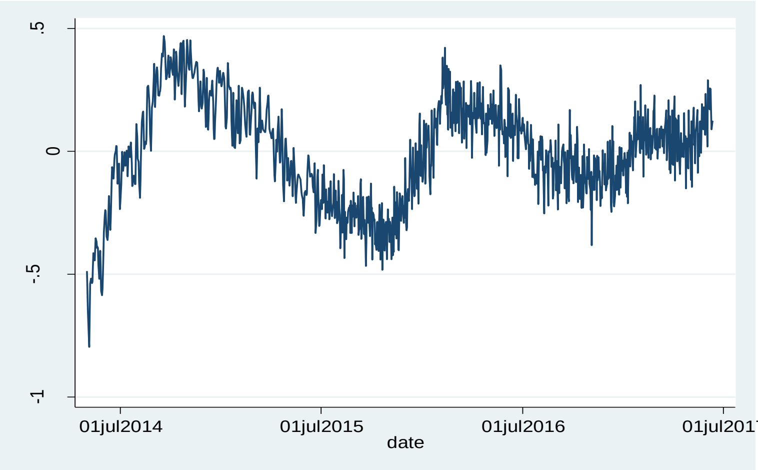
Hash Rate (HR)



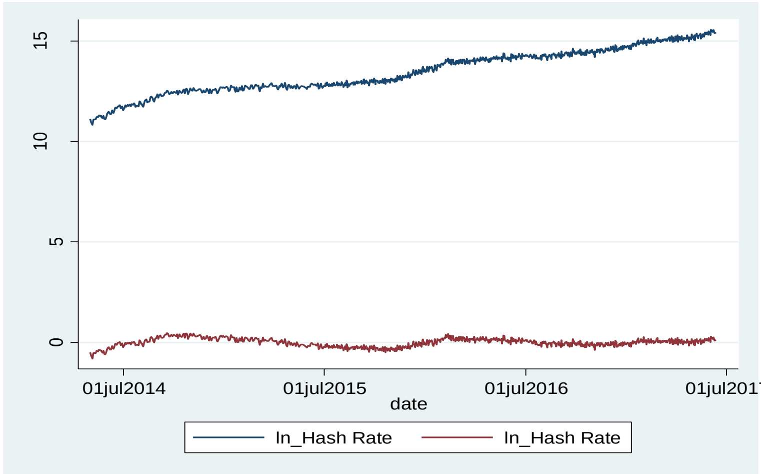
ln_HR (natural logarithm of HR)



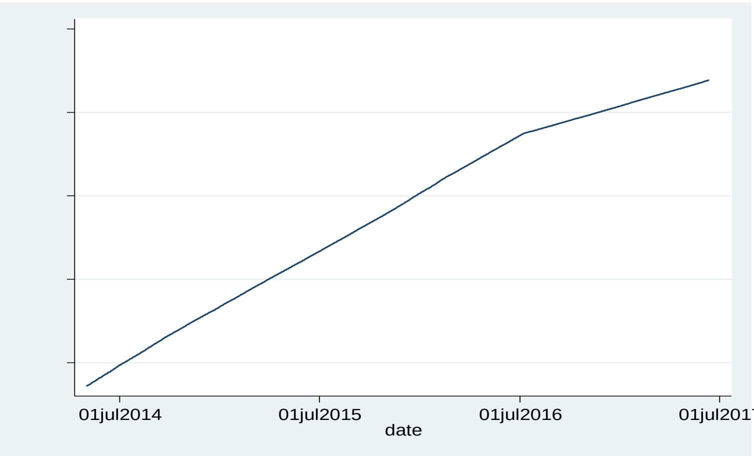
ln_HR_r (de-trended ln_HR)



