COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS



RELATIONSHIP BETWEEN BITCOIN AND TRADITIONAL CURRENCIES

MASTER THESIS

2019 Bc. Michal ZÁHRADNÍK

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

RELATIONSHIP BETWEEN BITCOIN AND TRADITIONAL CURRENCIES

MASTER THESIS

Study programme:	Mathematical Economics, Finance and Modelling
Field of study:	1114 Applied Mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	Ing. RNDr. Peter Molnár, PhD.

Bratislava 2019 **Bc. Michal ZÁHRADNÍK**





Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Michal Záhradník
Study programme:	Mathematical Economics, Finance and Modelling (Single degree study, master II. deg., full time form)
Field of Study:	Applied Mathematics
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Relationship between Bitcoin and traditional currencies

Annotation: Cryptocurrency Bitcoin has experienced tremendous growth in terms of both popularity and market value. This master thesis will investigate the relationship between the Bitcoin price (in USD) and seven most traded currency pairs (AUD, CAD, CHF, EUR, GBP, JPY and NZD against US dollar). We will study comovement in price changes, and also volatility transmission between these currencies and Bitcoin.

Supervisor:	Ing. RNDr. Peter Molnár, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, DrSc.

Assigned: 30.01.2018

Approved: 30.01.2018 prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Vote of thanks: I would like to thank my supervisor Ing. Mgr. Peter Molnar, PhD. for professional help, ideas and suggestive comments without which this work would not be created.

Abstract

ZÁHRADNÍK, Michal: Relationship between Bitcoin and traditional currencies [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Ing. RNDr. Peter Molnár, PhD., Bratislava, 2019, 63p.

The main objective of this master thesis is to compare the market efficiency for Bitcoin and seven most traded currency exchange rates: AUD, CAD, CHF, EUR, GBP, JPY, and NZD traded against the US dollar using high-frequency data.

We began with an AR(1) model which examines efficiency using only one historical observation. Later we used more sophisticated statistical tests which take into account longer periods, such as The Ljung-Box test, The Runs test, The Bartels test, and The Automatic Variance ratio test. We will work with different frequencies and subintervals to produce more robust results.

Later we experimentally verified whether there is a long-term dependence on the high-frequency Bitcoin returns. We used the R/S and DFA Hurst exponent for that. Finally, we looked at the volatility transmission between the traditional currencies and Bitcoin using the E-GARCH model. The results are presented on the graphs and in the tables.

Keywords: Bitcoin, Currencies, High-frequency data, Effectivity testing, Ljung-Box test, Runs test, Bartels test, Automatic Variance ration test, Hurst exponent.

Abstrakt v štátnom jazyku

ZÁHRADNÍK, Michal: Závislosť medzi Bitcoinom a tradičnými menami [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Ing. RNDr. Peter Molnár, PhD., Bratislava, 2019, 63s.

Hlavným cieľom tejto diplomovej práce je za použitia vysokofrekvenčných burzových dát porovnať mieru trhovej efektivity pre Bitcoin a sedem najobchodovanejších menových kurzov: AUD, CAD, CHF, EUR, GBP, JPY a NZD obchodovaných voči americkému doláru.

Začali sme s AR(1) modelom, ktorý skúma efektivitu iba s prihliadnutím na jedno historické pozorovanie. Neskôr využívame komplexnejšie štatistické testy, ktoré zohľadňujú dlhšie obdobia ako: Ljung-Boxov test, Runsov test, Bartelsov test a Automatic Variance ratio test. Pre vytvorenie čo najrobustnejších výsledkov budeme pracovať s rôznymi frekvenciami a podintervalmi.

Neskôr sme experimentálne overili, či je prítomná dlhodobá závislosť vo vysokofrekvenčných výnosoch Bitcoinu. Na to sme použili R/S a DFA Hurstov exponent. Na záver sme nahliadli na prenos volatility medzi tradičnými menami a Bitcoinom použitím E-GARCH modelu. Výsledky prezentujeme na grafoch a v tabuľkách.

Kľúčové slová: Bitcoin, Menové kurzy, Vysokofrekvenčné dáta, Testovanie effektivity, Ljung-Boxov test, Runsov test, Bartelsov test, Automatic Variance ratio test, Hurstov exponent.

Contents

Introduction **9**

1 Cryptocurrency Bitcoin **11**

1.1 Safety of Bitcoin 11

1.2 Bitcoin as a currency 12

2 Data **14**

3 Methodology **16**

3.1 Linear regresion in time series 16

3.2 AR(1) model 17

3.3 Autocorrelation 17

3.4 Effective market and random walk hypothesis 17

3.5 Tests for independence 19

3.5.1 Ljung-Box test 19

3.5.2 Runs test 19

3.5.3 Bartels test 20

3.5.4 Automatic Variance Ratio Test 20

3.6 Hurst Exponent 22

3.6.1 R/S Hurst 22

3.6.2 DFA Hurst 23

3.7 GARCH model 25

3.7.1 EGARCH 27

3.7.2 News Impact Curve 28

3.7.3 Sign Bias Test, Negative Size Bias Test, Positive Size Bias Test 28

4 Results **30**

4.1 Descriptive statistics for minute returns 30

4.2 Correlation between Bitcoin and classic currency exchange rates 31

4.2.1 Moving correlation 32

4.3 1. Linear model for Bitcoin returns on exchange rate returns 33

4.4 2. Linear model of Bitcoin returns on exchange rate returns 35

4.4.1 Evolution of regression coefficients throughout the period 38
4.5 AR (1) model depending on data frequency 39
4.6 AR(1) model depending on time 42
4.7 Efficiency testing 44
4.7.1 Performing the Ljung Box test on returns. 45
4.7.2 Evolution of p-value of Ljungbox test over time 46
4.7.3 Performing Runs test on returns. 47
4.7.4 Evolution of p-value of Runs test over time 48
4.7.5 Performing the Bartels test on returns. 49
4.7.6 Evolution of p-value of Bartels test over time 50
4.7.7 Performing the Automatic Variance ratio test on returns. 51
4.8 The Hurst Exponent and long term dependence 52
4.9 The EGARCH MODEL and the evolution of the Laverage effect in time 54
4.9.1 Evolution of the News Impact Curve in time 56

Conclusion **58**

References **60**

Appendix A **63**

Introduction

Bitcoin appeared in 2009 at the time of the economic crisis. It was outlined in the article by Satoshi Nakamoto [8]. It was originated as an alternative to the national currencies based on peer-to-peer electronic transactions. Bitcoin payments go directly from one party to another without going through a financial institution. In the beginning, there was no interest in Bitcoin. It was just one of many new digital projects. It took several years for Bitcoin (in short BTC) to gain more attention.

In recent years, the general public has paid particular attention to Bitcoin especially during the period of rapid growth. The demand for Bitcoin has risen and as a result of that its price has increased. There is no underlying asset linked to Bitcoins price so a large number of articles nowadays deal with the issue of what is a real fair price for Bitcoin.

However, in our work we will focus on Bitcoin with comparison to the seven most traded currency exchange rates: AUD, CAD, CHF, EUR, GBP, JPY and NZD each traded against the US dollar. For these analyses, we will download high-frequency one-minute data for Bitcoin and currencies.

Since it is known that currencies as such have a high correlation between each other [14] we will investigate how they correlate with BTC. Then we will try to create a linear model which will predict Bitcoin returns using the currencies returns.

Afterward, we will use the AR(1) model as a model for Bitcoin returns and the currencies returns. We are going to display results for various data frequencies together with the evolution in the time.

Then we deploy a bunch of powerful statistical test to compare how bitcoin and traditional currencies stand up against the efficient market hypothesis developed in 1970 by Fama [10]. Efficiency testing for Bitcoin was firstly examined by Urquhart in (2016) on daily data. He found that Bitcoin is inefficient, but it is becoming more efficient over time. After him, many other authors tested the effectivity of Bitcoin, such as Bariviera [4] who used Hurst's exponent and Yonghong [7] who studied Time-

varying long term memory in the Bitcoin market. We will compare our result utilized using high-frequency data to theirs. Concept of using high-frequency data together with comparing the efficiency of Bitcoin to other currencies was not, to the best of our knowledge, been yet used.

The final issue which we will investigate in this thesis is volatility transmission from conventional currencies into the Bitcoin and comparing the evolution of the effect of the news on Bitcoins volatility.

This thesis is divided into 4 chapters. In the first one, we will introduce Bitcoin. The second will be about high-frequency data and the potential problems which arise with them. In the third chapter, there is a brief overview of the methodology and the statistical tests later used. The end of the fourth chapter is devoted to model for volatility modeling - GARCH. In the last 4th chapter we present the results of analyzes on various graphs.

1 Cryptocurrency Bitcoin

Bitcoin is the first decentralized cryptocurrency which remains markets leader. However, in the early days, Bitcoin was not very popular among the general public. It took several years for Bitcoin (in short BTC) to gain more attention. The exchanges on which Bitcoin was traded gradually began to emerge, and during this period, the public started to have a demand for the purchase, which of course, helped to increase its price.

However, in the first years, the absolute majority of BTC deals was secured by the at the time largest exchange called Mt. Gox. In 2014 this exchange made over 70% of all Bitcoin transactions. However Bitcoin experienced his first major problem because of this poorly secured exchange. When Mt. Gox has declared bankruptcy because over 800 000 BTC got stolen. At that time, this volume of Bitcoins was worth of more than 400 million US dollars. Many people afterwards lost their trust in Bitcoin as such. However, the problem was not with the protocol which protects Bitcoin. The only problem was with a third party. An exchange which was not well secured [1].

1.1 Safety of Bitcoin

Bitcoin is nowadays a safe way to transfer money across the globe. This is due to the invention of blockchain, a technology that BTC uses and on which most of today's cryptocurrencies works. Bitcoin's transactions work on three basic principles: decentralization, cryptography and blockchain.

Transactions verification is happening on many computers around the world at the same time. Everyone can join this controlling network. When one of the computers fails, it is replaced by another which makes Bitcoin decentralized and independent of specific authority which in the case of the traditional currency exchange rates are banks. During transaction unique addresses of senders, receivers and amount of bitcoin is sent to all computers involved in worldwide verification. Subsequently, this transaction is written into a blockchain, which is merely a digital accounting book. After writing into

blockchain it has to be verified. People whose computers do this verification process are called miners, after the verification, some of them are rewarded with a small amount of newly created coin for their exertion. The total number of bitcoins never exceeds 21,000,000 because the rewards for verifying blocks are decreasing over time.

Revolutionary cryptography is involved in verification processes thanks to what no one has been able to decrypt and produce his own fake Bitcoins so far. The Bitcoin was the first world-wide spread digitally currency whose security is not yet defeated. The entire system is secured by signing the transaction with a private key. A controller then verifies whether you have enough BTC at your address and whether your transaction is valid. With the breakthrough use of cryptography, it is possible to verify if the transaction is correct and only the signature is sufficient. The private key always remains private. Therefore, when you own a Bitcoin, you do not save physical data about coins, but only the private key that in the accounting book shows that you are the real owner of these Bitcoins [8].

1.2 Bitcoin as a currency

Currently, Bitcoin still does not function as a currency in the real sense of the word. The most payments are still being exchanged according to the current rate because prices for goods are fixed in US dollars. Products typically do not have an exact price in BTC. The reason why it is nowadays almost impossible is the high volatility of Bitcoin's price. Nevertheless, in 2017 the Japanese government officially introduced Bitcoin as the legal option for payments in Japan. Microsoft, for example, accepts payments in Bitcoin, though they can only be used to purchase games, movies and apps in Windows and Xbox stores. On the other hand with Bitcoins you can not buy a computer in their official store.

With Bitcoin it is challenging to replace daily payments in the domestic currency, for example in grocery stores, because the process of verifying a transaction can take several minutes. Bitcoin's great advantage is cheap and fast money transfers world wide, as well as its anonymity for these transactions. Because of the anonymity and irre-

versible nature of the transaction, it is often used for illegal online transactions.

The market value of Bitcoin as such is not guaranteed by any economic player, such as large world banks, the European Central Bank or the American Fed. It is not known that any of these institutions would interfere with the Bitcoin price. The Bitcoin price results only from the expected future profits of people who trade with it. Investors sentiment is therefore a crucial variable. Short-term investors, speculators, noise traders and part of the general public who are interested in high profits are the most numerous BTC traders [1].

2 Data

Currencies are traded 24 hours a day 7 days a week. A significant amount of the trades are happening automatically and speed is crucial. It decides whether investors make a profit or loss. Therefore, analyzes which try to model or predict Bitcoin's future price based on daily data can overlook important information. That is why we decided to use for our analyses high-frequency minute data. Compared to daily data, we have 1440 data points for each day. This makes the analysis more accurate, but computationally more complex.

We have downloaded Bitcoin price data from two sources, kaggle.com and bitcoincharts.com, for the period from 1/2012 to 1/2019. Both data sets contain gaps so we created a combined intersected data set with 16% fewer gaps. It was essential to verify that the datasets do match. The data show the price evolution on the cryptocurrency exchange called bitstamp.net. The selection of this exchange was driven by data quality. According to [15], the price of Bitcoin at the top four exchanges worldwide behaves similarly. Hence, our analysis should be independent of the Exchange's choice.

The seven most traded currency exchange rates (AUD/USD, CAD/USD, CHF/USD, EUR/USD, GBP/USD, JPY/USD and NZD/USD) were downloaded from page HistData.com. We have transformed both data sets to Unix Timestamp time format. Timestamp displays year, day and time in just single number - a number of seconds from midnight on January 1st, 1970. For the needs of working in the R environment, it is a significant advantage. Unfortunately, the data have gaps (mainly due to API issues or exchange technical problems). The majority of these missing gaps is not longer than one minute. These gaps are spaced almost evenly across the data set. Although every single time-series do not have many gaps, after making an intersection, in which we have available all the data, the number of gaps increases. After all, we have a vector of 2.6 million observations that is capturing approximately 71% of all minutes from 1/2012 to 1/2019. Some of the more significant gaps in the data can cause abnormalities - events of extreme observations when they are located at the times when the price has risen or soared sharply. Just then there may be a change between two data points

that could not have happened in one minute. Distribution of gaps depending on their length and count is presented in the following charts 1&2.

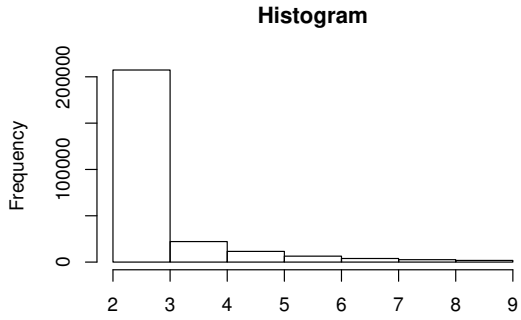

Figure 1: Frequency of missing gaps with size less than 9minutes.

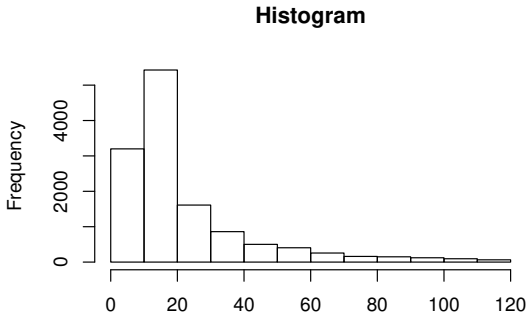

Figure 2: Frequency of missing gaps with size between 9 and 120 minutes.

To make sure that these gaps do not affect our analysis, we worked with a vector containing NA values wherever data was not available.

3 Methodology

In the next chapter, we present the basic properties of the time series, linear regression, the hypothesis of an efficient market, tests for independence of returns within the time series and GARCH models which we will use. In evolution of this chapter, we used information mainly from sources [10],[11],[12] and [18],[19],. . . ,[29].

First we test data for stacionarity using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test with the null hypothesis that a time series variable is trend-stationary against the alternative hypothesis of a unit root [17].

Time series for Bitcoin prices and currency exchange rates show signs of non-stationarity (p-values of KPSS test less than 0.01), so we produce stationary data using logarithmic transformation:

$$r_t = \log \frac{P_t}{P_{t-1}}, \tag{3.1}$$

Where r_t is a return of time series and $\log(P_t)$ is a natural logarithm of price at the time t.

