BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF BUSINESS AND MANAGEMENT

FAKULTA PODNIKATELSKÁ

INSTITUTE OF ECONOMICS

ÚSTAV EKONOMIKY

BITCOIN VOLATILITY ESTIMATION

ODHAD VOLATILITY BITCOINU

BACHELOR'S THESIS

BAKALÁŘSKÁ PRÁCE

AUTHOR Petr-Lev Bařina

AUTOR PRÁCE

SUPERVISOR Ing. Jiří Luňáček, Ph.D., MBA

VEDOUCÍ PRÁCE

BRNO 2022

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF BUSINESS AND MANAGEMENT

Assignment Bachelor's Thesis

Department:	Institute of Economics
Student:	**Petr-Lev Bařina**
Supervisor:	**Ing. Jiří Luňáček, Ph.D., MBA**
Academic year:	2021/22
Study programme:	Business Economics

Garantka studijního programu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává Bachelor's Thesis s názvem:

Bitcoin volatility estimation

Characteristics of thesis dilemmas:

Introduction
Objectives of the work, methods, and procedures of processing
The theoretical basis of the work
Analysis of the current condition
Business solutions
Conclusion
List of literature
Attachments

Objectives which should be achieve:

This work aims to create and evaluate a model that will indicate bitcoin volatility on daily data using a suitable statistical instrument. The result will be to compare multiple models based on different techniques and numbers of variables.

Basic sources of information:

BOYLE, Patrick. Statistics for the Trading Floor: Data Science for Investing. 1. London: Independently published, 2020. ISBN 9798644826551.

OLSON, David L. a Desheng WU. Predictive Data Mining Models. 1. Singapore: Springer, Singapore, 2017. ISBN 978-981-10-2543-3.

FRANCES, Philip H. a Richard PAAP. Periodic Time Series Models. 1. Oxford: OUP Oxford, 2004. ISBN 978-0-19-152926-9.

SCHÄR, Fabian a Aleksander BERENTSEN. Bitcoin, Blockchain, and Cryptoassets: A Comprehensive Introduction. 1. Cambridge, MA: MIT Press, 2020. ISBN 978-0-262-53916-6.

Deadline for submission Bachelor's Thesis is given by the Schedule of the Academic year 2021/22

In Brno dated 28.2.2022

L. S.

prof. Ing. Alena Kocmanová, Ph.D. Branch supervisor	doc. Ing. Vojtěch Bartoš, Ph.D. Dean

Abstrakt

Tato bakalářská práce se zaměřuje na ekonometrické modelování a predikci volatility Bitcoinu. První částí práce je teorie statistických vlastností časových řad a modely interpretující tyto vlastnosti. Praktická část je zaměřena na modelování, predikci a hodnocení přesnosti predikcí. Klíčovými modely jsou model klouzavého průměru, autoregresní model, ARCH a GARCH. Poslední částí je shrnutí výsledků a návrhy na zlepšení.

Abstract

This bachelor thesis focuses on econometrics modelling and prediction of Bitcoin volatility. The first part of the thesis is the theory of statistical properties of time series and models interpreting these properties. The practical part focuses on modelling, prediction, and evaluation of the accuracy of the predictions. The Key models are the Moving average model, the Autoregressive model, ARCH and the GARCH. The last part is a summary of results and proposals for improvement.

Klíčová slova

Bitcoin, prognóza, model, časové řady, regresní analýza, ARCH, GARCH

Keywords 3.

Bitcoin, forecast, model, time series, regression analysis, ARCH, GARCH

Bibliographic citation

BAŘINA, Petr-Lev. *Odhad volatility Bitcoinu* [online]. Brno, 2022 [cit. 2022-05-04]. Dostupné z: https://www.vutbr.cz/studenti/zav-prace/detail/131784. Bakalářská práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav ekonomiky. Vedoucí práce Jiří Luňáček.

Affidavit of Originality

This declaration is to attest that the submitted bachelor thesis is original and was prepared exclusively by me. I declare that the citation of the used sources is entirely stated in the list of references, and I did not violate in any way another author's copyrights (in the sense of Act No. 121/2000 Coll., on copyright and related rights with copyright).