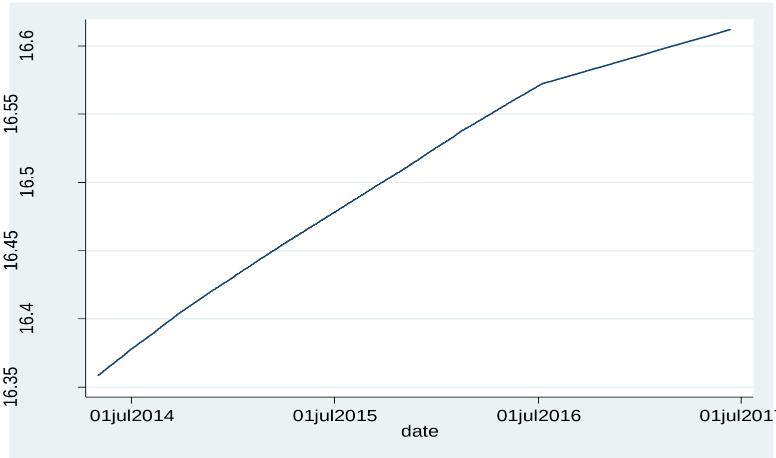
Both ln_HR and ln_HR_r



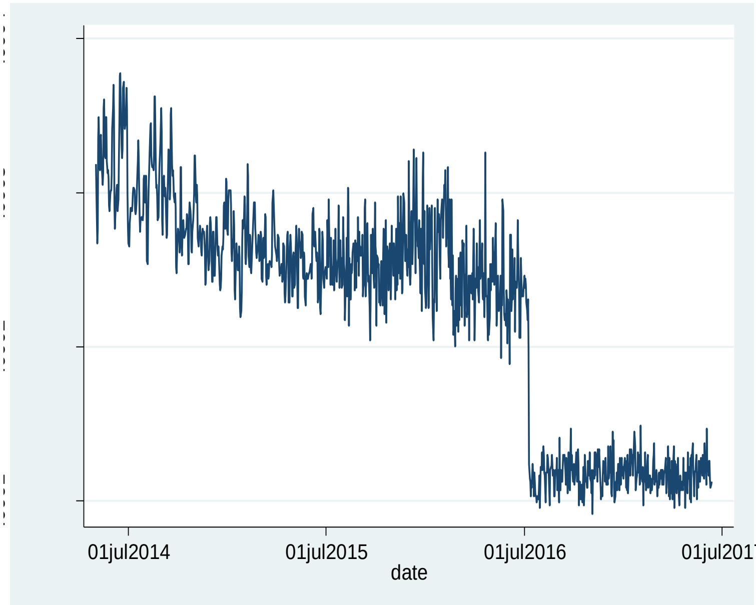
Supply (S)



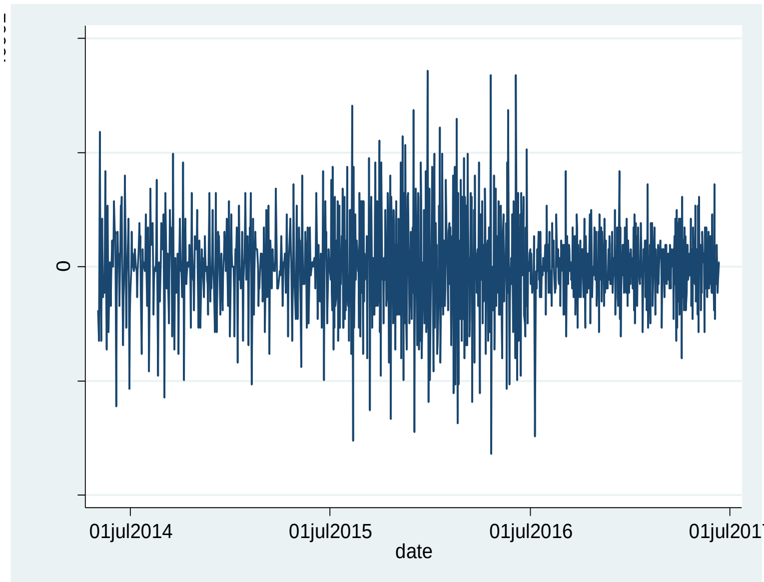
ln_S (natural logarithm of S)



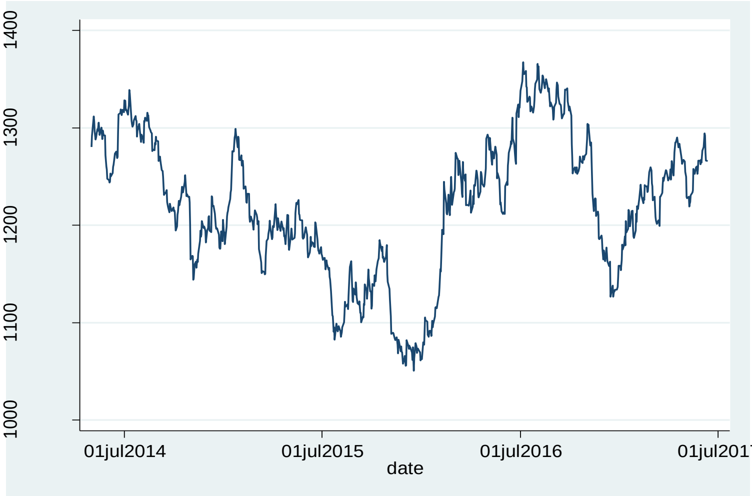
D1.ln_S (first difference of ln_S)



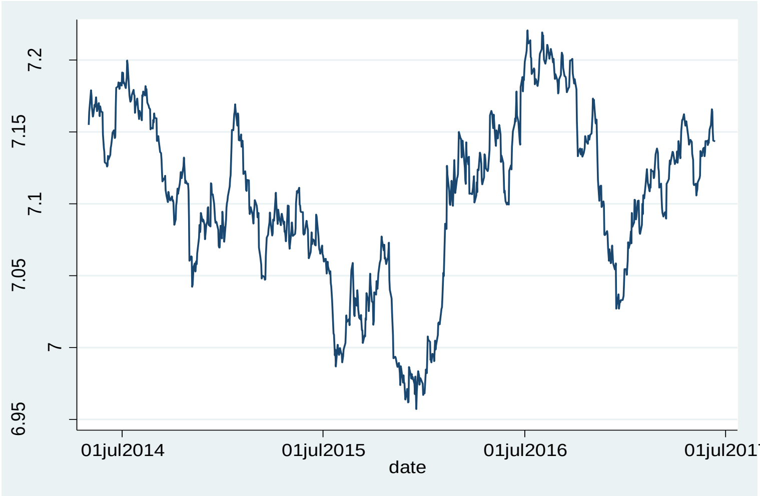
D2.ln_S (second difference of ln_S)