3.1 Linear regresion in time series

The linear time series $\{(y_1, \ldots, y_t); t \in \mathbb{N}\}$ model can be expressed as:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_m x_{m,t} + \epsilon_t \tag{3.2}$$

where $x_{i,t}$ is the value of the i-th explanatory variable at the time t. Error term at time t is expresed with ϵ_t and $\beta_0, \ldots, \beta_m$ are parameters of model. With time index t .

3.2 AR(1) model

AR(1) model is a linear regression model for the time series, which lineary predicts output variable depending on its previous value.

$$r_t = \beta r_{t-1} + u_t \tag{3.3}$$

where β is a constant and u_t is white noise.

3.3 Autocorrelation

Autocorrelation [12] is the correlation of a variable with a delayed copy of itself at the different times. It is also known as a serial correlation. The size of autocorrelation coefficients R_k may be calculated using:

$$R_k = \frac{\sum_{t=1}^{n-k}(r_t - \hat{r})(r_{t+k} - \hat{r})}{\sum_{t=1}^{n}(r_t - \hat{r})^2} \tag{3.4}$$

The number of the time steps k between the variables is known as the lag. We will work with the estimated autocorrelation functions acf and with the estimated partial autocorrelation functions pacf. These functions help to describe the patterns within a given data series. They are often used as a guide to ARIMA order that might fit the data. An autocorrelation function is a useful tool for detecting the randomness in, for example, returns. For random data, autocorrelations and partial autocorrelation should be near zero for any and all time-lag separations. The presence of one or more significantly non-zero autocorrelations might suggest non-random process [12].

3.4 Effective market and random walk hypothesis

Let us with regard to [11] define P_t as stock price at time t and $p_t = \log P_t$ as a stochastic process defined using the recursive relation:

$$p_t = \mu + p_{t-1} + \epsilon_t \tag{3.5}$$

$$r_t = p_t - p_{t-1} = \mu + \epsilon_t \tag{3.6}$$

Where r_t is a return for the period of length one, μ is a constant drift parameter and ϵ_t is the random error. The ϵ_t $\forall t$ are assumed to be white noise - independently and identically distributed (i.i.d.) and having a normal distribution.

The essence of the random walk hypothesis is that the errors term (disturbances) ϵ_t are serially uncorrelated. Hence it is not possible to forecast future returns using past returns [11].

An efficient market hypothesis (EMH) is one of the fundamental theories of financial economics, according to which asset prices fully reflect all the information already available. A market should only react to new information. It is not possible in the long term to buy undervalued stocks and to sell overvalued. When the market is efficient more money can be earned only by chance or by investing in more risky/volatile assets [10, 11].

There are three types of market efficiencies. In the scope of this thesis, we will deal with a so-called weak form in which investors cannot make a lasting profit by using previous public information. We assume that the return series follow a random walk. Classic financial derivatives, such as stocks or bonds, usually meet the criteria of this hypothesis [11]. However there have been several cases in history, where the hypothesis of an efficient market has been violated. For example, the monthly data for the Swedish stock market for the years 1919-1990 showed a positive autocorrelation of up to one year.

3.5 Tests for independence

In the time series analyses is not always clear whether a pattern is present in the data. That is why it is appropriate to deploy some mathematical methods to verify whether any form of data dependency exists. We use a set of tests: the Ljung-Box test, the Runs test, the Bartels test, the Automatic Variance Ratio test, the Hurst exponent that could provide a comprehensive view of the existence of dependency in data.

3.5.1 Ljung-Box test

The Ljung-Box test [30] is based on the Box-Pierce test. It tests whether the correlation of a random process with delayed copy of itself is significantly different from zero. Test hypothesis:

H0: Correlations in the time series data are a result of randomness.

Calculation of test statistics:

$$Q(k) = n(n+2)\sum_{i=1}^{k}\frac{\hat{\rho_i^2}}{n-i} \tag{3.7}$$

Where $\hat{\rho_i^2}$ are accumulated sample autocorrelations at lag i and k is the number of lags being used. Under null hypothesis test statistics $Q(k)$ follows $\chi^2_{1-\alpha,k}$ distribution, where α is a significance level, and k is the degree of freedom.

3.5.2 Runs test

The runs test [18] is a non-parametric statistical test which can be used to determine if a data set was created randomly.

H0: Each sample value from a sequence is independently drawn from the same distribution.

The run is defined as a consecutive series of positive values or negative values. The number of positive or negative values is the length of the run. In a random data set, the probability that the r_i value is larger or smaller than the r_{i-1} value follows a binomial

distribution, what creates the fundamentals for Runs test with test statistics:

$$Z = \frac{R - \overline{R}}{S_R} \tag{3.8}$$

Where R is the observed number of runs, $\overline{R}$, is the expected number of runs, and S_R is the standard deviation of the number of runs. The values of $\overline{R}$ and S_R are computed as follows:

$$\overline{R} = \frac{2n_1n_2}{n_1 + n_2} + 1 \tag{3.9}$$

$$S_R^2 = \frac{2n_1 + n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \tag{3.10}$$

with n_1 and n_2 denoting the number of positive and negative values in the tested series. The runs test rejects the null hypothesis if $|Z| > Z_{1-\alpha,2}$ at significance level α. The value of the test statistics Z asymptotically follows, as the number of data increases, the normalized normal distribution.

3.5.3 Bartels test

Bartels test [19] is another non-parametric test for serial correlation in a sequence of observations. The test statistics are identical to the Von Neumann Ratio test [19] but it is calculated on the ranks. The advantage of the Bartest test versus Von Neumann Ratio test is that it works better for large data sets with which we will work. The null hypothesis is identical to previous tests and the test statistic has the form:

$$RVN = \frac{\sum_{i=1}^{n-1}(R_i - R_{i+1})^2}{\sum_{i=1}^{n}(R_i - \overline{R})^2} \tag{3.11}$$

where R_i is the rank of the observation r_i. For the number of data $n > 100$, the null hypothesis test statistics follows normal distribution: $N(2, \frac{20}{5n+7})$.

3.5.4 Automatic Variance Ratio Test

Automatic Variance Ratio Test (AVR test) is the variance ratio test with holding period value chosen "automatically" by a data dependent procedure [20]. The main advantage of the AVR test is that it does not require an assumption about the underlying distribution. Test hypothesis about the return series is:

H0: r_t is serially uncorrelated.

We want to verify if price movements P_t are the result of a random walk process. To achieve this returns r_t has to be i.i.d process. However, we will consider P_t to be random walk when r_t will be serially uncorrelated [20].

With the use of consistent variance ratio estimator $\widehat{VR(k)}$:

$$\widehat{VR(k)} = 1 + 2\sum_{i=1}^{T-1} m(k)\hat{\rho(i)} \tag{3.12}$$

where $\hat{\rho(i)}$ and $m(x)$ are defined as:

$$\hat{\rho(i)} = \frac{\sum_{t=1}^{T-i} r_t r_{t+i}}{\sum_{t=1}^{T} r_t^2} \tag{3.13}$$

$$m(x) = \frac{25}{12\pi^2 x^2}\left[\frac{sin(\frac{6\pi x}{5})}{\frac{6\pi x}{5}} - cos(\frac{6\pi x}{5})\right] \tag{3.14}$$

the standardized variance ratio test statistics $VR(k)$ can be calculated as:

$$VR(k) = \sqrt{\frac{T}{k}}\frac{(\widehat{VR(k)} - 1)}{\sqrt{2}} \tag{3.15}$$

$VR(k)$ converges, under null hypothesis, in distribution to N(0,1) as $T \to \infty$, $k \to \infty$ and $\frac{T}{k} \to \infty$.

The variance ratio test defined in [20] depends on the election of lag k. It is well known that the selection of the truncation point can have a tremendous impact on the estimated values on the variance ratio test statistics. Therefore calculating optimal truncation points, which meet all the previously specified conditions, is by [20] done in the following steps :

1. Choose an approximating AR model for r_t.

2. Estimate the optimal truncation points using formula 6.2 in [21].

3. Using optimal truncation points, calculate the variance ratio test statistics.

3.6 Hurst Exponent

Long-term memory is an essential element in the behavior of the time series. Financial series are generally not only dependent on a few previous observations. Possible long-term dependence in the data may not be revealed by the previous tests. It would be a good idea to evaluate the larger observation interval as a whole. The Hurst exponent [22] is often used for such evaluation. Hurst exponent was originally introduced in hydrologic time series measuring the height of the Nile river.

3.6.1 R/S Hurst

Let $\{X_k\}$ be a sequence of not necessarily independent random variables and we define Y_n as the partial sum $Y_n = X_1 + \dots X_n$. Than Adjusted Range $R(n)$ was originally defined by Hurst as:

$$R(n) = \max_{1 \le k \le n} \left\{ Y_k - \frac{k}{n} Y_n \right\} - \min_{1 \le k \le n} \left\{ Y_k - \frac{k}{n} Y_n \right\} \tag{3.16}$$

Using $R(n)$ Hurst later defined normalized adjusted range called R/S(n) (Rescaled Adjusted Range statistic):

$$R/S(n) = \frac{\max_{1 \le k \le n} \left\{ Y_k - \frac{k}{n} Y_n \right\} - \min_{1 \le k \le n} \left\{ Y_k - \frac{k}{n} Y_n \right\}}{\sqrt{\frac{1}{n} \sum_{k=1}^{n} (X_k - \frac{1}{n} Y_n)^2}} \tag{3.17}$$

$R/S(n)$ is averaged over the regions $[X_1, X_n], [X_{n+1}, X_{2n}] \dots [X_{(f-1)t+1}, X_{f*n}]$ where $f = floor(\text{number of observations}/n)$. Hurst exponent scales as time increases by a power law:

$$R/S(n) = cn^H \tag{3.18}$$

C is a constant and H is Hurst exponent. To obtain estimated Hurst exponent - a slope of the fitted regression line through values of (R/S) versus n on logarithmic axes. Obtaining the estimator in this way is a maximum likelihood method.

3.6.2 DFA Hurst

Detrended Fluctuation Analysis is a method first introduced in 1994 in [23]. It is an extension to the R/S method. The Hurst exponent calculated by DFA can, in contrast to the Hurst exponent calculated by R / S, acquire values greater than 1. The value greater than one indicates a non-stationary series, which needs to be detrended before continuing.

The DFA calculation with which we will continue to work comes from [24]. At the beginning, it is necessary to divide the time series $\{X_i\}_{i=1}^{N}$ into d sub-series, each of which will have the length n. For all these sub-series, we calculate cumulative sum:

$$Y_{i,m} = \sum_{j=1}^{i} X_{j,m} \tag{3.19}$$

Where $i = 1, \ldots, n$. Then we place a regression line over the points $(Y_{1,m}, \ldots, Y_{n,m})$:

$$\hat{Y}_m(t) = a_m t + b_m \tag{3.20}$$

Where $t = 1, \ldots, n$. Than we calculate the standard deviation of the time series:

$$F(m) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y\text{–}i, m - a_m(i) - b_m)^2} \tag{3.21}$$

Afterward, we calculate mean from these standard deviations across all d subseries:

$$F(n) = \frac{1}{d}\sum_{m=1}^{d} F(m) \tag{3.22}$$

In the end, as with the R/S method, the Hurst exponent value can be obtained by employing the simple regression line: $\log(F(n)) \sim Hlog(n) + \epsilon$. Where the slope of this line H is called DFA Hurst exponent.

1. $H = 0$ indicates an uncorrelated random series
2. $0 < H < 0.5$ indicates anti persistent series (mean-reverting). A down value is more likely to be followed by an up value.

3. $0.5 < H < 0.1$ indicates persistent series (trend reinforcing). Next value is more likely to be the same as the previous one.

A big slice of the financial market has values of Hurst exponent above 0.5. A value significantly below 0.5 indicates a mean reversion process. It means that values tend to come back to the mean and it is relatively easy to create a universal profitable strategy.

An example of such a strategy is to consider two moving average processes of different lengths. The profitable long-term strategy is to sell (short), when the larger of these moving averages goes below the smaller one, and buy vice-versa. As shown in the Figure 3.

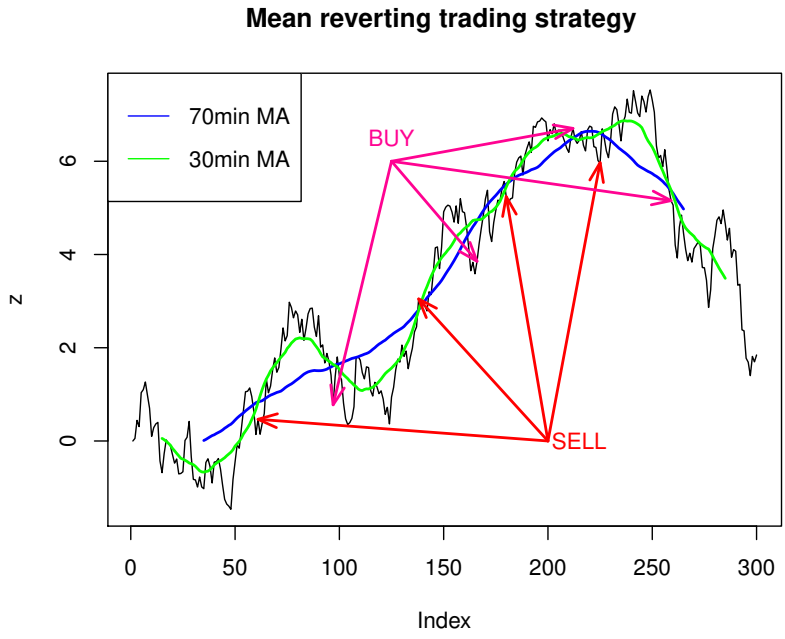

Figure 3: Dummy example of profitable trading strategy.

No conditional distribution is defined for the Hurst exponent. Thus, a standard statistical test for comparing critical values is out of the question. In practice, different authors often consider values outside the $0.45 - 0.65$ range to be significant [2]. For more accurate results, the following confidence intervals are in [24] obtained using bootstrapping:

Table 1: Confidence intervals for R/S.

confidence level:	lower bound	upper bound
90%	$0.5 - \exp(-7.35 \log(\log M) + 4.06)$	$0.5 + \exp(-7.07 \log(\log M) + 3,75)$
95%	$0.5 - \exp(-7.33 \log(\log M) + 4.21)$	$0.5 + \exp(-7.20 \log(\log M) + 4,04)$
99%	$0.5 - \exp(-7.19 \log(\log M) + 4.34)$	$0.5 + \exp(-7.51 \log(\log M) + 4.58)$

Table 2: Confidence intervals for DFA.

confidence level:	lower bound	upper bound
90%	$0.5 - \exp(-2.99 \log(M) + 4.45)$	$0.5 + \exp(-3.09 \log(M) + 4.57)$
95%	$0.5 - \exp(-2.93 \log(M) + 4.45)$	$0.5 + \exp(-3.10 \log(M) + 4.77)$
99%	$0.5 - \exp(-2.67 \log(M) + 4.06)$	$0.5 + \exp(-3.19 \log(M) + 5.28)$

Where $M = \log_2(n)$ and n is a lenght of subseries. For both cases we need to consider only subseries greater than 50 [24].

3.7 GARCH model

Bitcoin's behavior on the financial market has undergone several changes since its inception in 2009. The popularity of bitcoin has been further promoted by media attention. Many cryptocurrency exchanges have been founded, and the access of ordinary people to Bitcoin has become easier.

From the long-term point of view, until 2017, the price was rising. Constantly growing number of Bitcoins in circulation together with the decentralization and independence of Bitcoin prices from the real underlying asset resulted in a generally higher Bitcoins volatility. As most financial assets, bitcoin returns are known to exhibit volatility clustering and heteroskedasticity [16]. The use of high-frequency data enhances these properties even more. We can assume that there is an equilibrium level of volatility to which the short-term increased or decreased volatility converges (see Figure 4).

In an efficient market the growth in volatility is associated with the increase in risk and investors should naturally move towards less risky assets.