In Brno, 05.05.2022

…………………………

Signature

Acknowledgement

I want to express my gratitude to Patrick Boyle, professor of finance at King's College London and Youtuber, for introduction to the field of econometrics and the book that helped me understand the theory used in this thesis.

Also, thanks to Savva Shanaev, lecturer in Finance at Newcastle Business School and YouTuber, for helping me with models construction in Excel.

Lastly, thanks to Ing. Jiří Luňáček, Ph.D., MBA, for guidance through this thesis.

Index

INTRODUCTION

Bitcoin is the first cryptocurrency existing only since 2009. In this short time, it gained immense popularity for its unique properties. However, there are many opinions on Bitcoin in markets. Some call it future reserve currency, internet gold or a total scam. Non the less, investors are seeking ways to include this new asset class in portfolios, which raises the question of how to incorporate Bitcoin in portfolios or hedge against its volatility successfully.

Financial econometrics is a science discipline trying to understand economic and finance phenomena through quantitative analysis. It uses statistical tools to do so. Due to significant quantities of data in financial markets, it is recommended to use the software. In some cases, even using the software can be time demanding because of the amount of data required to analyze. Appropriate pick of model and descriptive variables is a challenge as well. That is why it is recommended to use the most straightforward models possible, trying to avoid using too many unimportant variables.

This thesis is trying to conclude the nature and dynamics of Bitcoin's daily price change time series and how to predict its next movement. This prediction can be helpful for retail and institutional investors trying to expose themselves to Bitcoin and Bitcoin derivatives and help them assign a level of risk for this asset. This thesis's four critical models are the moving average model, autoregression model, ARCH and GARCH.

The first part discusses theoretical starting points about Bitcoin and the main forces behind Bitcoin price creation. The modelling process and its underlying statics rules because models can be used with restrictions based on statistical properties of observed data. Also, verification of models and accuracy of their prediction is described.

The second part is the creation of the models in software Microsoft Excel, deriving Excel functions from equations described in the first part of the thesis. Then evaluating these models based on the theoretical part.

In the third part are the findings described in the second part evaluated.

The last part is a suggestion for improvement.

PROBLEM DEFINITION AND WORK OBJECTIVES

Objectives

The aim of this thesis is to construct and evaluate a model that will predict Bitcoin´s volatility on daily data using a suitable statistical tool. The thesis will focus on comparing multiple models that are based on different procedures and numbers of variables.

Solution methodology

Scientific literature will be used to fulfil the set goals. The correct number of parameters has to be chosen for the successful construction of models. The critical part is computing models based on previous theory and findings in software Microsoft Excel. The four main models are the moving average model, autoregression model, ARCH and GARCH. These four models and their variations will be compared, and if their accuracy is verified, they will be recommended for the prediction of Bitcoin volatility.

Data analysis based on statistics and the econometrics method is crucial in this thesis. Comparing the models and their prediction accuracy is essential in econometrics modelling. Different comparison methods will be used to comprehend changing number of parameters and the difference between the construction of the models.

The practical part is the method of description used to describe the process of construction of the models in software Microsoft Excel. To avoid overfitting, parameters are chosen purely based on previous tests and not with the goal of making the best prediction evaluation results.

1 THEORETICAL PART

This part described the problematics of Bitcoin, statistics, and econometrics models.

1.1 Bitcoin

Bitcoin is a peer to peer decentralized payment network and a token used to settle the transactions on this network. Thanks to complete decentralization using blockchain technology and the token's deflationary properties is a unique type of financial asset (Schär, Berentsen, 2020).

Since its introduction in 2009, when the first Bitcoin was mined, its price went from 0 to a maximum of over 63000 USD in 2021. Moreover, while Bitcoin was seen by investors only as wild speculation in previous years, by 2020, numerous big financial institutions started investing in Bitcoin, and a significant derivatives market grew around it, reaching as high as 2,7 trillion USD in late 2020 (Coindesk.com, ©2022).

1.1.1 Adaptive market hypothesis

This hypothesis tries to combine principles of efficient market hypothesis and behavioural finance. The adaptive market hypothesis states that markets and people are mainly rational, but there are times when investors can overreact in times of growing market volatility (Lo, 2004).