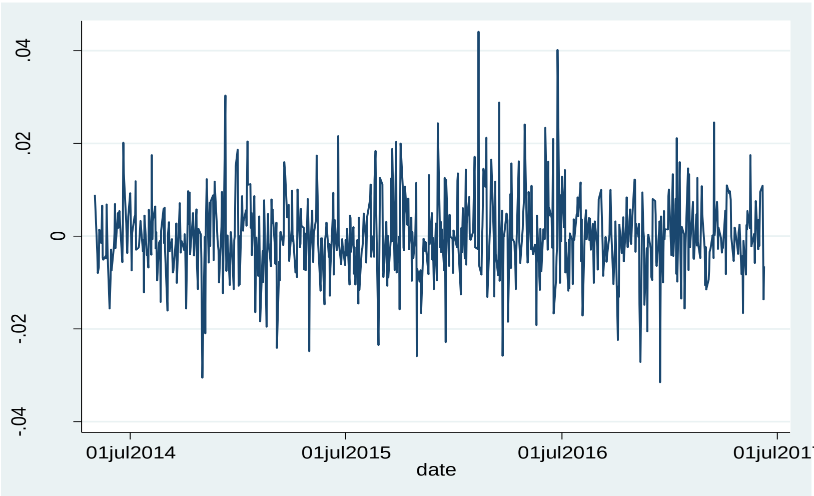
Gold Price (GP)



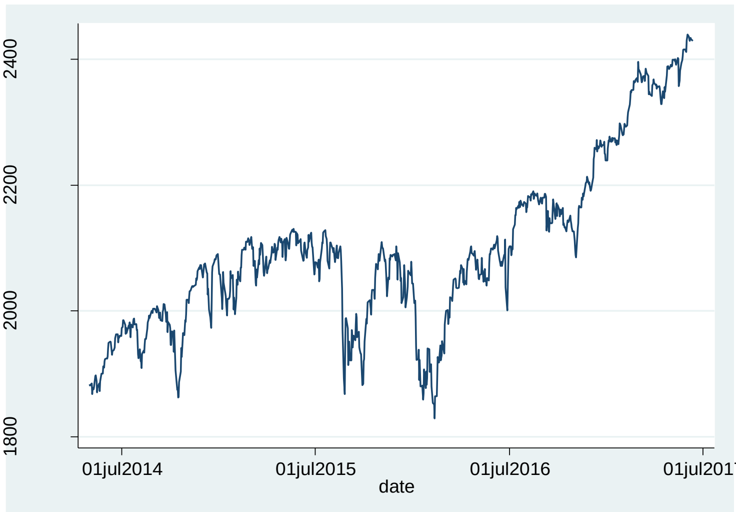
ln_GP (natural logarithm of GP)



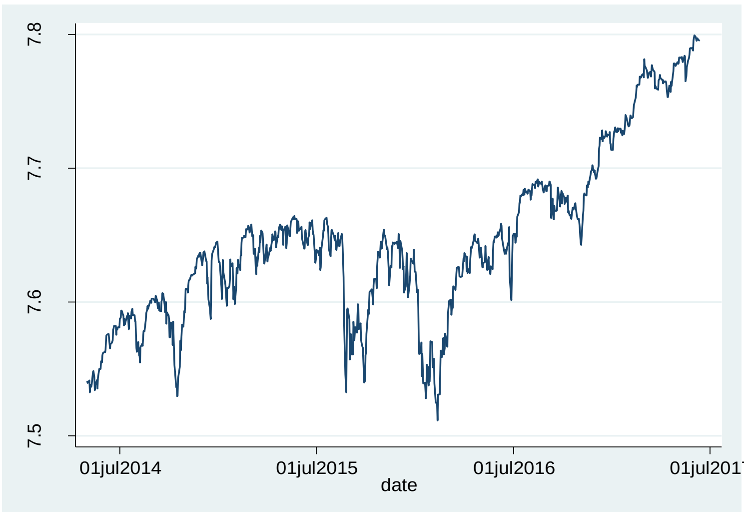
D1.ln_GP (first difference of ln_GP)