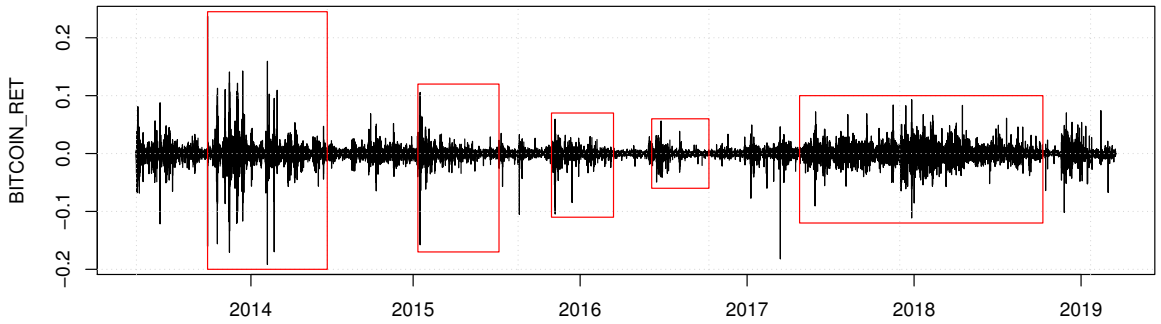

Figure 4: Volatility clustering on Bitcoin 30min. returns.

The features mentioned above can be taken into account using Generalized Autoregressive Conditional Heteroskedasticity model GARCH(p,q) first defined in 1986 in [25] as:

$$y_t = x_t' b + \epsilon_t \tag{3.23}$$

$$\epsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \tag{3.24}$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \tag{3.25}$$

where ω, $\alpha_1, \ldots, \alpha_q$ a $\beta_1, \ldots, \beta_p$, which to ensure a positive dispersion must satisfy:

$$\omega, \alpha_p, \beta_q > 0, \alpha_1, \ldots, \alpha_{p-1}, \beta_1, \ldots, \beta_{q-1} \geq 0 \tag{3.26}$$

to meet stationarity:

$$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1 \tag{3.27}$$

Garch models use the equation (3.23) to model conditional mean and equation (3.25) to model conditional variance. According to [27], GARCH (1,1) is in the most cases sufficient i.e. higher orders are not necessary. The effect of a return shock on current volatility in GARCH(1,1) decreases geometrically over time.

The main benefit of the GARCH model is its ability to allow for time-varying volatility. The disadvantage is in the assumption that positive and negative news have the same effect on volatility. The positive news is considered to be $\epsilon_{t-i} > 0$ (unexpected price increase). Conversely $\epsilon_{t-i} < 0$ (price drop unpredicted by the model) suggests the arrival of bad news on the market. Modeling this kind of asymmetry can be especially interesting for Bitcoin. Sentiment and uncertainty have a great impact on Bitcoin as it is not clear what its true value is. Therefore, working with a model where we use a symmetric constant for past ϵ_t might be inappropriate.

The high value of $|epsilon_t|$ implies the magnitude of these messages in the sense that they produce a tremendous unexpected change.

3.7.1 EGARCH

E-GARCH model published in 1991 by Nelson [29] assumes that an unpredicted drop in price (lousy news) increases predictable volatility more than a similarly substantial, unpredicted increase in price (good news).

EGARCH has the form:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \beta \log(\sigma_{t-1}^2) + \sum_{i=1}^{p} \alpha_i \frac{|\hat{\epsilon_{t-i}}| + \Upsilon \hat{\epsilon_{t-i}}}{\sigma_{t-i}} \tag{3.28}$$

where ω, $\alpha_1, \ldots, \alpha_q$ a $\beta_1, \ldots, \beta_p$ are positive constants parameters, ϵ_{t-i} are residuals and Υ is leverage coefficient. Negative information generally has a worse effect, producing higher volatility than positive effects, so we can assume that the coefficient Υ for the level of asymmetry $\frac{\hat{\epsilon_{t-i}}}{\sigma_{t-i}}$ will be negative. The great value of $|\hat{\epsilon_{t-i}}|$ implies the magnitude of these messages in the sense that they produce a great unexpected change.

3.7.2 News Impact Curve

GARCH(1,1) model and EGARCH(1,1) model could be compared by looking on the effect of the news on the conditional heteroskedasticity. By plotting news impact curve relation of past returns (shocks / news) e_{t-i} and current volatility σ^2 we can see how these information incorporate into volatility estimates for these models [28]. An example of such a curve is given in Figure 5.

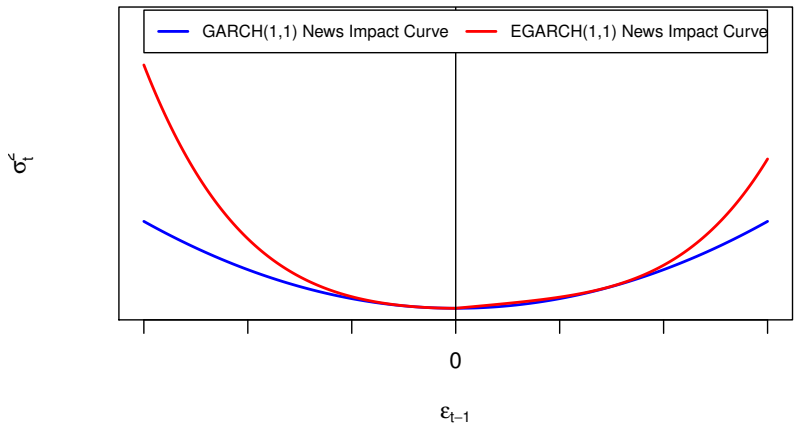

Figure 5: News impact curve for GARCH(1,1) and EGARCH(1,1)

GARCH model differs from the EGARCH model in two main aspects. The GARCH model does not allow good news and bad news to have a different impact on volatility. The GARCH model does not allow large shocks to have a considerably larger impact on volatility.

3.7.3 Sign Bias Test, Negative Size Bias Test, Positive Size Bias Test

In practice it is useful to test whether the model, we are working with, is adequate. To make sure that there is no underestimation of the volatility not predicted by the model after bad news and there is no overestimate of the volatility not predicted by the model after good news, it is recommended to employ Sign Bias Test [28].

Negative Size Bias Test [28] is used to compare the effect of small and large negative return shocks on volatility not predicted by the model. Positive Size Bias Test is the

same for positive shocks. These tests work with the hypothesis:

H0: The model for a given effect on volatility cannot be improved by adding historical variables with which the model is not currently working.

Rejecting this hypothesis suggests that the original model might be misspecified. These tests are in R-enviroment part of ru-Garch package.

4 Results

In the following section we investigate the Bitcoin and currencies as such. The logs returns are positive with excess kurtosis and negative skewness. The return series are not normally distributed and display fat-failed behavior.

4.1 Descriptive statistics for minute returns

Bitcoin/USD and CHF/USD have the most considerable returns. The Swiss franc has been weakening against the US dollar around seven months since mid-2014. This period was followed by an extremely sharp increase in value and a decline again. Higher values of descriptive statistics for CHF (especially kurtosis) are the result of this period.

Statistics	Mean	Std.Dev	Median	Min	Max	Skewness	Kurtosis
BTC/USD	$1,96 * 10^{-6}$	0.002	0	-0.11	0.12	0.07	88.2
AUD/USD	$-6.87 * 10^{-9}$	0.0002	0	-0.016	0.011	-0.02	185.2
CAD/USD	$6.40 * 10^{-9}$	0.0001	0	-0.013	0.016	0.21	311.8
EUR/USD	$6.61 * 10^{-9}$	0.0001	0	-0.014	0.017	-0.06	405.3
GBP/USD	$5.61 * 10^{-9}$	0.0001	0	-0.032	0.013	-8.13	2366.3
CHF/USD	$8.82 * 10^{-9}$	0.0001	0	-0.051	0.059	17.66	22000.7
JPY/USD	$-1.85 * 10^{-8}$	0.0002	0	-0.013	0.02	3.74	926.6
NZD/USD	$1.4 * 10^{-8}$	0.0001	0	-0.05	0.022	-6.77	2326.3

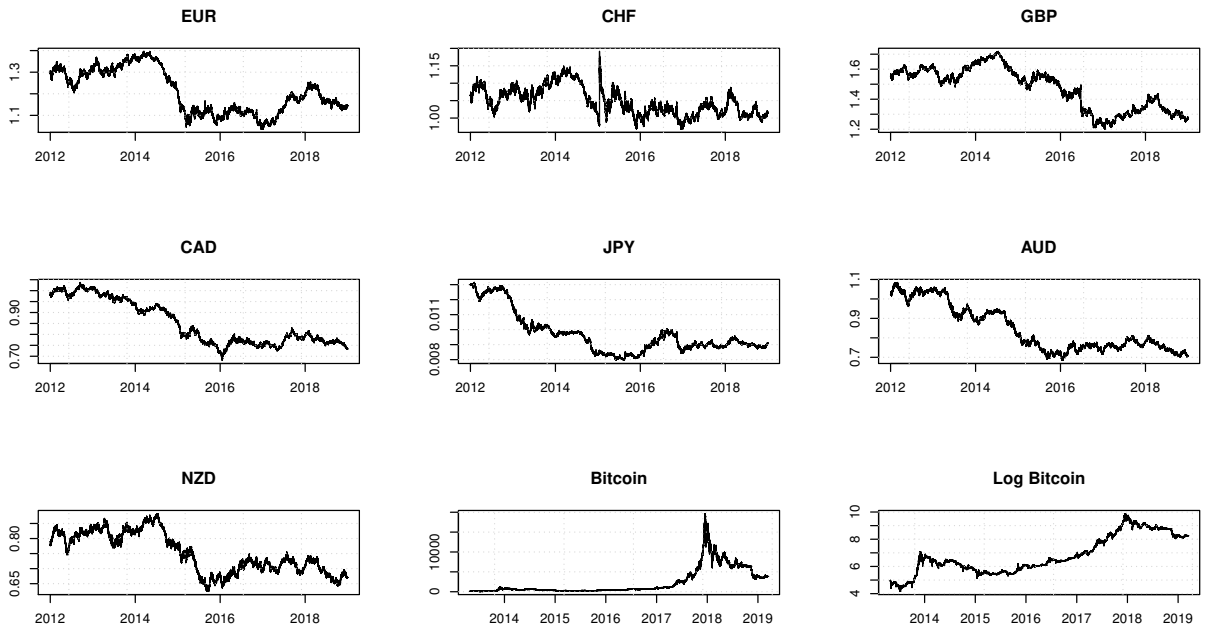

Figure 6: Correlation size and significance for returns.

4.2 Correlation between Bitcoin and classic currency exchange rates

Correlation of Bitcoin's returns with other market assets is generally small somewhere between -0.1 and 0.1 in local periods. This is something we expected because if the correlation was higher, it would be possible to replicate the Bitcoin portfolio with less risky assets.

As we can see in Figure 7, the exchange rates correlate more significantly among each other. Color-filled fields indicate a correlation which is by cor.test evaluated as significant (at a 10% significance level). Correlations are calculated on 6-hour data on the entire data sample from 2013 to 2019. EUR/USD correlates negatively with all other currencies except Bitcoin.

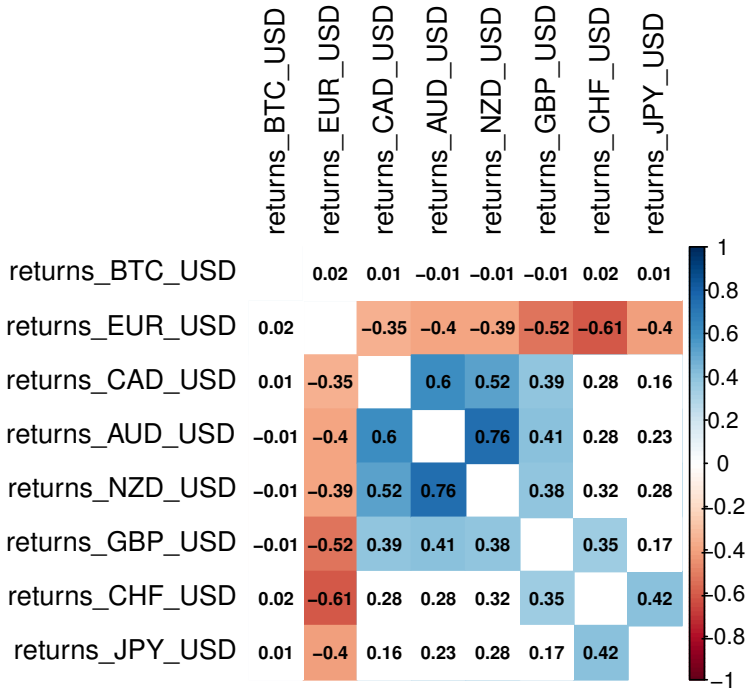

Figure 7: Correlation size and significance for returns.

4.2.1 Moving correlation

In the following charts we present a moving correlation over time using overlapping two months period and a two-days sliding window. The correlation was estimated on the data with a 12-hour frequency. The rolling correlations estimated on the one-minute data frequency can be found in AppendixA. The results, for minute data are approximately the same, but the speed of change in between positive and negative correlations is faster. The correlation ranges from -0.02 to 0.02 for minute data.

The correlation gradually changed over time and there are periods where it is non-zero. From the middle to the end of 2014 there is an area in which the correlation is significantly non-zero for all exchange rates. Generally, in some shorter periods, the price of Bitcoin against the US dollar is correlated with the other currency exchange rates. Therefore we try to capture these impacts in our first linear model.

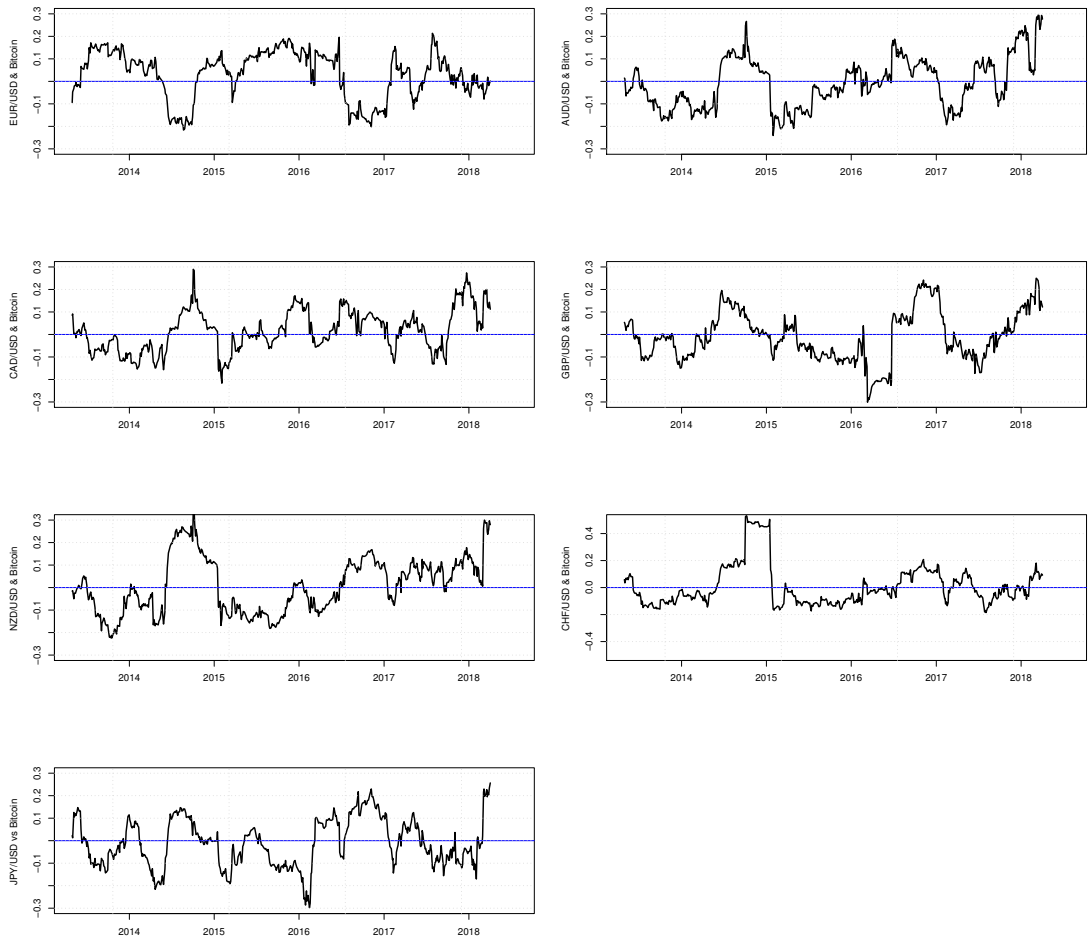

Figure 8: Floting correlation in between the currency exchange rates and Bitcoin.