External and internal factors influence the price discovery of Bitcoin. On the external, we can have an exchange rate, inflation, market cycles or bad and good news like banning Bitcoin mining. Internal factors are fundamental processes in the Bitcoin network, such as hash rate influencing the speed of mining new blocks or inflation that dictates how many new Bitcoins are in mined blocks as a reward for miners. Every approximately four years, this reward is halved, halving inflation of the Bitcoin token. Over time Bitcoin will become a deflationary asset as there are only 21 million tokens to be mined (Schär, Berentsen, 2020). Also, the lack of regulation and the crypto derivatives market can influence the price. The volatility of Bitcoin is probably the result of the nature of internal factors that differ from internal factors of other assets like stocks and bonds (Alexander, Heck, 2020).

Whatever causes this significant volatility in Bitcoin, many findings suggest that the trading of Bitcoin tokens is getting more efficient over time, supporting the adaptive market hypothesis. There are times of greater and smaller efficiency, but over time volatility of Bitcoin declines while efficiency goes up (Khuntia, Pattanayak, 2018).

1.2 Mathematical modelling

Mathematical modelling uses different mathematical operations to represent and understand data sets and real-world phenomena. Sometimes models can help us understand the complexity of a given problem and can help us predict what we can expect in the future (Olson, Wu, 2017).

1.3 Statistics and econometrics

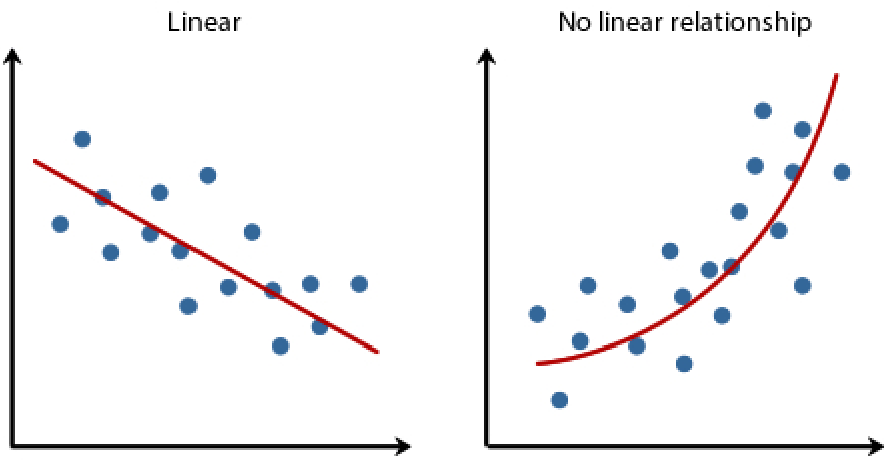
We live in a world full of randomness and probability. Statistics is the discipline of using mathematical techniques that helps us collect, analyze, and present data. Modern statistics was developed at the end of the 19^{th} century and came from the older mathematical field of probability.

Econometrics is a science field using tools of mathematics and statistics in the economic field. Economics and finance have always been about data, so financial institutions were one of the first adopters of statistics and econometrics.

When trying to understand and predict some phenomenon, a model is made by presenting a simplified version of an actual situation. The model then shows the main features of our phenomenon while not taking irrelevant features into account (Boyle, 2020).

1.4 One factor linear regression

It is one of the basic mathematical models that allows a linear relationship between the dependent variable Y and one independent or explanatory variable X to be found (Olson, Wu, 2017). For a set of independent variables, we can estimate the average value of Y as a conditional expectation $E(Y|X)$ or as $\mu(X)$. If $\mu(X)$ is not constant and varies with X, then dependent variable Y has regression on independent X. This variability may be linear or nonlinear (Altman, Krzywinski, 2015).



Picture 1: **Linear relationship in data set**
(SALT DATA LABS, © 2021)

Simple linear regression is a straight line.