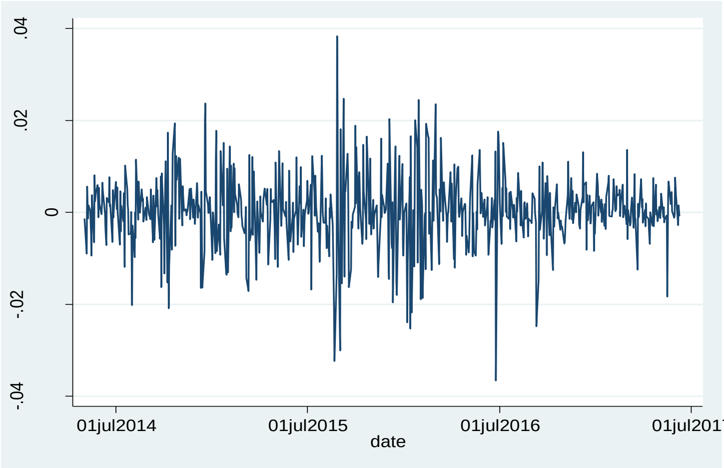
S&P 500 (SP500)



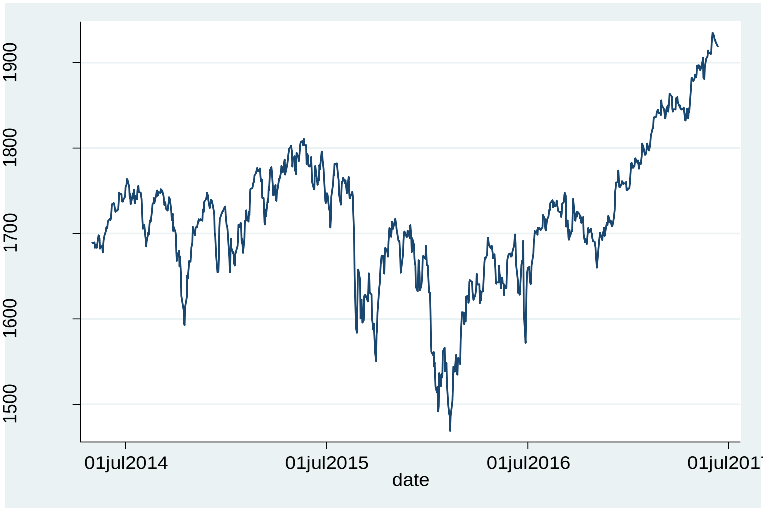
ln_SP500 (natural logarithm of SP500)



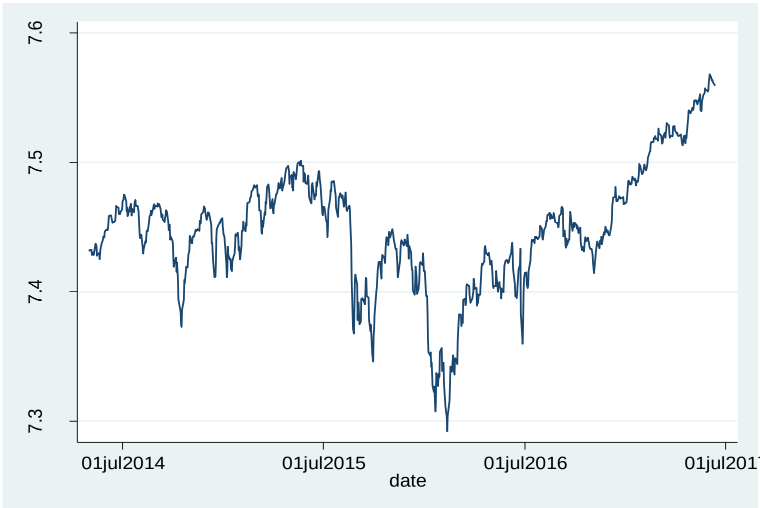
D1.ln_SP500 (first difference of ln_SP500)



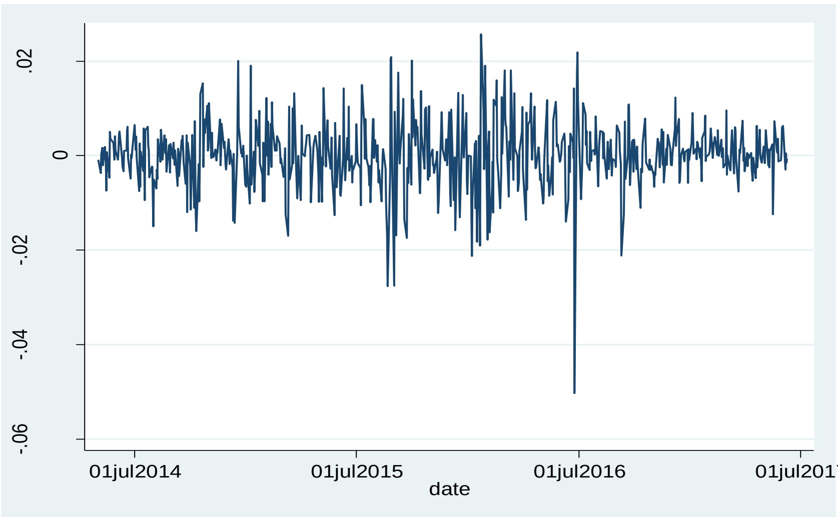
MSCI world (MSCI_WORLD)