4.3 1. Linear model for Bitcoin returns on exchange rate returns

Our first linear was model calibrated on the minute data:

$$r_i^{BTC/USD} = \beta_0 + \beta_1 r_i^{EUR/USD} + \beta_2 r_i^{NZD/USD} + \beta_3 r_i^{CAD/USD} + \beta_4 r_i^{AUD/USD} + \dots \quad (4.1)$$

$$\cdots + \beta_5 r_i^{GBP/USD} + \beta_6 r_i^{JPY/USD} + \beta_7 r_i^{CHF/USD} + \beta_8 r_i^{VolumeBTC} + \epsilon_i$$

The model is due to data gaps and fluctuating correlations (which are on the long time average equal to almost zero) calibrated only on the week period commencing on April 26, 2013. Overall we had 10071 observations. The reason, why we chose to use the calmer period of Bitcoin's earlier existence is, that the price (in that time = 112 US dollars) fluctuated less than later in the times of extreme increases since 2017. This basic model thus had a better chance of succeeding. Significant for 5% significance level were 4 parameters. With sequential calibration of the model over the next weeks we can see significant parameters alternating. They are not all significant at one moment due to high correlation between them. There are also times when Bitcoin return does not correlate with currency exchange rates. The model is sensitive to calibration for single minute returns and does not work well for periods longer than two weeks due to changing dependencies. For 6-hour data we can go up to $1 - 3$ months. This result broadly agrees with what the authors came up with in [16]. They report using Bayes' changepoint concept on average one change in distribution from which Bitcoin returns come from every 1.5 months.

The R-squared of this model is 0.027, which means, that the model captured a total of 2.7% of the total data variability, which is not a good result. On the other hand, those parameter estimates β, that came out significant, are relatively high considering using high-frequency minute data.

```
Call:
lm(formula = BITCOIN_RET[a:b] ~ EUR_USD_RET[a:b] + NZD_USD_RET[a:b] +
    CAD_USD_RET[a:b] + AUD_USD_RET[a:b] + GBP_USD_RET[a:b] +
    JPY_USD_RET[a:b] + CHF_USD_RET[a:b] + BTC_VOLUME_MA[a:b])

Residuals:
      Min        1Q    Median        3Q       Max
-0.084435 -0.001516 -0.000036  0.001539  0.223766

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -2.381e-05  6.288e-05  -0.379   0.7050
EUR_USD_RET[a:b]   -1.777e+00  7.435e-01  -2.391   0.0168 *
NZD_USD_RET[a:b]   -2.508e-01  3.887e-01  -0.645   0.5187
CAD_USD_RET[a:b]    5.329e-01  7.028e-01   0.758   0.4483
AUD_USD_RET[a:b]   -2.260e+00  5.318e-01  -4.250 2.16e-05 ***
GBP_USD_RET[a:b]   -4.139e-01  6.432e-01  -0.644   0.5198
JPY_USD_RET[a:b]   -1.390e+00  3.484e-01  -3.991 6.64e-05 ***
CHF_USD_RET[a:b]   -4.297e-01  7.367e-01  -0.583   0.5597
BTC_VOLUME_MA[a:b] -5.184e-04  3.400e-05 -15.249  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.006312 on 10071 degrees of freedom
Multiple R-squared:  0.02746,Adjusted R-squared:  0.02669
F-statistic: 35.55 on 8 and 10071 DF,  p-value: < 2.2e-16
```

The estimated autocorrelation function (acf, see Figure 9) of the residuals from model (4.1) signals a negative correlation with the lag 1 residual variable ϵ_{t-1}. Correlation with returns from previous minutes can be seen on the estimated partial autocorrelation function (pacf, see Figure 9). P-values sometimes exceed rejection value 0.05 even for higher lags. We think, it might be a sign of long-term dependence, which could be later revealed by Hurst's exponent.

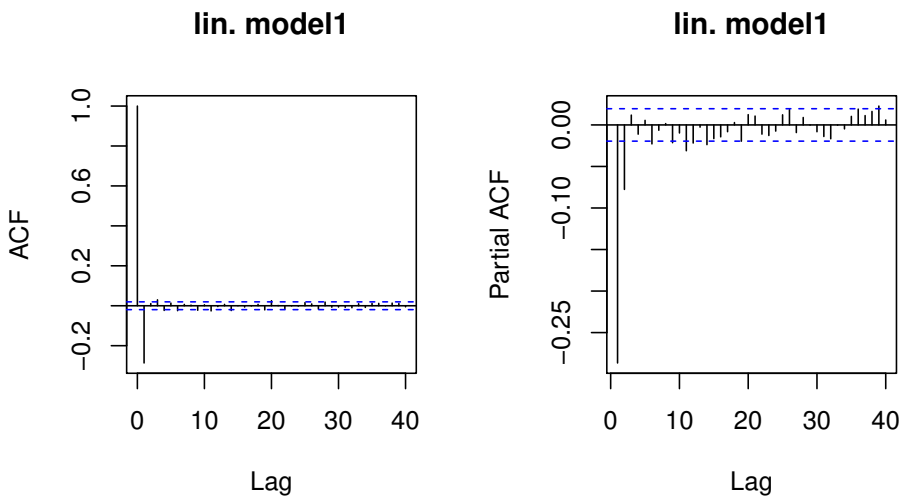

Figure 9: Acf and Pacf of residues from Model 4.1.

4.4 2. Linear model of Bitcoin returns on exchange rate returns

The second model is based on the first model 4.1 with added one explanatory variable the MA(1) variable (the value of the previous $i-1$ residue ϵ_{i-1}). This variable helps to explain pieces of information which is likely to affect the Bitcoin's price, but did not affect prices for currencies such as negative reports about Bitcoin's future or other financial indicators, that some of the noise traders follow. It might suggest, that Bitcoin is on minute data ineffective. This observation has led us to examine more closely the effectiveness of Bitcoin.

Observing the estimated pacf on Figure 10 from Model 4.1 may make sense to add also the explanatory variable AR(1) (the return from previous $i-1$ minute r_{i-1}). The addition of AR (1) or MA (1) variable among explanatory variables of this model seems to be in terms of its improvement and reduction of residue dependence equivalent. Adding both these variables makes little sense as R-squared improves only marginally - by 0.1%. The estimated pacf and acf are approximately the same for these alternatives. After all we managed to reduce the dependence shown by Acf. The estimated pacf does not show significant improvement.

The main advantages of adding a new variable are the R-square increase to almost 11% and making the residuals from model independent. Runs test for them do not reject the hypothesis of independence on 5% significance level.

```
Call:
lm(formula = BITCOIN_RET.a.b. ~ EUR_USD_RET.a.b. + NZD_USD_RET.a.b. +
    CAD_USD_RET.a.b. + AUD_USD_RET.a.b. + GBP_USD_RET.a.b. +
    JPY_USD_RET.a.b. + CHF_USD_RET.a.b. + BTC_VOLUME_MA.a.b. +
    lag1MA, data = regresion_data_2)

Residuals:
      Min        1Q    Median        3Q       Max
-0.084581 -0.001845  0.000066  0.001974  0.223413

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -2.289e-05  6.024e-05  -0.380 0.703942
EUR_USD_RET.a.b.   -1.580e+00  7.123e-01  -2.218 0.026561 *
NZD_USD_RET.a.b.   -9.115e-03  3.724e-01  -0.024 0.980474
CAD_USD_RET.a.b.    5.723e-01  6.733e-01   0.850 0.395344
AUD_USD_RET.a.b.   -2.418e+00  5.095e-01  -4.746  2.1e-06 ***
GBP_USD_RET.a.b.   -1.894e-01  6.162e-01  -0.307 0.758527
JPY_USD_RET.a.b.   -1.270e+00  3.339e-01  -3.805 0.000143 ***
CHF_USD_RET.a.b.   -3.884e-01  7.058e-01  -0.550 0.582117
BTC_VOLUME_MA.a.b. -5.019e-04  3.257e-05 -15.408  < 2e-16 ***
lag1MA             -2.870e-01  9.552e-03 -30.042  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.006047 on 10069 degrees of freedom
Multiple R-squared:  0.1074,Adjusted R-squared:  0.1067
F-statistic: 134.7 on 9 and 10069 DF,  p-value: < 2.2e-16
```

Sign of dependence shown on the acf and pacf from the second model (see Figure 10) may lead to the need for improvement using a more complex model, such as the GARCH model.

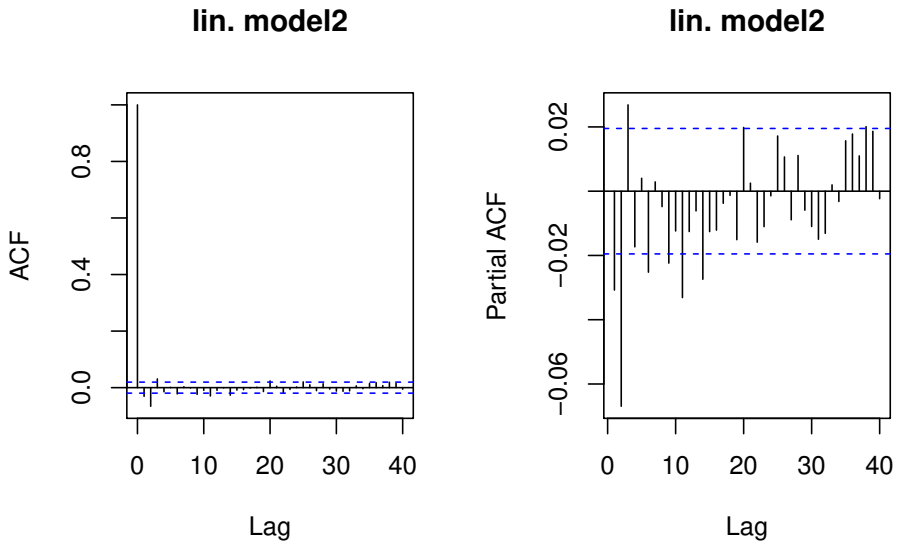

Figure 10: Acf and Pacf of residues from 2. linear model.

Comparison between fits of these models are displayed in Figure 11. Model two is remarkably better in terms of fitting in more volatile subperiods.

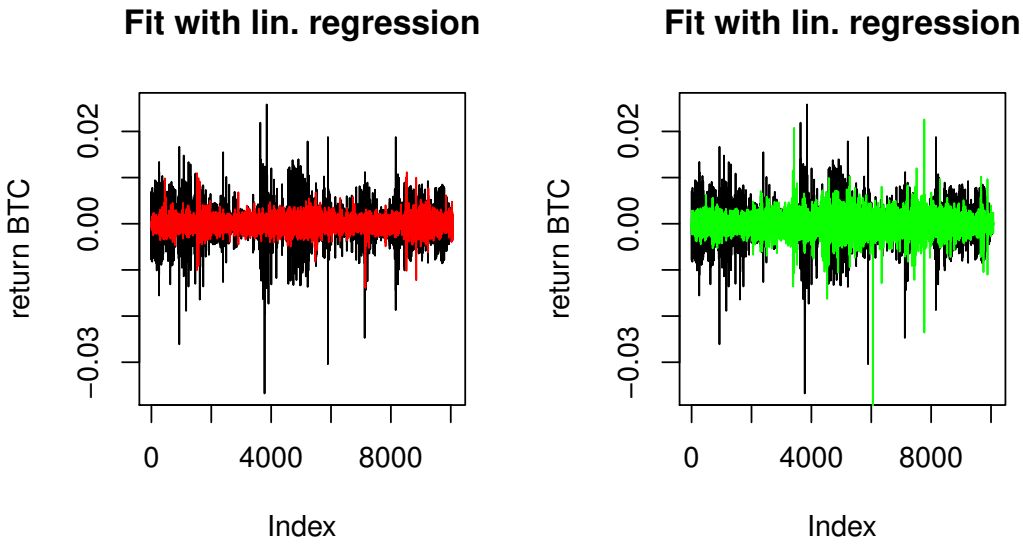

Figure 11: Comparison of fit in between 1. model and 2. model.

4.4.1 Evolution of regression coefficients throughout the period

The regression coefficients are non-constant over time in previous models. Using 6-hour data we will try to summarize the evolution of Bitcoin/USD dependence from the currency exchange rates over time. The six-hour data seems to be the most appropriate alternative for this kind of analysis. They are still high frequency, but at the same time, they allow the model to be configured for a longer period of time, contain less noise and thus make the results more credible.

Currencies returns are strongly correlated, therefore we need to simplify our original model (4.1) to reveal the real effect, that every single currency has on the Bitcoin returns:

$$r_i^{BTC/USD} = \beta_0 + \beta_1 r_i^{currency/USD} + \epsilon_i \tag{4.2}$$

We monitored the significance and evolution of the individual coefficients. To achieve more robust results we configured the model on the intervals that have been stretched in the different ways from 1 week to 3 months. The only credible parameter was the one whose result was consistent across different intervals. Timeline results are shown in the following Figure (12):

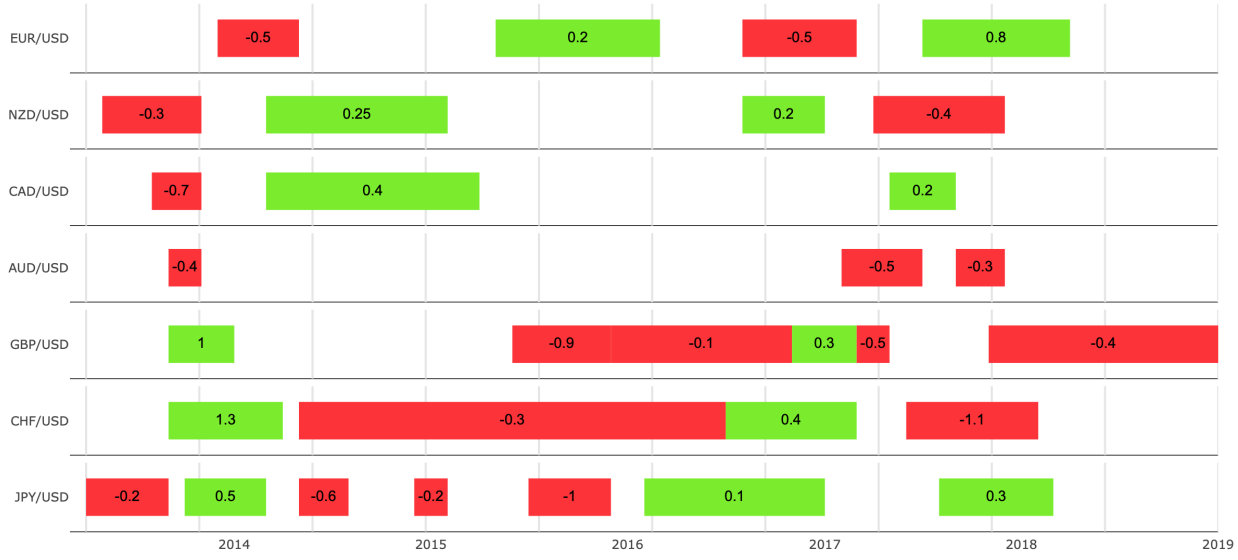

Figure 12: Evolution of regression coefficients of model (4.2) over time.

The period of higher correlation between the Bitcoin returns and the currency returns at the turn of 2013 and 2014 (see Figure 8) is in agreement with the significance of the parameters from the simplified model (4.2) over this time.

The most persistent relationship in our analysis seems to be between Bitcoin and GBP, CHF and JPY. The sign of dependence has changed few times, but their pattern has been almost always present in price evolution of Bitcoin.

Explanatory variables with the positive sign may indicate a weakening or strengthening of the US dollar, which has transferred to the currency exchange rates and Bitcoin. For example, if the dollar weakened and the currencies and the Bitcoin grew then returns will be positive with positive relation between them.

The long-term negative dependence between Bitcoin and Swiss Frank at the time of the fall of the Swiss franc could be caused by the transfer of some capital from Frank (considered to be safe storage - often referred as "safe heaven") into Bitcoin, resulting in price increases.

Countries in Asia have emerged as leaders in the usage of cryptocurrencies. The largest daily turnover of Bitcoin is in Japan yen. Therefore decline/increase in the fundamental value of JPY may affect the Bitcoin price more than other currencies. JPY was another currency with which Bitcoin's dependence is relatively consistent in our analyses.