Equation 1: **Simple linear regression**
(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$E(Y|X) = \mu(X) = \beta_0 + \beta_1 Y$$

Where:

β_0 is the *Y* intercept of the line, the predicted value of *Y* when *X*=0

β_1 is the slope

The deviation of the model:

Equation 2: **Error term in linear regression**
(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$Y = \mu(X) + \varepsilon, \mu(X)$$

So, the linear regression model can be written as:

Equation 3: **Simple linear regression model**
(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

Now to forecast y_t at unobserved x_t a simple substitution of values can be done to estimate regression at that point.

To estimate β_0 and β_1 is a commonly used least-squared estimator, which minimizes the sum of squared errors or residuals (*SSE*). If (x_t, y_t) are values of *X* and *Y* and $(\hat{x}_t, \hat{y}_t)$ points on the estimated regression line in time *t*, where $\hat{y}_t = m(x_i) = b_0 + b_1 x_t$ (Altman, Krzywinski, 2015).

Equation 4: **Sum of squared residuals**

(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$SSE = \sum (y_t - \hat{y}_t)^2$$

(Altman, Krzywinski, 2015)

The estimates are given by:

Equation 5: **Estimation of $\boldsymbol{\beta_0}$ coefficient**

(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$b_0 = \bar{Y} - b_1\bar{X}$$

Equation 6: **Estimation of $\boldsymbol{\beta_1}$ coefficient**

(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$b_1 = \frac{r\sigma_X}{\sigma_Y}$$

Where $r = r(X, Y)$ is the correlation coefficient of X and Y.

Equation 7: **Estimation of $\boldsymbol{\beta_1}$ coefficient**

(Source: Processed according to Altman, Krzywinski, 2015, p. 999)

$$b_0 = \bar{Y} - \frac{r(X, Y)\sigma_X}{\sigma_Y}\bar{X}$$

1.4.1 Multiple Linear regression

Adding more than one independent variable to a regression model does not mean it will automatically offer a better prediction. This effect is called overfitting when an independent variable may be added under a bias. Also, if not careful, some of the variables might be related to each other, influencing the model prediction in an unwanted way. This phenomenon is called multicollinearity. Therefore, when choosing the correct independent variables, they should be correlated to our dependent variable but not to each other (Olson, Wu, 2017).

After picking the correct regressors, our multiple linear regression model is similar to one factor linear regression. It is the total of our linear parameters (Hyndman, Athanasopoulos, 2018).

Equation 8: **Multiple linear regression**

(Source: Processed according to Hyndman, Athanasopoulos, 2018)

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} + \varepsilon_t$$

Now with each independent variable, it is harder to find the value of every constant β_n so usually, an estimated multiple regression equation is used.

Equation 9: Estimation of multiple linear regression

(Source: Processed according to Boyle, 2020, p. 217)

$$\widehat{y_t} = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_n x_{nt}$$

Where $b_0, b_1, \ldots, b_n$ are estimates of $\beta_0, \beta_1, \ldots, \beta_n$ and $\widehat{y_t}$ the predicted value of the dependent variable with the error term assumed to be zero. Each coefficient b_n is interpreted as the estimated change in $\hat{y}$ corresponding to a one-unit change in a variable while all other variables are constant (Boyle, 2020).

1.5 Time series

Time series is a type of set where the data $d_1, d_2, \ldots, d_n$ is recorded chronologically, preferably in constant time intervals. In finance, time series records the movement of price data for a given asset. Although there is no minimum or maximum amount of time, time series might record daily, hourly, or tick data. We distinguish two basic types of time series: discrete and continuous (Frances, Paap, 2004). The indicator's value in the discrete-time series indicates the status of the indicator, for example, the price of a stock for a given day, while the indicator's value in the continuous-time series arises throughout the time interval, for example, the amount of sold products per year.

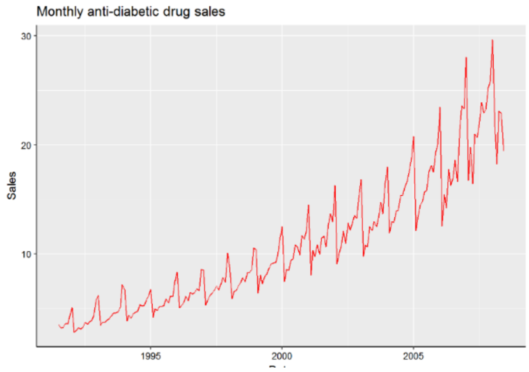
In time series is essential to consider randomness in the probability distribution. However, we do not know all variables influencing the value of our time series. Because of that, we will be using models that take randomness into account.