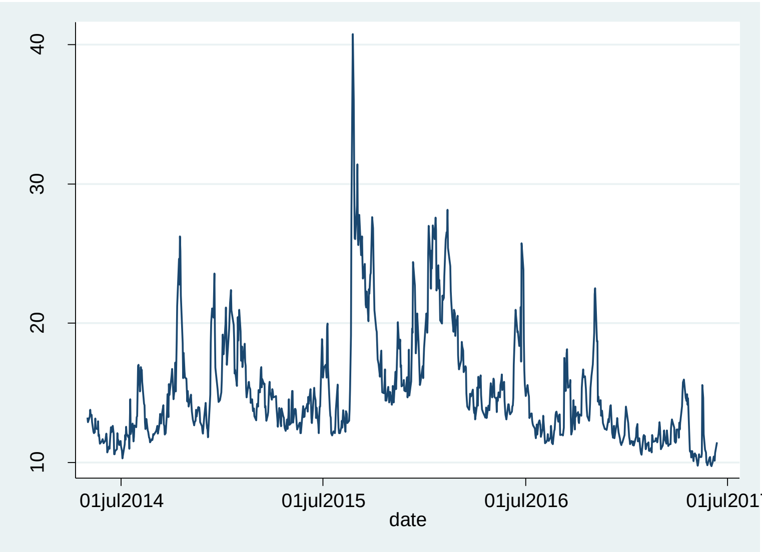
ln_MSCI_WORLD (natural logarithm of MSCI_WORLD)



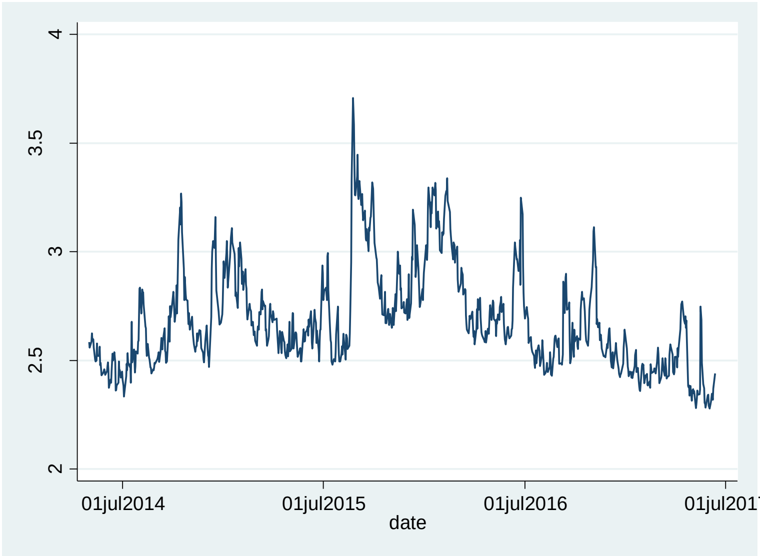
D1.ln_MSCI_WORLD (first difference of ln_MSCI_WORLD)



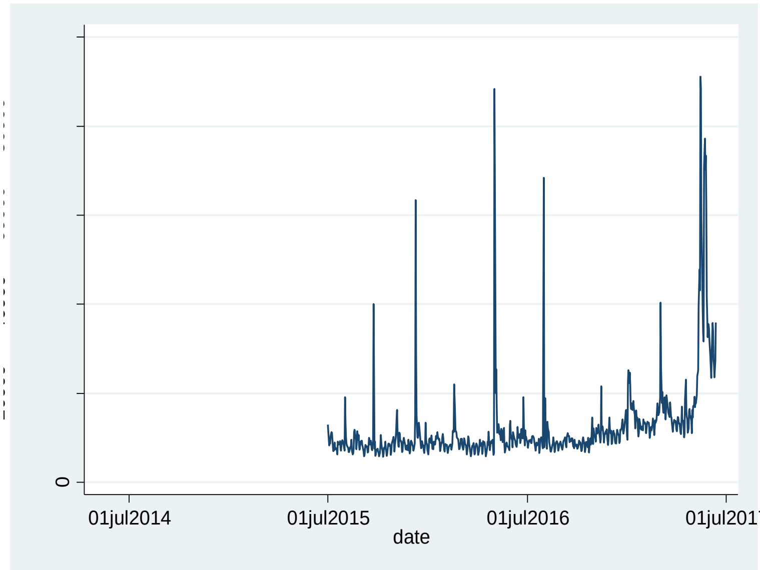
VIX index (VIX)