If we should propose some trading strategy based on this analysis, not necessarily profitable, it would be buying/selling in the opposite direction to CHF or GBP.

4.5 AR (1) model depending on data frequency

Loking at the predictability of the stock market prices (returns), unless they follow a random walk (which is equivalent to returns driven by the white noise), considering the market without transaction costs and with infinite liquidity - at any time it is possible to buy/sell any quantity for a given price there might be a strategy making money for investors. The question is whether there is a speed difference between the Bitcoin market and the conventional currency market in responses to a change in a situation or new information arriving. On an inefficient market the new information affects the price at multiple consecutive intervals. We now undertake an empirical investigation about how well can the AR(1) model predict the Bitcoin and currency exchange rates returns for various frequencies. The AR(1) model will model both Bitcoin and currency exchange rates while gradually reducing the data frequency. The results are shown in the following figures:

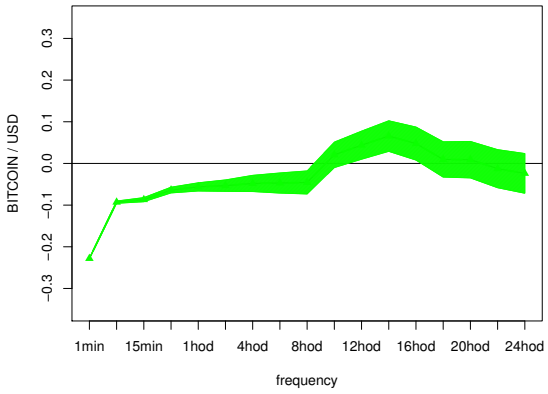

Figure 13: Frequency dependent AR coefficient for the Bitcoin.

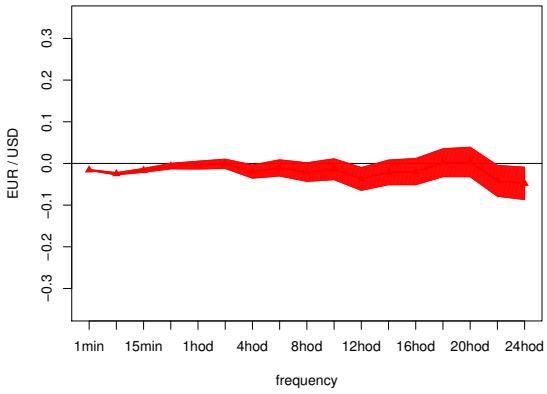

Figure 14: Frequency dependent AR coefficient for the Euro.

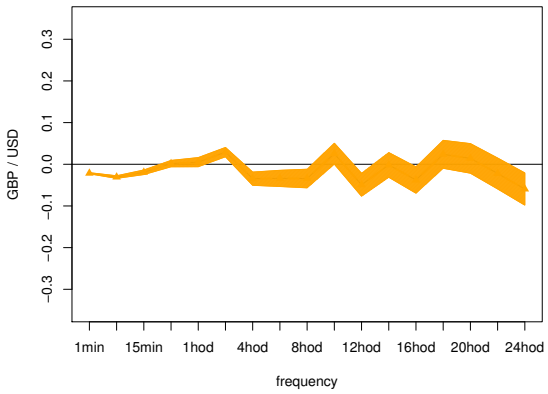

Figure 15: Frequency dependent AR coefficient for The pound sterling (GBP).

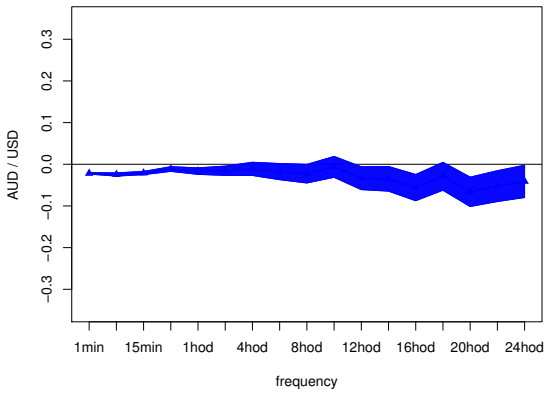

Figure 16: Frequency dependent AR coefficient for the Australian dollar

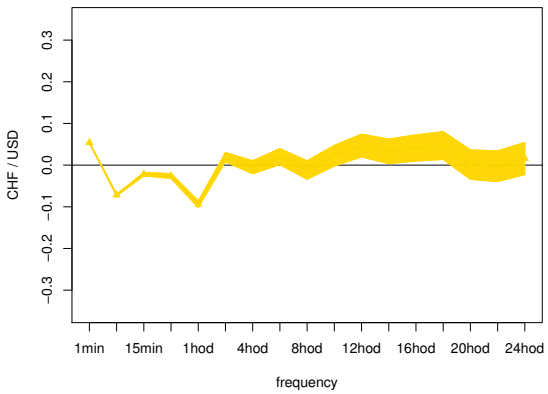

Figure 17: Frequency dependent AR coefficient for the franc.

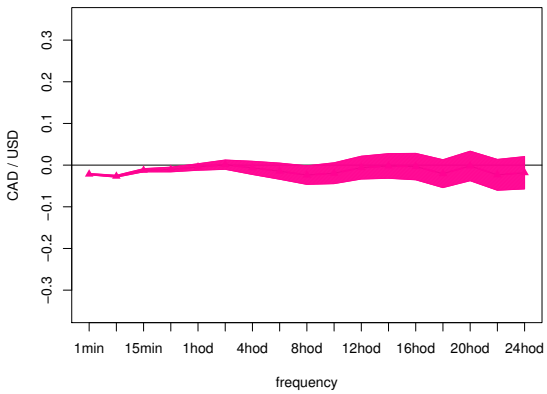

Figure 18: Frequency dependent AR coefficient for the Canadian dollar.

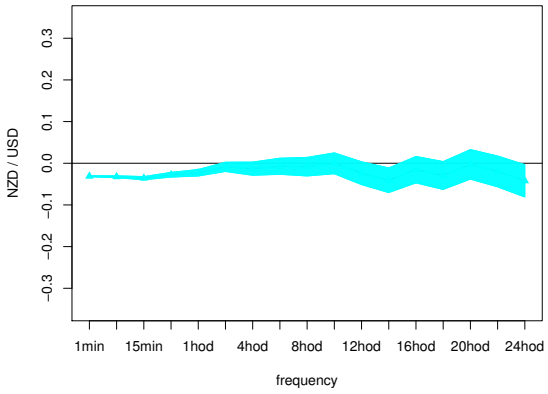

Figure 19: Frequency dependent AR coefficient for the New Zealand dollar.

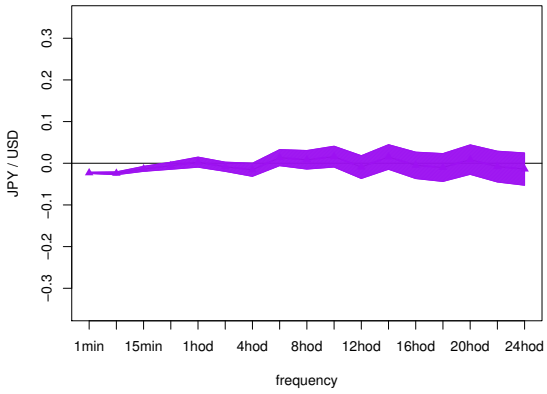

Figure 20: Frequency dependent AR coefficient for the yen (JPY).

It turns out, that currency exchange rates tend to be according to AR(1) model already relatively effective even at minute data. The AR(1) coefficient together with a confidence interval varies for them around zero, while for Bitcoin the estimated coefficient for the minute data is significantly non-zero (-0.22). It can be seen as something contrary to the efficient market hypothesis. For minute data this dependency may not be associated with the ability to predict the market. Inefficiency may be related to a low in liquidity in the market. The Bitcoin market does not have a liquidity provider and bid-ask spread grows when liquidity leaves. The bid-ask spread perhaps causes a higher negative value of AR (1) on the minute data. If multiple requests of the same type arrive at the market in one minute, the realizable price is biased in one direction. After this offset the opposite requests will probably come on the market and the price will go in the opposite direction. However, with the increasing time span, the bid-ask has no longer such a big impact on the realizable price. Bitcoin inefficiency from 5-minute data to 8-hour data is something, that is not a common phenomenon for the conventional currency market.

Bitcoin becomes more and more efficient with the decreasing frequency. At 12-hour data, AR (1) is no longer needed for modeling BTC returns.

An estimation of previous coefficients was made on the full data sample from 2012 to 2019. We will now focus on how dependence on one minute data evolved over time.

4.6 AR(1) model depending on time

The ability to predict Bitcoin returns using the AR(1) model appears to be related to the efficiency and liquidity. It seems meaningful to investigate the evolution of the estimated AR(1) coefficient, because the amount of liquidity on the market was not constant over time. We worked with one-minute data frequency. Minute data in previous analyzes points to the most considerable inefficiency of Bitcoin. The model was gradually configured at half-year subintervals. The results together with confidence intervals are plotted in the following figures. The confidence intervals were small, so the estimated AR (1) coefficients are credible.

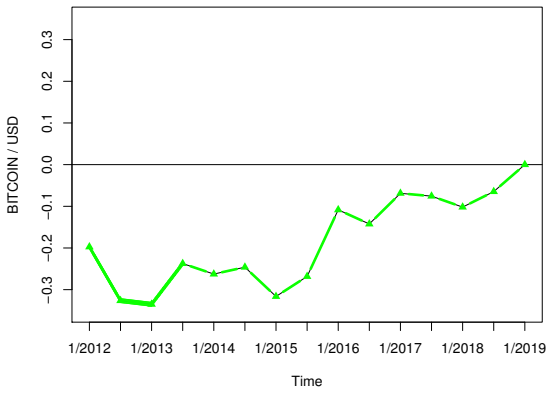

Figure 21: Evolution of AR(1) coeff. for Bitcoin in time.

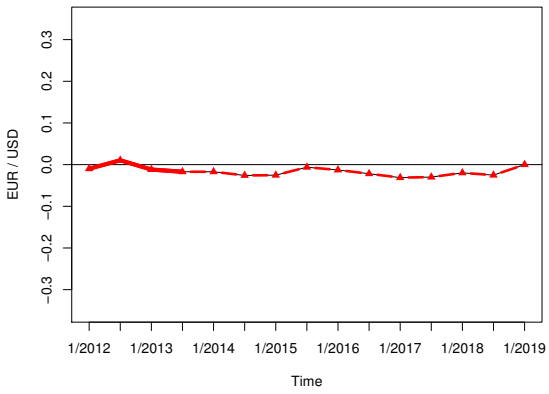

Figure 22: Evolution of AR(1) coeff. for the Euro in time.

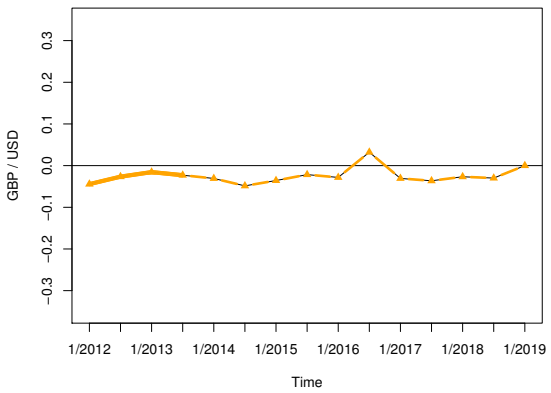

Figure 23: Evolution of AR(1) coeff. for the pound sterling (GBP) in time.

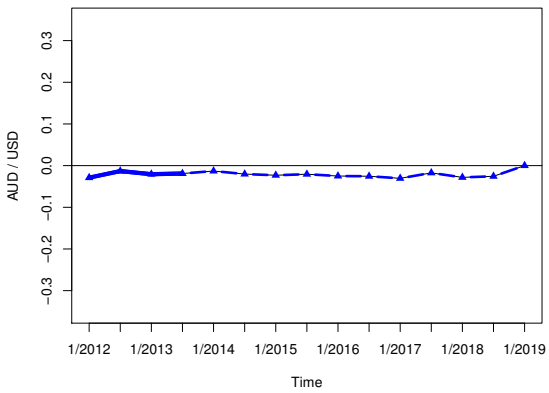

Figure 24: Evolution of AR(1) coeff. for the Australian dollar in time.

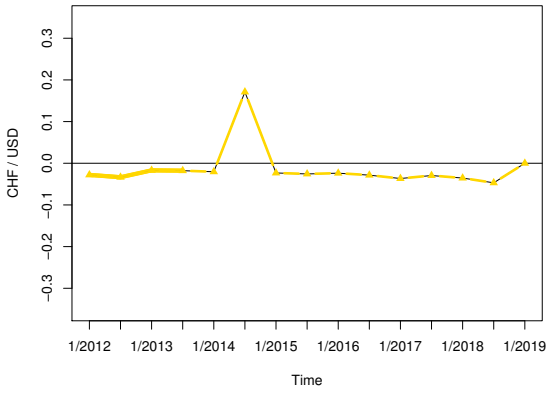

Figure 25: Evolution of AR(1) coeff. for the franc in time.

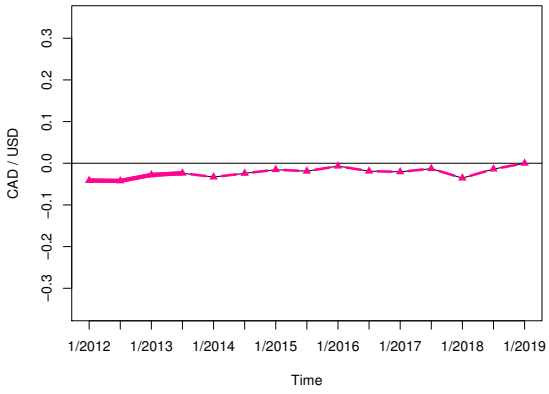

Figure 26: Evolution of AR(1) coeff. for the Canadian dollar in time.

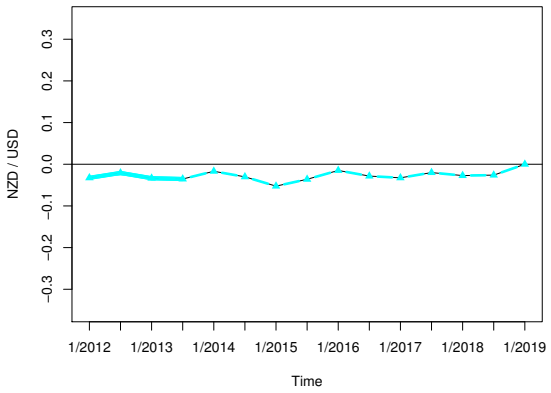

Figure 27: Evolution of AR(1) coeff. for the New Zealand dollar in time.

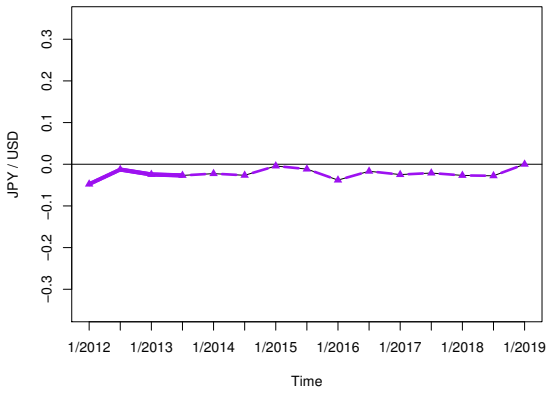

Figure 28: Evolution of AR(1) coeff. for the yen (JPY) in time.

The inefficiency of Bitcoin is in the beginning relatively strong, but over time gradually disappears. At the beginning of 2019 Bitcoin seems to be as effective as currency exchange rates on minute data. We expected this as new investors together with liquidity come to the market. On the contrary, we expected, that during the period of steep increases in the price, that Bitcoin experienced at the end of 2013 and throughout 2017, the value of AR (1) would approach zero (efficiency increases), but it did not. Effectiveness was rather constant in the times of the high growth. The currency exchange rates experience constant efficiency - low predictability using the AR (1) model. The only exception is the CHF, which grew sharply from 2014 to 2015 and than fell.

4.7 Efficiency testing

In the past authors have done a great deal of the effectiveness testing for Bitcoin using different approaches. As far as we know, nobody jet utilized the concept of high-frequency data in their analyses. The AR(1) coefficient is just one possible approach to the efficiency monitoring. So we decided to deploy more powerful statistical tests for testing the weak form efficiency of Bitcoin. We will try to capture the dynamics of Bitcoin for various frequencies and display the evolution in time.

Each of the tests we use works with different test statistics and offer another perspective on the data dependency. For the longer-term models, it is recommended to use the more sophisticated tests to examine whether the correlation in data is due to chance. We used Ljung-Box test with the null hypothesis of no autocorrelation. The Runs test determines whether the sequences of signs in returns are the result of randomness. The Bartels test in a nonparametric manner checks if a serial correlation is present in returns. Automatic Variance Ratio searches for an optimal lag to detect the presence of dependence.

Bitcoin prices were relatively stable except a rise in late 2013 and before peaking dramatically in 2017. We examine the weak-form informational efficiency of Bitcoin over the full sample period. As well we split our sample into four subsamples (see Figure 37) in order to determine whether the level of efficiency has varied over time. We tried to divide the period into the parts so that the Bitcoin properties in the area were about the same:

1. From 2012 to mid-2014. In this period the price was relatively small. Than Bitcoin goes through one sharp increase in the price from 100 USD to 1000 USD. Afterwards there is a gradual decline.

2. From mid-2014 to the beginning of 2017. Bitcoin prices grew very slowly and generally, this period is not very volatile.

3. from the beginning to the end of 2017. During this period Bitcoin experienced the biggest boom in terms of growth as well as the supply of liquidity to the market.

4. from the beginning to the end of 2018. There is a period of decline in Bitcoin value. The market has dried out and liquidity is leaving.

In each case, we report the corresponding p-values on graphs.

4.7.1 Performing the Ljung Box test on returns.

The Ljung Box's test tests null hypothesis: *time series is without autocorrelation.* The Bitcoin returns as well as the currency exchange returns looks correlated for minute data, as frequency decreases pass into uncorrelation. This transition is significantly earlier (starts on data with 15minute frequency) for currencies. Bitcoin seems to be effective by means of Ljungbox test from 12hour data).

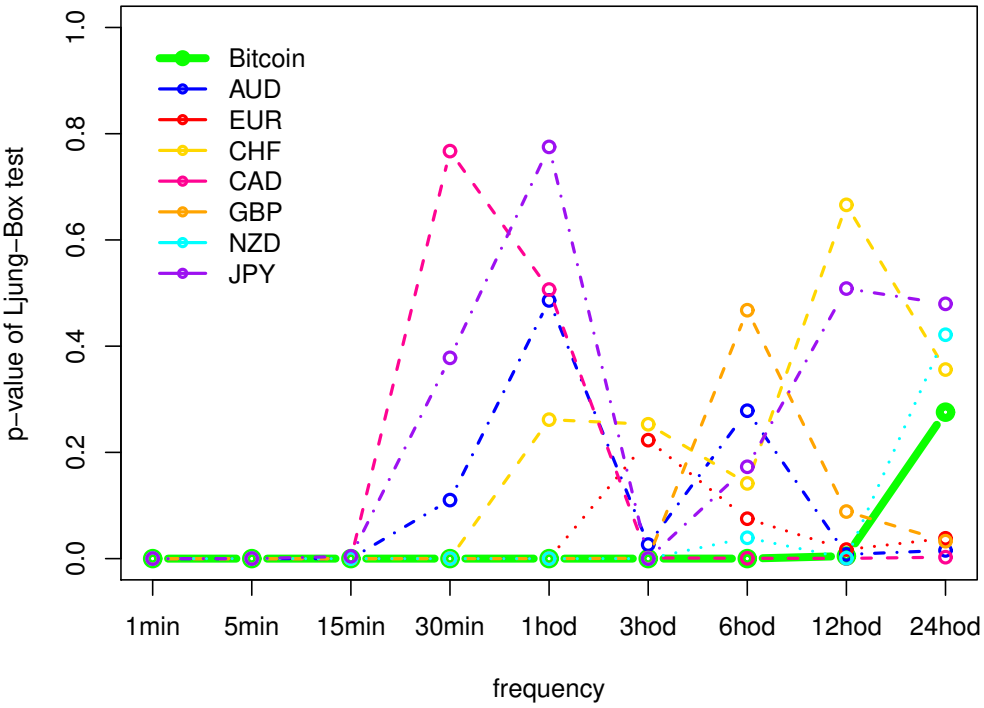

Figure 29: P-values of Ljung-Box test depending on data frequency.

The Figure 29 shows the p-values of the Ljung-Box test for the entire period.

On the next page we show the results for the four different periods. As can be seen, the efficiency tested by the Ljung-Box test for currencies did not changed so much. The effectivity of Bitcoin based on the rejection of the null hypothesis of no correlation by Ljung-Box test is similar in the first and third periods. Bitcoin starts showing signs of efficiency from 3 and 6 hour data. For the second and fourth periods it is later for 12-hour data.

4.7.2 Evolution of p-value of Ljungbox test over time

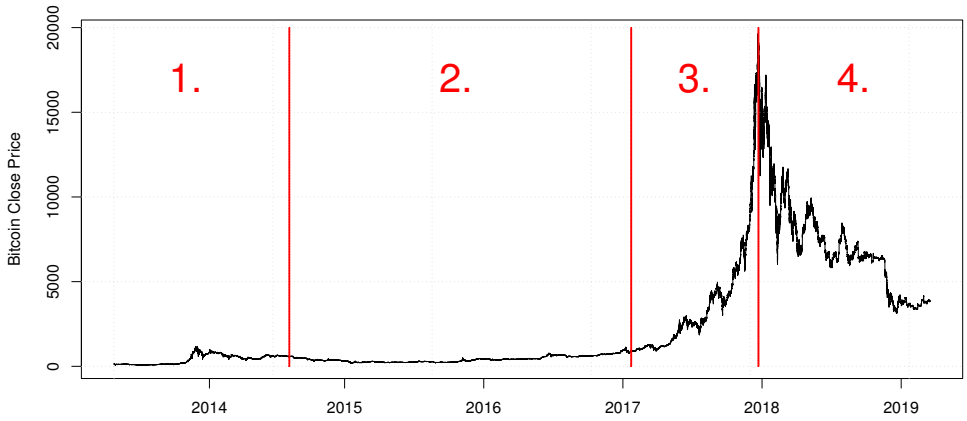

Figure 30: Division into subintervals.

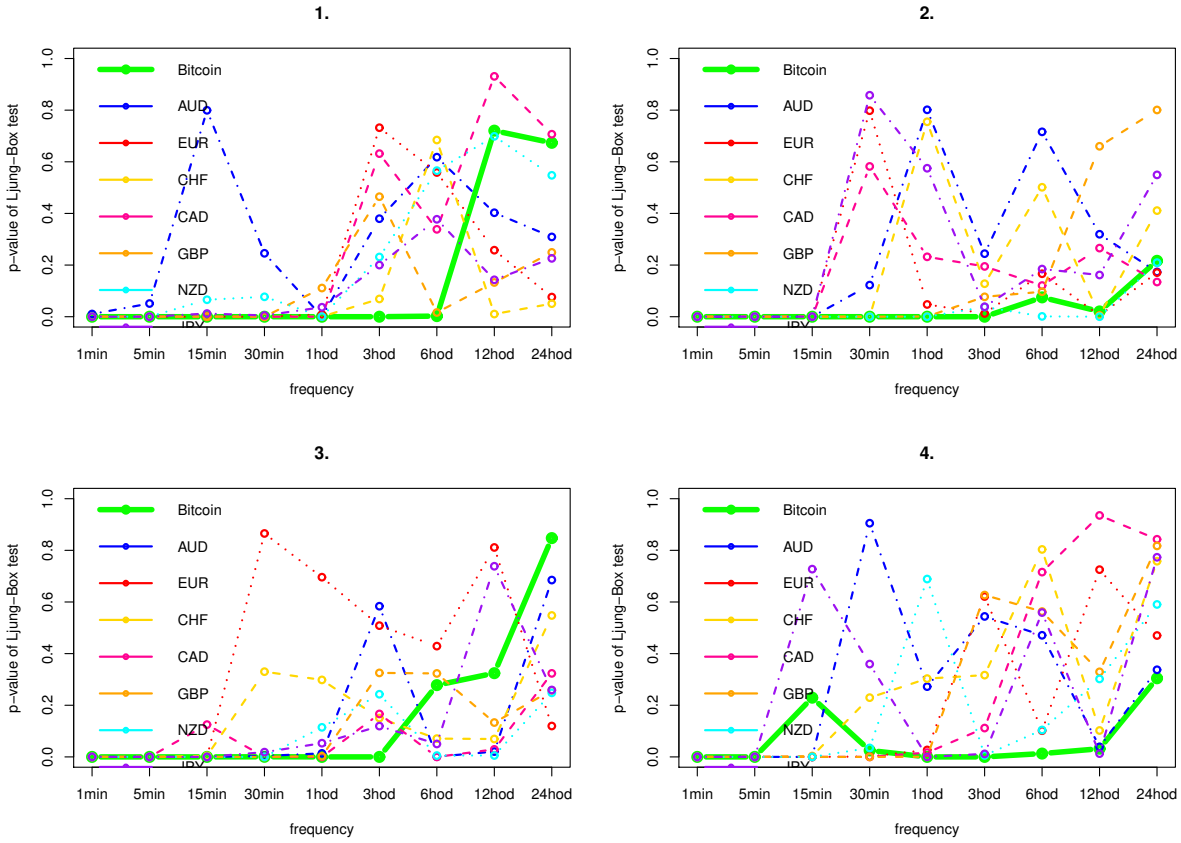

Figure 31: Division into subintervals.

4.7.3 Performing Runs test on returns.

Runs test was employed to determine whether returns are independent, by looking on the signs of returns. The null hypothesis for the Runs test is: *Each sample value from a sequence is independently drawn from the same distribution.*

Runs test seems to be simpler and more conservative than Ljung-Box test in rejecting the null hypothesis. Test does not reject null hypothesis indicating no autocorrelation on the full sample even for the currencies until 6hour data. After all Bitcoin still seems to be less effective than conventional currencies.

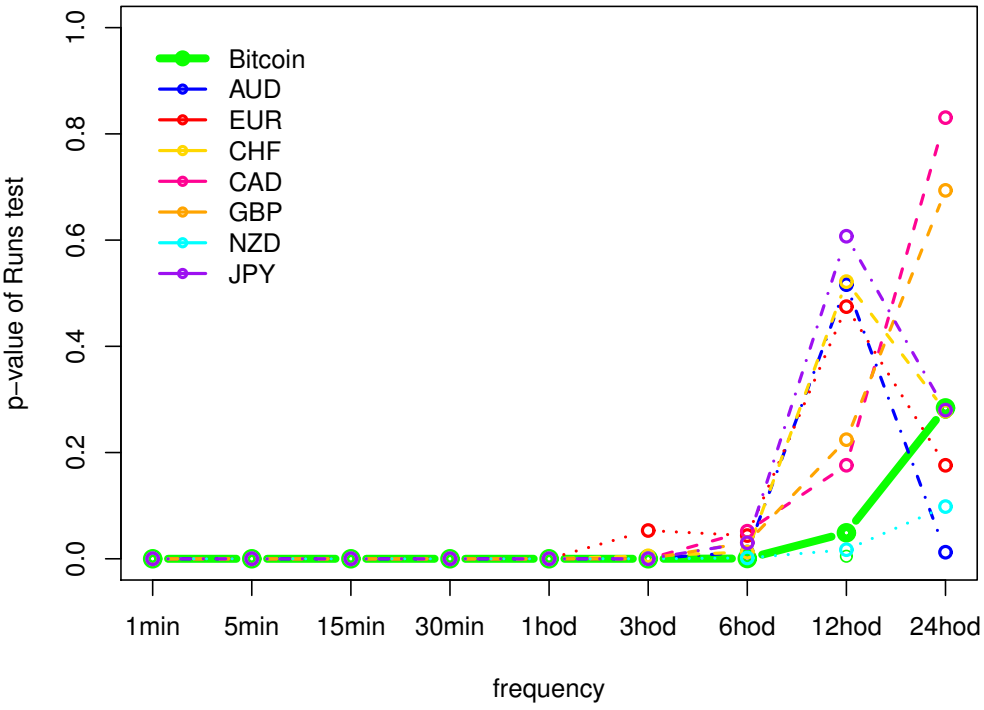

Figure 32: P-values of Runs test depending on data frequency.

As with the Ljungox test we can see, that the test results are in the first + the third subsamples and second + fourth subsamples approximately the same. The test on the first and third subsample period shows no evidence against the null hypothesis. It indicates no significant difference between Bitcoin and currencies in terms of effectivity. The second and fourth periods shows smaller effectivity for Bitcoin.

4.7.4 Evolution of p-value of Runs test over time

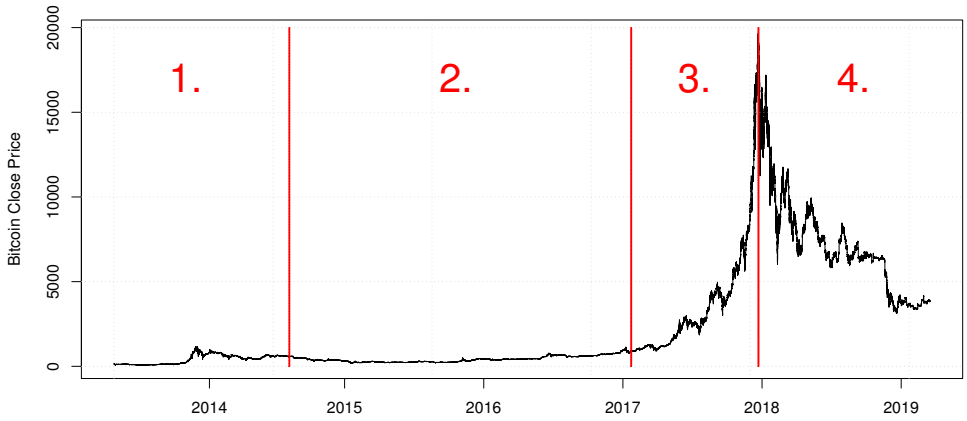

Figure 33: Division into subintervals.

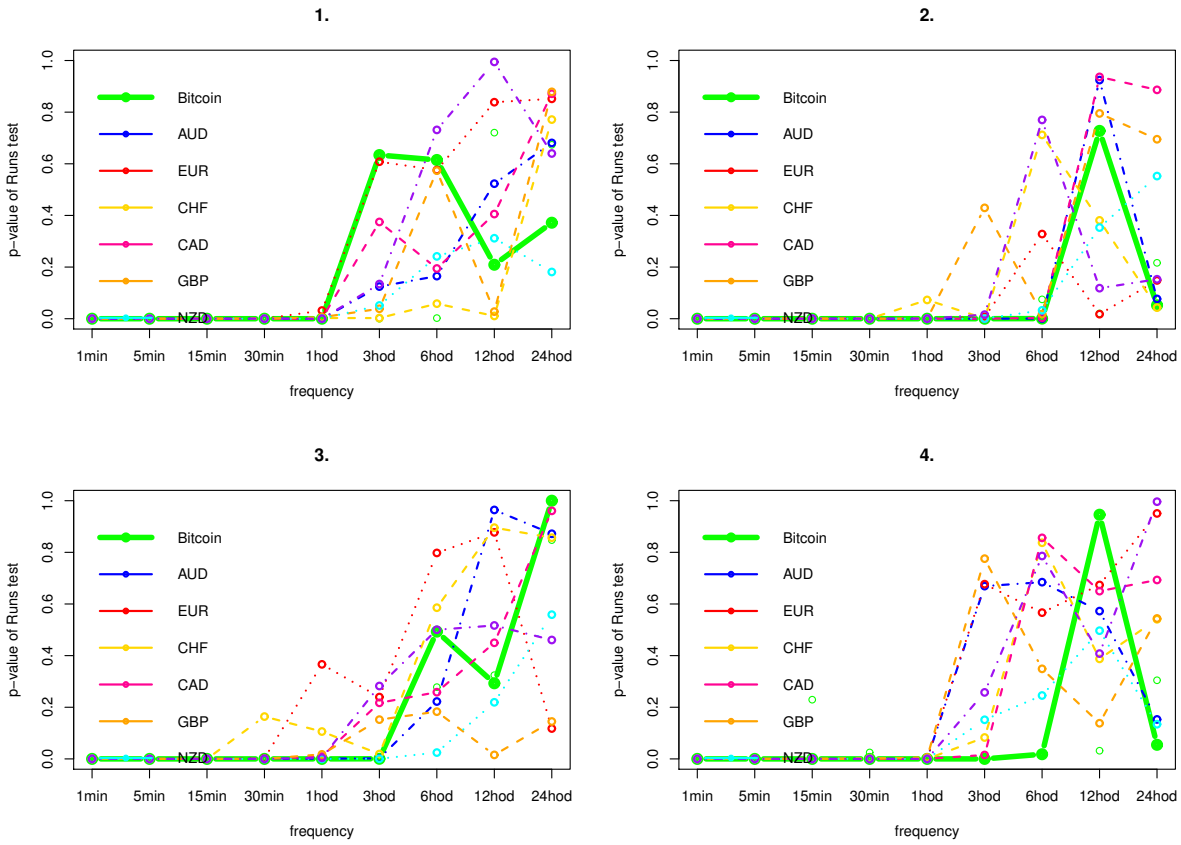

Figure 34: Division of Bitcoin into subintervals.