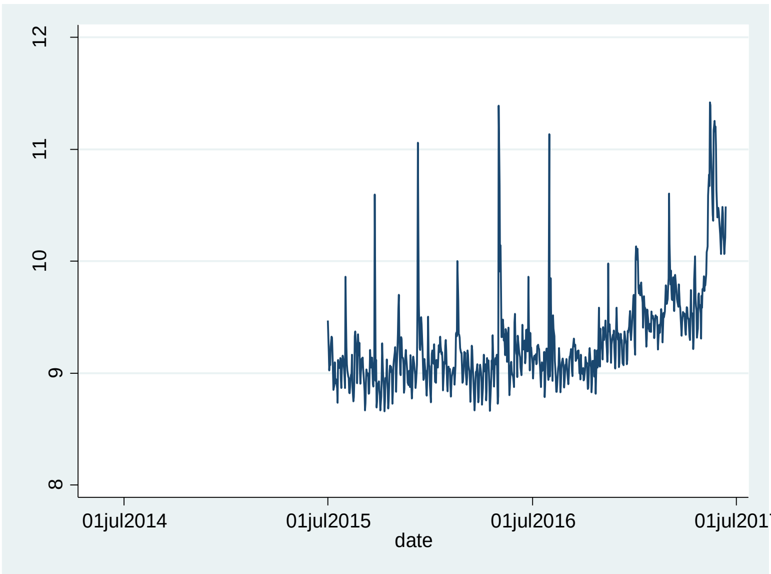
ln_VIX (natural logarithm of ln_VIX)



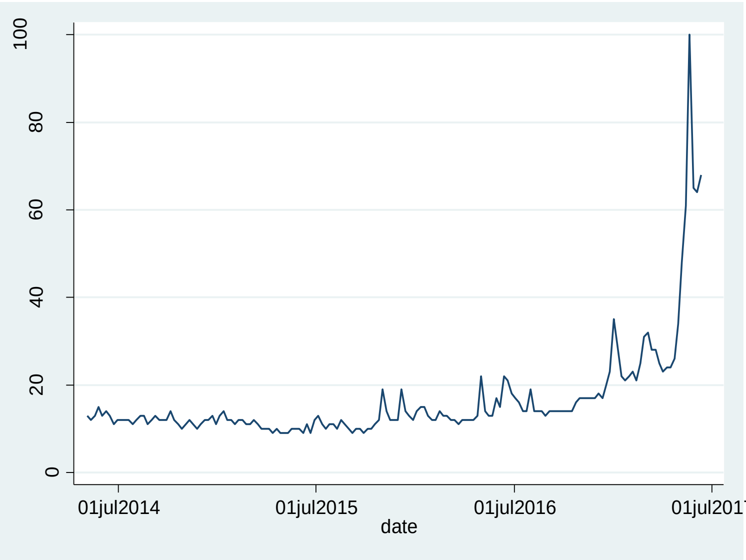
Wikipedia Views (daily)



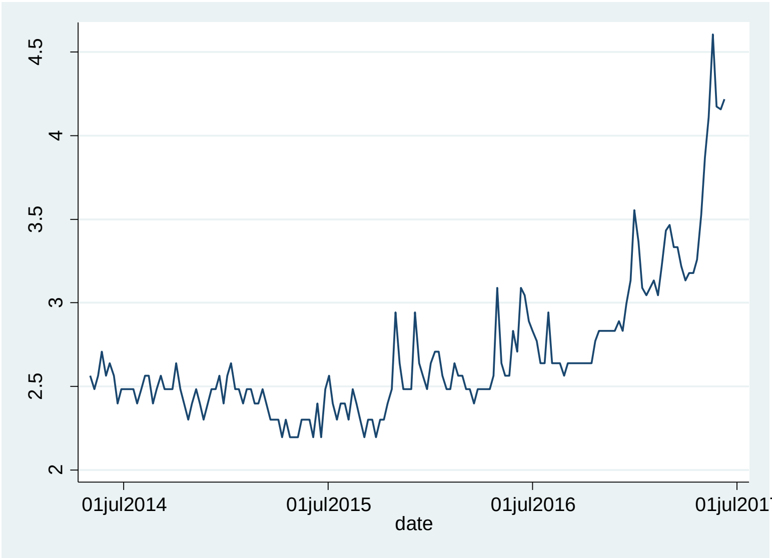
ln_WikipediaViews (natural logarithm)



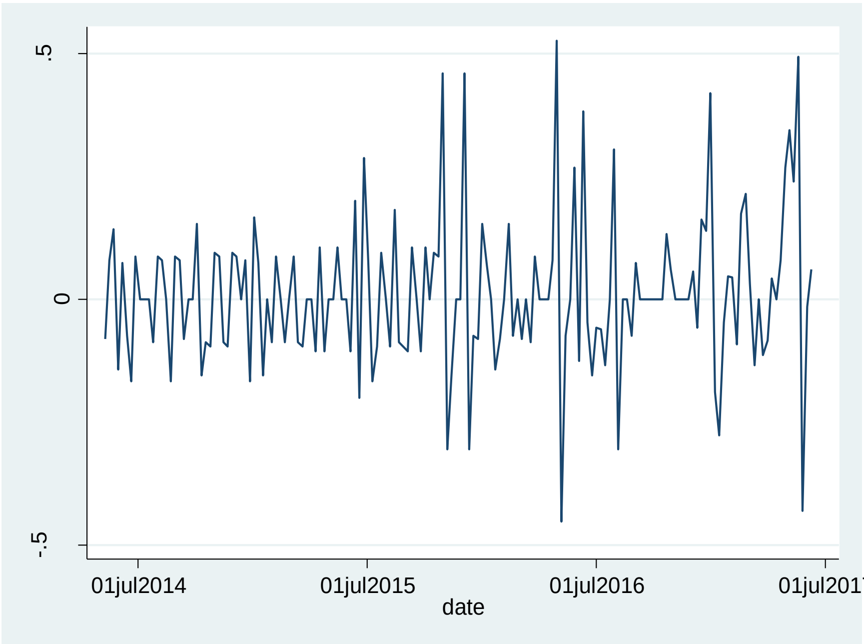
Google Trends (weekly)



ln_GoogleTrends_weekly (natural logarithm)