4.7.5 Performing the Bartels test on returns.

As with the Runs test the Bartels test is another non-parametric test used for testing of independence as a null hypothesis. The results on the whole undivided data sample show very similar properties for Bitcoin as for currencies. On splitted subsamples Bitcoin seems to be significantly ineffective at the last interval.

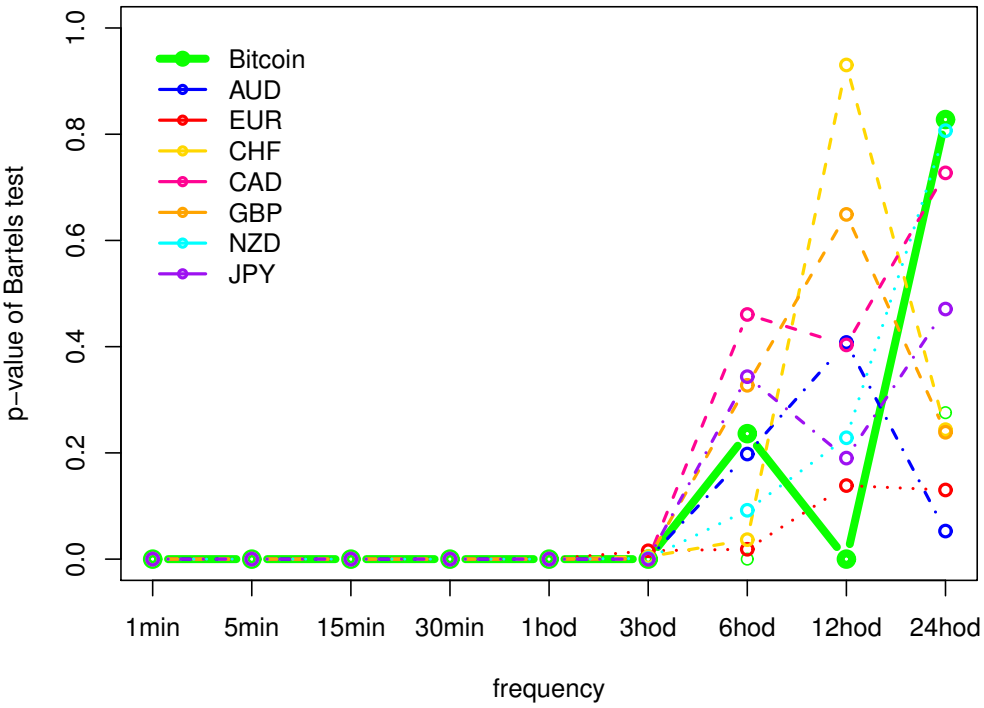

Figure 35: P-values of Bartels test depending on data frequency.

4.7.6 Evolution of p-value of Bartels test over time

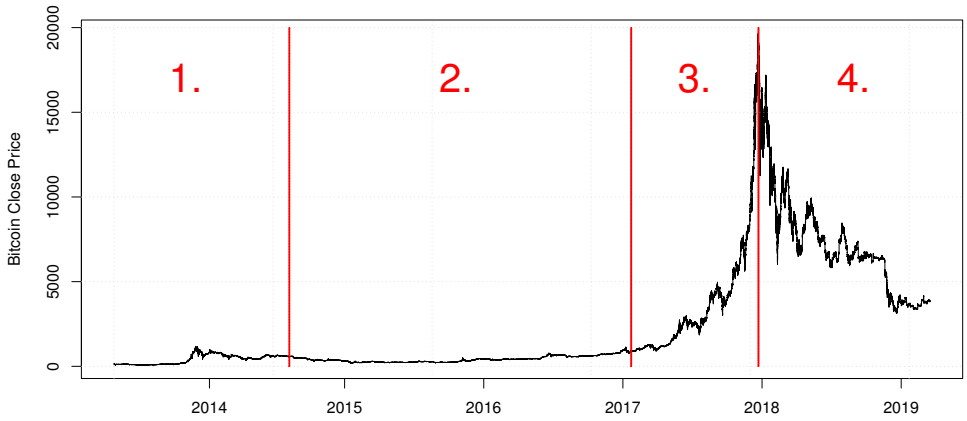

Figure 36: Division into subintervals.

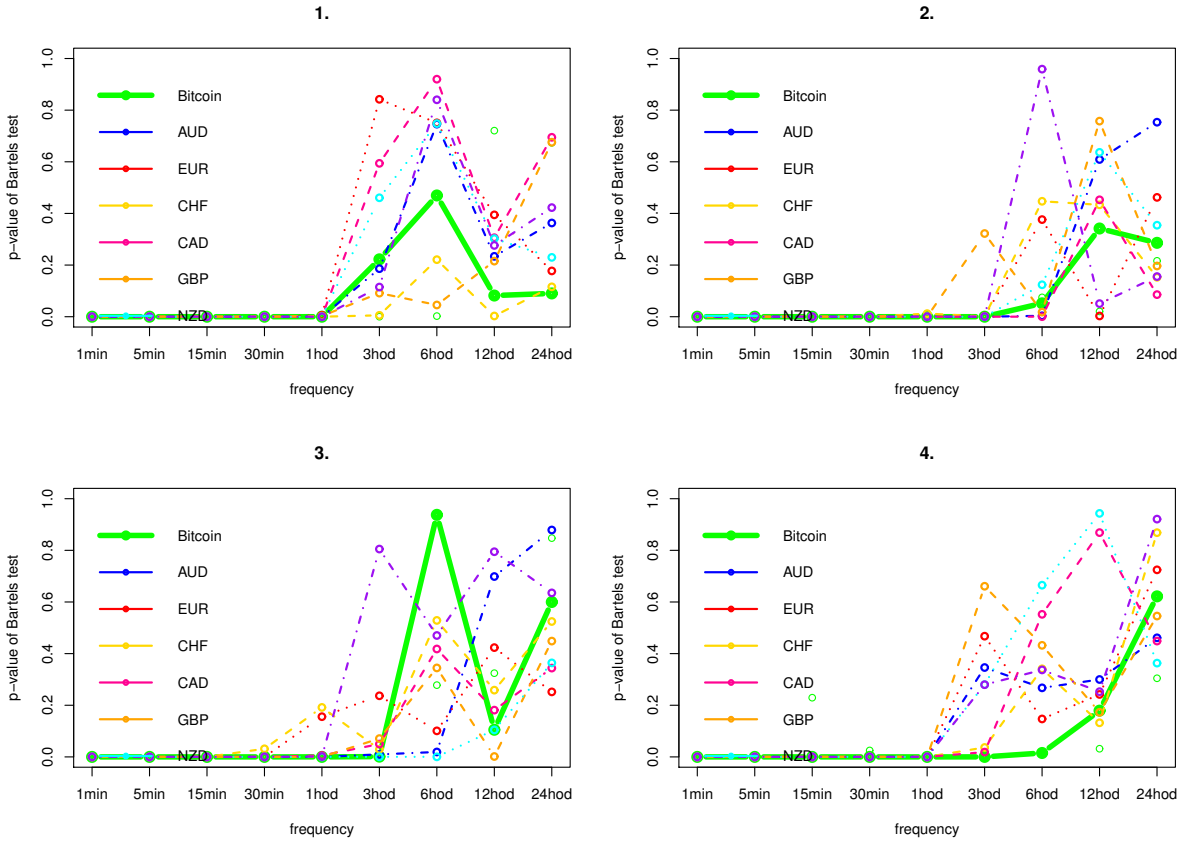

Figure 37: Division into subintervals.

4.7.7 Performing the Automatic Variance ratio test on returns.

The last test, which we used to measure efficiency in Bitcoin and currencies, is the Automatic Variance Ratio Test (AVR test). It checks in null hypothesis: *Wheather the random walk hypothesis holds for returns.* The AVR test is more complicated and therefore we were able to work with 30-minute frequency data at least. Automatic variance ratio test chooses the length of the dependence optimally to find a possible correlation in returns. In this case we choose to display test statistics instead of displaying P-values to display the difference between Bitcoin and currencies easier. The rejection region comes from a standardized normal distribution.

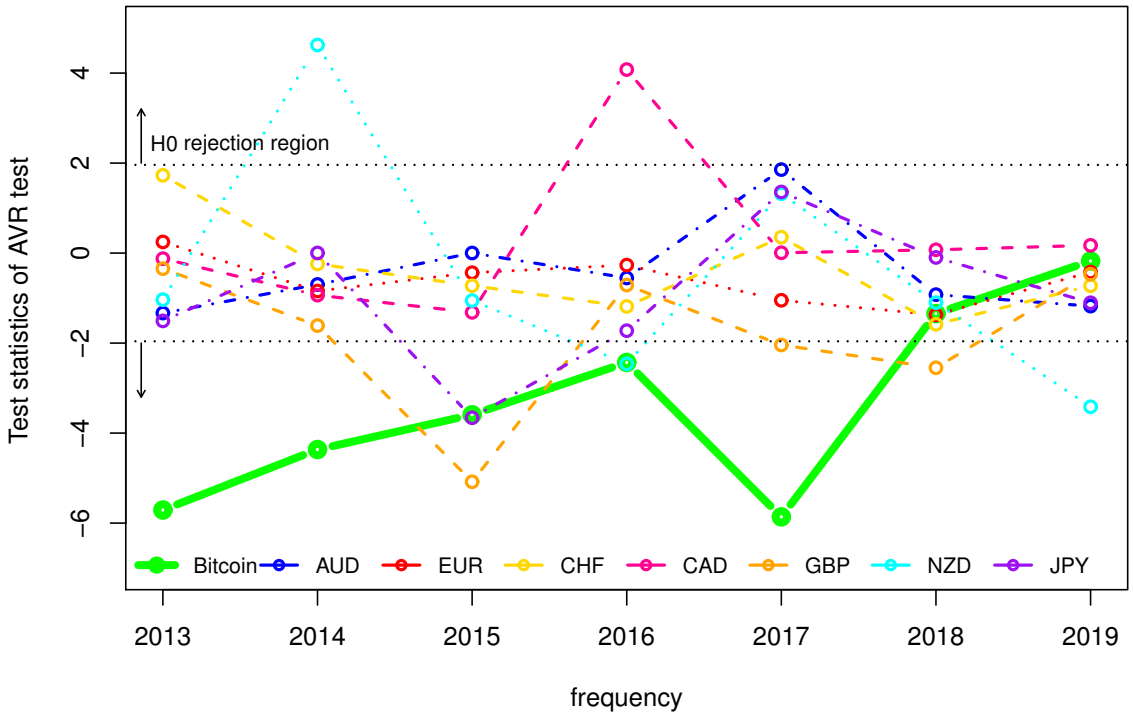

Figure 38: P-values of Bartels test depending on data frequency.

The AVR test does not reject a random walk hypothesis for currencies returns except for a few isolated cases. In the case of a test statistic deviation to the rejection area this deviation is smaller for currencies than for Bitcoin. It can be seen that test statistics of the AVR test for Bitcoin gradually becomes more and more effective. Signs of ineffectiveness that were present in previous tests in the period from early 2018 have disappeared. The AVR test evaluates the period from 2018 to 2019 for Bitcoin as the most effective.

4.8 The Hurst Exponent and long term dependence

The long-term dependence analysis in time series, which was originally applied to the hydrological problems, appears to be meaningful at the financial time series as well [4]. According to [7] investors might obtain a consistently high return by using the existence of long-term dependency in the returns series of the financial assets.

We calculated the Hurst exponent for our full data sample using two methods: R/S and DFA. The calculation is based due to computational complexity on 30-minute data with applying rolling window approach using a one-year time window and two-day shift. According to [7] there might be a problem with calculating the Hurst exponent only on the sub-samples and no overlapping intervals because they may not capture full dynamics and dependency in the data. Therefore we have used this type of approach.

R/S Hurst exponent:

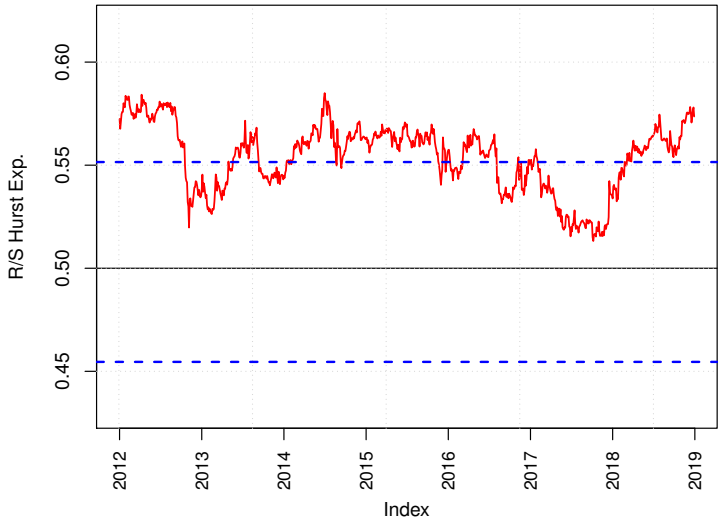

Figure 39: R/S Hurst exp. with conf. interval calculated using rolling window approach.

The results of the successively calculated Hurst exponent are compared with a 95 percent confidence interval (from 0.45 0.55 to) by calculation based on the formulas from Tables 1 and 2. Also in this case it appears that Bitcoin has been showing the signs of the long-term dependence in returns in the beginning of the investigated period and then from mid-2014 until the beginning of the volatile period in 2017.

The DFA hust exponent shows on Figure 40 few intermediate evidences of the anti-

persistence. This short-term dependency came only in four places and only for a short period of time. Over time it disappears and DFA Hurst exponents bounds inside confidence interval.

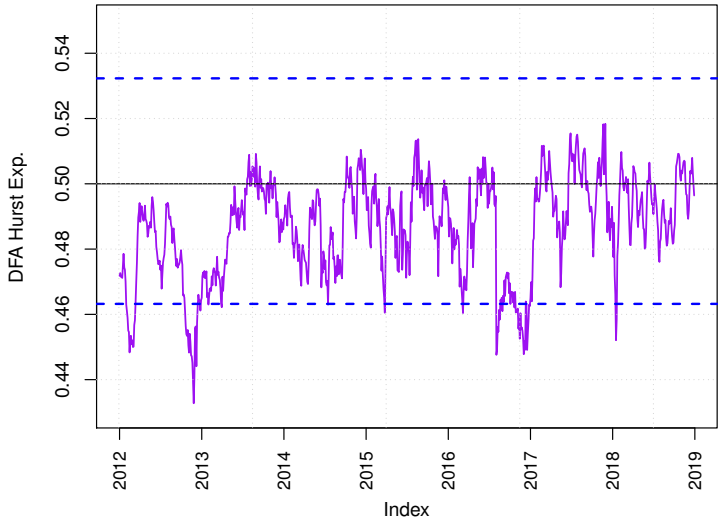

Figure 40: DFA Hurst exponent calculated using rolling window approach.

4.9 The EGARCH MODEL and the evolution of the Laverage effect in time

The linear models, we have been initially working with, were not perfect in the terms of the data modeling. The squared residuals from the second linear model were still considered as dependent by the Ljung-Box test. This is usually one of the indications to use the GARCH model to explain the volatility. The volatility clustering is an obvious effect on the financial market, which also affects Bitcoin [16]. Many people have already worked on Bitcoin modeling using GARCH models. As far as we know, no one has ever explored, what the transmission in the volatility between the currency exchange rates and Bitcoin using the high-frequency data.