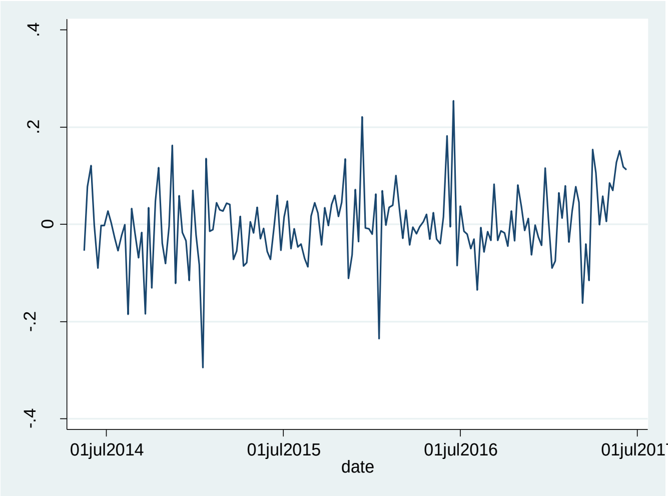
D1. ln_GoogleTrends_weekly (first difference of ln_GoogleTrends_weekly)



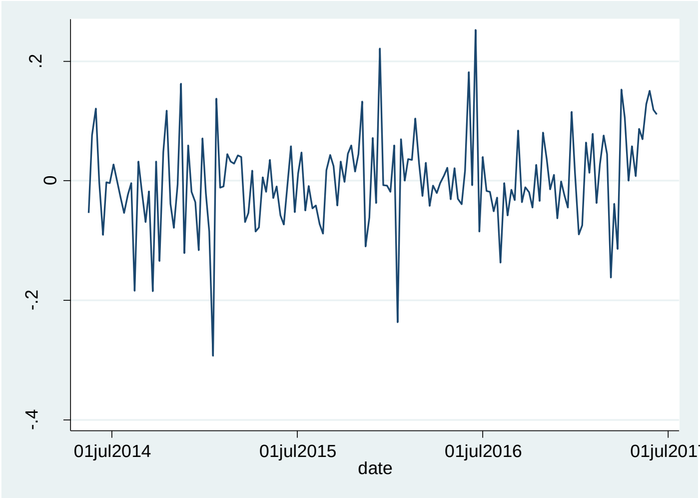
Charts B

Graphs depicting residuals in weekly OLS analysis.

For the S&P500 index:



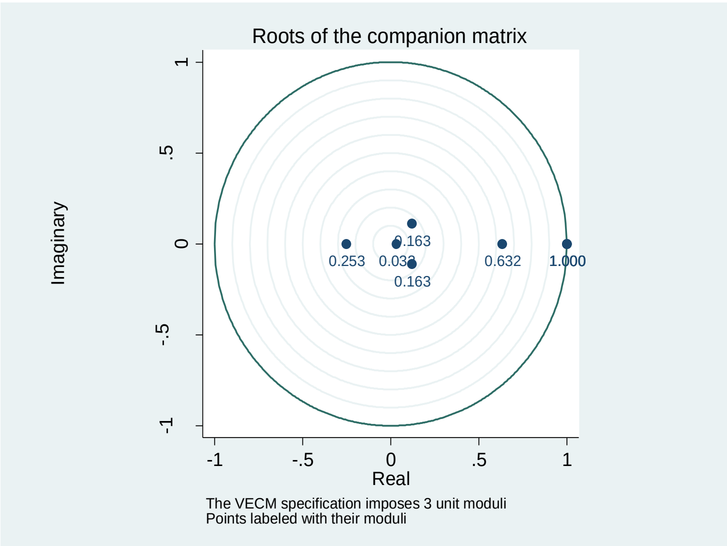
For the MSCI-world index:



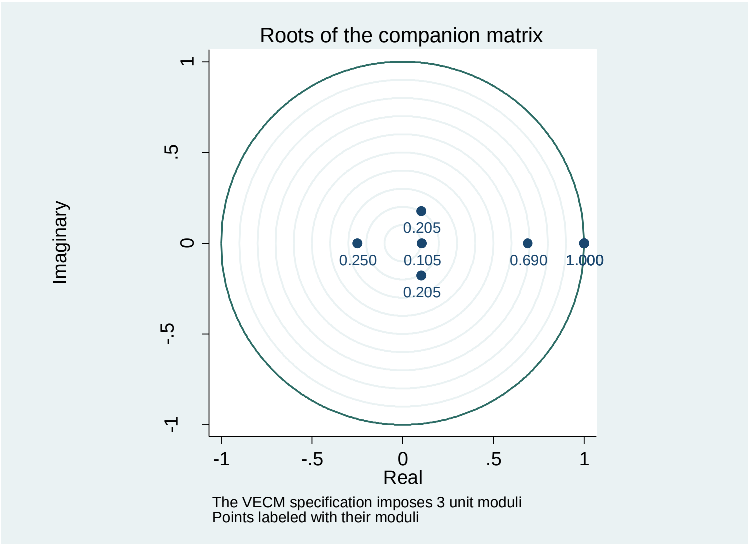
Charts C

Graphs depicting results of the VECM Stability Tests.

For the S&P500 index:



For the MSCI-world index:



References:

Asteriou, D., & Hall, S. G. (2015). *Applied econometrics*. Palgrave Macmillan.

Baek, C., & Elbeck, M. (2015). Bitcoins as an investment or speculative vehicle? A first look. *Applied Economics Letters*,*22*(1), 30-34.

Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012, February). Bitter to better—how to make bitcoin a better currency. In International Conference on Financial Cryptography and Data Security (pp. 399-414). Springer, Berlin, Heidelberg.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Bouoiyour, J., & Selmi, R. (2014). What Bitcoin looks like?.

Bouoiyour, J., & Selmi, R. (2015). What Does Bitcoin Look Like?. Annals of Economics & Finance, 16(2).

Bouoiyour, J., Selmi, R., & Tiwari, A. (2014). Is Bitcoin business income or speculative bubble? Unconditional vs. conditional frequency domain analysis.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?. Finance Research Letters, 20, 192-198.

Brandom, R. ,2014. Overstock.com is now accepting bitcoin. The Verge, Jan 18, 2014.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society, 1287-1294.

Breusch, T. S., & Godfrey, L. G. (1980). A review of recent work on testing for autocorrelation in dynamic economic models. University of Southampton.

Brito, J., & Castillo, A. (2013). Bitcoin: A primer for policymakers. Mercatus Center at George Mason University.

Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press.

Buchholz, M., Delaney, J., Warren, J., & Parker, J. (2012). Bits and Bets, Information, Price Volatility, and Demand for Bitcoin. Economics, 312.

Cheung, A., Roca, E., & Su, J. J. (2015). Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. Applied Economics, 47(23), 2348-2358.

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of BitCoin price formation. Applied Economics, 48(19), 1799-1815.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

DeLeo, M. J., & Stull, W. (2014). Does the Velocity of Bitcoins Effect the Price Level of Bitcoin. Temple University.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.

Durbin, J., and G. S. Watson. (1950). Testing for serial correlation in least squares regression. I. Biometrika 37: 409-428.

Dyhrberg, A. H. (2016). Hedging capabilities of bitcoin. Is it the virtual gold?. Finance Research Letters, 16, 139-144.

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar–A GARCH volatility analysis. Finance Research Letters, 16, 85-92.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

Franzen, C., ,2014. Zynga tests bitcoin payments for seven online games. The Verge, Jan 18, 2014.

Garcia, D., Tessone, C. J., Mavrodiev, P., & Perony, N. (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. Journal of the Royal Society Interface, 11(99), 20140623.

Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D. N., & Giaglis, G. M. (2015). Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M. C., & Siering, M. (2014). Bitcoin-asset or currency? revealing users' hidden intentions.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424-438.

Gronwald, M. (2014). The Economics of Bitcoins--Market Characteristics and Price Jumps.

Hamilton, J. D. (1994). Time series analysis (Vol. 2). Princeton: Princeton university press.

Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press on Demand.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. Economics Letters, 158, 3-6.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Scientific reports, 3, 3415.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PloS one, 10(4), e0123923.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. Journal of econometrics, 54(1-3), 159-178.

Li, X., & Wang, C. A. (2016). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. Decision Support Systems, 95, 49-60.

Moler, C. B. (2004). Numerical computing with MATLAB. Society for Industrial and Applied Mathematics.

Murphy, E., Murphy, M., & Seitzinger, M. (2015). Bitcoin: Questions, answers, and analysis of legal issues. Congressional Research Service.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2), 335-346.

Polasik, M., Piotrowska, A. I., Wisniewski, T. P., Kotkowski, R., & Lightfoot, G. (2015). Price fluctuations and the use of Bitcoin: An empirical inquiry. International Journal of Electronic Commerce, 20(1), 9-49.

Segendorf, B. (2014). What is bitcoin. Sveriges Riksbank Economic Review, 2, 71-87.

Takemoto, Y., & Knight, S. (2014). Mt. Gox files for bankruptcy, hit with lawsuit. Reuters, Tokyo.

Urquhart, A. (2016). The inefficiency of Bitcoin. Economics Letters, 148, 80-82.

Wikileaks, (2013). Transcript of secret meeting between Julian Assange and Google CEO Eric Schmidt, Jan 7, 2014

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In Handbook of digital currency (pp. 31-43).