Increasing the volatility/tension in the currency market may potentially affect the Bitcoins volatility. To measure the volatility of the currencies we have created an index. From the 15-minute returns for each currency we calculated its 6-hour volatility. Than the average of theese volatilities could be considered as the volatility index.

We will add this newly created variable vol_{t-1} to E-GARCH(1,1) model as the external explanatory variable for model configured using 6-hour data:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \beta \log(\sigma_{t-1}^2) + \sum_{i=1}^{p} \alpha_i \frac{|\hat{\epsilon_{t-i}}| + \Upsilon \hat{\epsilon_{t-i}}}{\sigma_{t-i}} + \delta * \log(vol_{t-1}) \tag{4.3}$$

The model was estimated at 4 subintervals equal to the intervals from the efficiency testing. We added ruturns from the currencies as the external variable to the model for conditional mean. Results are shown in Table 3. The model is configured with using the Student residuals in ARMA modeling. Since the student residuals seemed to us to be the best choice because they better explain the heavier tails of the Bitcoin returns. On QQ plot standardized rezidiuals lays on the line.

Table 3: Estimated E-GARCH coeff. for subperiods.

	1.subperiod	2. subperiod	3.subperiod	4. subperiod
mu	5.25e-05	$0.00036^{\star}$	$0.0028^{\star}$	-0.0012
ar1	$-0.74^{\star}$	$0.99^{\star}$	$0.43^{\star}$	$-0.18^{\star}$
ma1	$0.74^{\star}$	$-0.99^{\star}$	$-0.41^{\star}$	$0.12^{\star}$
mxreg1	0.046	$-0.15^{\star}$	-0.17	0.29
mxreg2	-0.092	$-0.044^{\star}$	-0.47	0.70
mxreg3	-0.34	0.010	0.025	-0.17
mxreg4	$0.56^{\star}$	-0.019	-0.088	0.13
mxreg5	-0.031	0.010	0.29	-0.60
mxreg6	$-0.42^{\star}$	$-0.095^{\star}$	-0.22	-0.0028
mxreg7	0.35	$0.043^{\star}$	0.17	0.44
omega	$-0.93^{\star}$	$-0.11^{\star}$	$-0.28^{\star}$	$-0.021^{\star}$
alpha1	-0.052	0.0036	-0.049	$-0.062^{\star}$
beta1	$0.84^{\star}$	$0.98^{\star}$	$0.95^{\star}$	$0.99^{\star}$
gamma1	$0.81^{\star}$	$0.44^{\star}$	$0.35^{\star}$	$0.024^{\star}$
delta	0.0059	-0.021	-0.053	0.0001
skew	$0.94^{\star}$	$0.977^{\star}$	$0.98^{\star}$	$0.943^{\star}$
shape	$2.213^{\star}$	$2.17^{\star}$	$2.62^{\star}$	$3.91^{\star}$

5 % significance level is displayed by *.

The volatility transfer coefficient δ between currencies and Bitcoin came out in this case not significant for all periods under examination. Currency volatility does not affect Bitcoin volatility. AR(1) and MA(1) coefficient, on the other hand, were significant for all subintervals.

4.9.1 Evolution of the News Impact Curve in time

The E-GARCH model takes into account the different impact of the positive and negative news ϵ_t on the volatility. The increase in volatility is on the market seen as an increased risk. Bitcoin is particularly sensitive to this risk since it is not clear what is the real value of Bitcoin. We will be interested in the evolution of the impact of the positive and negative information on volatility for each subperiod. The capturing of the impact of positive and negative information on the volatility seems to be good since The Sign Bias Test, the Negative Size Bias Test and the Positive Size Bias showed no evidence against the null hypothesis: *the model is adequate.*

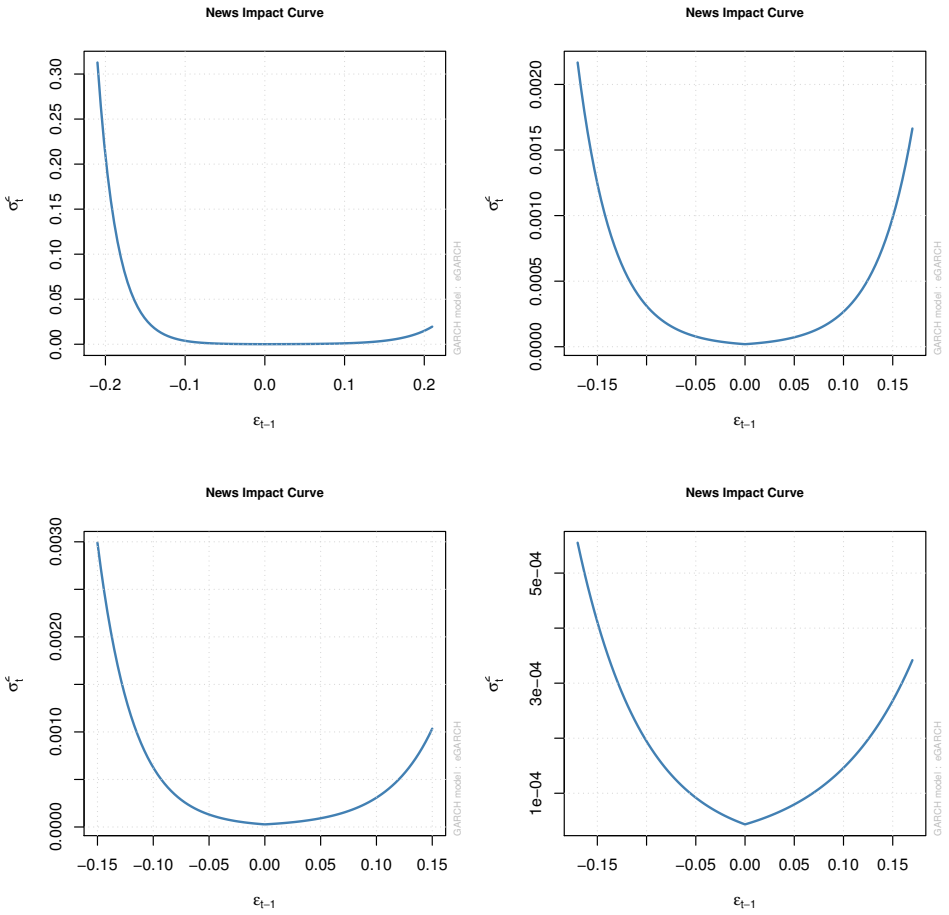

Figure 41: News Impact Curve for subintervals.

In the first period negative information $\epsilon_{t-1} < 0$ has a tremendous impact on the volatility.

In the second period, the effect of negative news on the volatility was approximately the same as the effect of positive news. In general, equally comprehensive information have caused less volatility increase than in the first period.

In the third period of the extraordinary growth of Bitcoin the market is becoming more liquid and at the same time more sensitive to negative news. A large number of investors expect a "bursting" of the bubble. The effect of positive news on the volatility is relatively small.

In the fourth period after 2018, even the arrival of smaller news causes a substantial increase in the volatility.

Conclusion

The main essence of this thesis was to examine and based on various statistical tests compare the effectivity of Bitcoin with the effectivity of the currencies for different frequencies and various subintervals using high-frequency data.

To this objective, we were inspired by the linear model (4.1) in which we predicted Bitcoin returns by the currencies returns. We saw significant dependency on one-minute data, between return and return from the previous minute. Possibility to predicting future returns based on past information is contrary to The Hypothesis of an Efficient Market [10] by which prices in the financial market are driven by Brown's move, and thus returns should be an independent white noise.

We looked at how well can returns from Bitcoin and returns from the currencies modeled by simple AR(1). The negative relationship, estimated on one-minute data probably results from a low liquid market, on which there is a large bid-ask spread. It does not necessarily mean the presence of the mean reversion.

Several authors who tested Bitcoin efficiency on daily data point out that Bitcoin is becoming more effective over time. Therefore, we have decided to verify whether this is the also case with high-frequency data and we compared these results with the results for conventional currencies.

The AR(1) model uses only one historical variable. A large amount of older dependence may not be noticed. Therefore we applied a more sophisticated model for effectivity testing. We used the Ljung-Box test, Runs test, Bartels test and Automatic Variance ratio test. These test were applied for the full period and to the subperiods as well. Each test rejected the efficiency of Bitcoin at various frequencies, particularly Bitcoin was inefficient at higher frequencies. Bitcoin was more inefficient at the beginning when the liquidity was lower. According to the AVR test, the effectivity of Bitcoin grows over time. From 2018 to 2019 Bitcoin seems to be as effective as the currencies.

The long-test dependence is often present in financial time series, which may not be revealed by previous tests. Therefore, we will submit 30-minute high-frequency data to the Hurst exponent analysis. The R/S analysis shows inefficiencies for the entire test period. The only exception was for two intervals. At the end of 2013 and during the rapid growth in 2017. Again, we had confirmed that in times when the Bitcoin market was not so liquid

there is a decrease in efficiency.

One of the reasons why we decided to model Bitcoin's returns in the last section of this work using GARCH models is that in the second linear models the squared residuals from the second linear model were dependent. Usually, GARCH model deprives this dependence. We decided to model 6 hours returns. Among the external regressors for conditional mean, we added currency returns. We added a currency volatility index to the conditional volatility model. However, the volatility of currencies did not affect Bitcoin volatility.

The volatility of Bitcoin is affected by both positive and negative news. We were interested in the development of the impact type and size of information on volatility. Therefore, we have used News Impact Curve estimated on subintervals, which makes it evident that Bitcoin reacts more to negative news in a subperiods with higher liquidity.

After all the fact that Bitcoin returns are non-independent and do not confirm the Efficient Market Hypothesis does not mean that they are predictable.

References

[1] Aalborg, H. A., Molnár, P., Vries, J. E. d.: *What can explain the price, volatility and trading volume of Bitcoin?*, Finance Research Letters (2018),

[2] Urquhart, A.: *The inefficiency of Bitcoin*, Economics Letters 148(2016), 80-82

[3] Nadarajah, S., Chu, J.: *On the inefficiency of Bitcoin*, Economics Letters 150 (2017), 6-9

[4] Bariviera, A., F.: *The inefficiency of Bitcoin revisited: A dynamic approach*, Economics Letters 161 (2017), 1-4

[5] Tiwari, A.K., Jana, R.K., Das, D., Roubaud, D.: *Informational efficiency of Bitcoin - An extension*, Economics Letters 163 (2018), 106-109

[6] Brauneis, A., Mestel, R.:*Price discovery of cryptocurrencies: Bitcoin and beyond*, Economics Letters 165 (2018), 58-61

[7] Yonghong, J., He, N., Weihua, R.: *Time-varying long-term memory in Bitcoin market*, Economics Letters 25 (2018), 280-284

[8] Nakamoto, S.: *Bitcoin: A Peer-to-Peer Electronic Cash System*, ONLINE (13.1.2019): https://bitcoin.org/bitcoin.pdf

[9] Kristoufek, L.: *Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era*, Scientific Reports 3 (2013), 3415-end

[10] Fama, E. F.: *Efficient Capital Markets: A Review of Theory and Empirical Work*, The Journal of Finance 25 (1970), 383-417

[11] Frennberg, P., Hannson, B.: *Testing the random walk hypothesis on Swedish stock prices:*, Journal of Banking and Finance 17 (1993), 175-191

[12] Cowpertwait, P.S.P, Metcalfe, A.V.: *Introductory Time Series with R*, Springer Science, New York, 2009

[13] Melicherčík, I., Olšarová, L., Úradníček, V.: *Kapitoly z finančnej matematiky*, Epos, Bratislava, 2005, ISBN 80-8057-651-3

[14] Kevin.: *What is Bitcoin's Correlation With Other Financial Assets?*, Working paper on internet, ONLINE (13.1.2019): https://www.signalplot.com/what-is-bitcoins-correlation-with-other-financial-assets/

[15] Tonnesen, E. J., Bergloff, S., Molnár, P., Overli, M.: *The Determinants of Bitcoin Liquidity*, Working paper on internet, ONLINE (29.4.2019): https://brage.bibsys.no/xmlui/handle/11250/2574692

[16] Molnár, P., Thies, S.: *Bayesian change point analysis of Bitcoin returns*, Finance Research Letters 27 (2018), 223-227

[17] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y.: *Testing the null hypothesis of stationarity against the alternative of a unit root*, Journal of Econometrics 54 (1992), 159-178.

[18] Bradley, V. J.: *Distribution-Free Statistical Tests*, Prentice-Hall, 1968, ASIN: B0006BVE3K, Chapter 12

[19] Bartels, R.: *The Rank Version of von Neumann's Ratio Test for Randomness*, Journal of the American Statistical Association 77 (1982), 40-46

[20] Choi I.: *Testing the Random Walk Hypothesis for Real Exchange Rates*, Journal of Applied Econometrics 3 (1999), 293-308

[21] Andrews, D. W. K.: *Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation*, Econometrica 3, Vol. 59, (1991), 817-858

[22] Hurst, H.E.: *Long-term storage capacity of reservoirs*, Transactions of American Society of Civil Engineers, Vol. 116, (1951)

[23] Peng, C. K., Buldyrev, S. V., Simons, M.: *Nature and Fractals*, Physics Rev, Vol. 168 (1994)

[24] Weron, R.: *Estimating long range dependence finite properties and confidence intervals*, Journal of Stochastic Methods, Vol. 1-5 (2001)

[25] Bollerslev, T.: *Generalized Autoregressive Conditional Heteroskedasticit*, Journal of Econometrics, Vol. 31 (1986), 307-327

[26] Marček, D.: *Volatility Modelling and Forecasting Models of High Frequency Financial Data: Statistical and Neural Approach*, Ekonomický časopis, Vol. 62 (2014), 133-149

[27] Hansen, P. R., Lunde, A.: *A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?*, Journal of applied econometrics, Vol. 20, No. 7 (2005), 873-889

[28] Engle, R. F., Ng, K. V.: *Measuring and Testing the Impact of News on Volatility*, Journal of Finance, Vol. 48 (1993)

[29] Nelson, D. B.: *Conditional Heteroskedasticity in Asset Returns: A New Approach*, Econometrica, Vol. 59, No. 2 (1991), 347-370

[30] Ljung, G. M., Box, G. E. P.: *On a measure of lack of fit in time series models*, Biometrika, Vol. 65, No. 2 (1978), 297–303

Appendix A

This correlations were calculated using moving 30day time window with 1day shift.

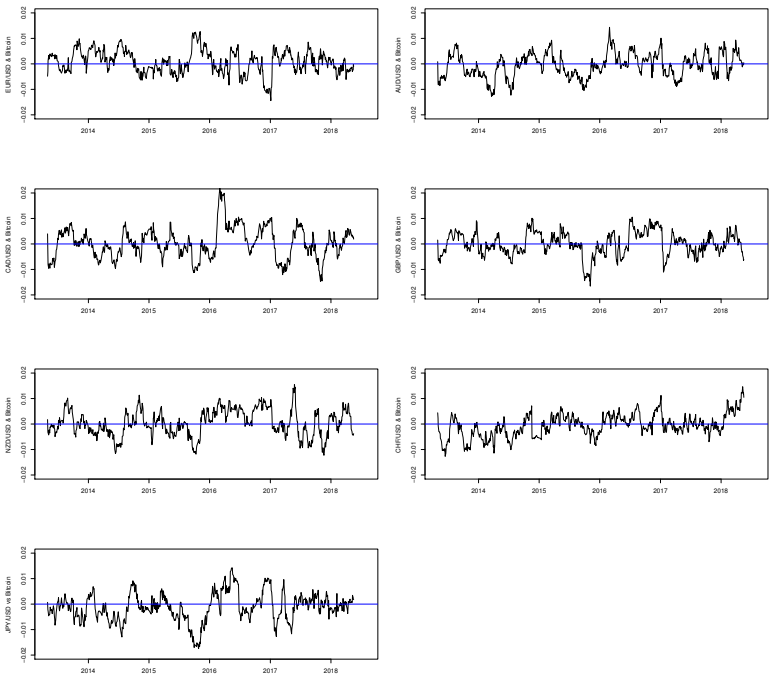

Figure A.42: Floting correlation in between the currency exchange rates and Bitcoin calculated on minute data.