The data should be recorded and presented to provide us with an easy understanding and future work with the data. Therefore, our first step should be visualizing the data using a graph. This can help us orient what type of time series we are dealing with and possibly find some repeating patterns or anomalies (Kirchgässner, Wolters, Hassler, 2012).

1.5.1 Time series analysis

Time series analysis is the technique of analysing data recorded in time series to understand past movements and sometimes help us with future predictions.

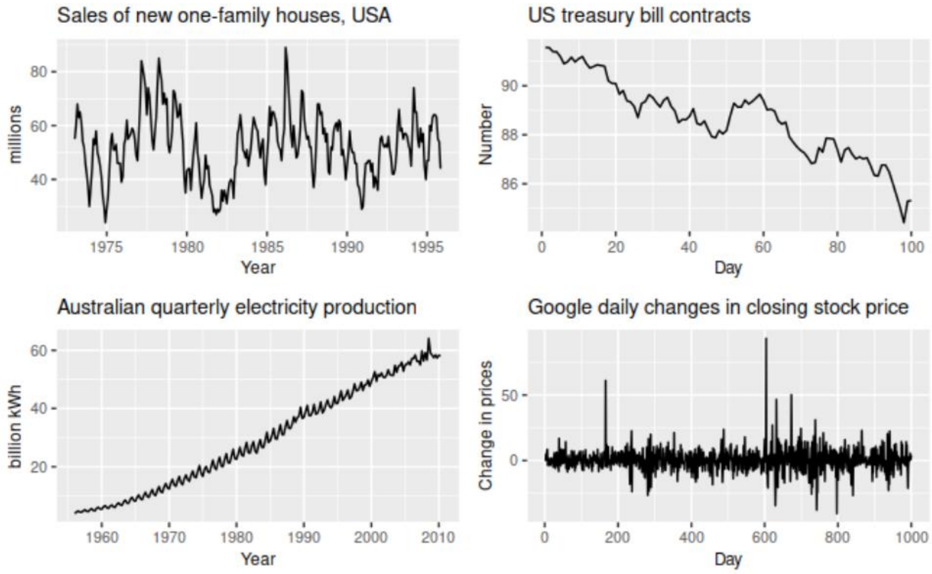
Seasonality is one of the patterns that can be found in time series. Seasonality occurs regularly, for example, in the same month of the year or day of the week. Frequency is a number of regularly recorded intervals of the same length that it takes for seasonality to occur again (Frances, Paap, 2004).



Picture 2: Example of seasonality

(Source: Hyndman, Athanasopoulos, 2018)

Trends can also be observed in time series, which is stable growth or decreased values in numerous consecutive observations (Frances, Paap, 2004).



Picture 3: Examples from top left: seasonality and cycle, trend, seasonality and trend, white noise

(Source: Hyndman, Athanasopoulos, 2018)

1.5.2 Stochastic process

A random process or stochastic process is a set of random values in the same probability space $(\Omega, \mathcal{F}, P)$, indexing them with t from the set $(T \in R)$ where t represents time.

Parameter *t* allows us to split the stochastic process in continuous time into a stochastic process in discrete time, where *T* is a time interval.

The stochastic process may be represented as a function of two variables, *t* and ω. Where *t* is the parameter of time and ω is a random event. The stochastic process considers the joint probability distribution of values taken at different times. For $\{x_1, x_2, \ldots, x_n\}, x \in X$ taken at the time $\{t_1, t_2, \ldots, t_n\}, t \in T$ the dynamic characteristics of the stochastic process may be defined as *n*th order probability distribution function(Godfrey, 1980).

Equation 10: Stochastic process function
(Source: Processed according to Godfrey, 1980)

$$F(X_1, X_2, \ldots, X_n) = P[x_1 \leq X_1, x_2 \leq X_2, \ldots, x_n \leq X_n]$$

1.5.3 Data distribution

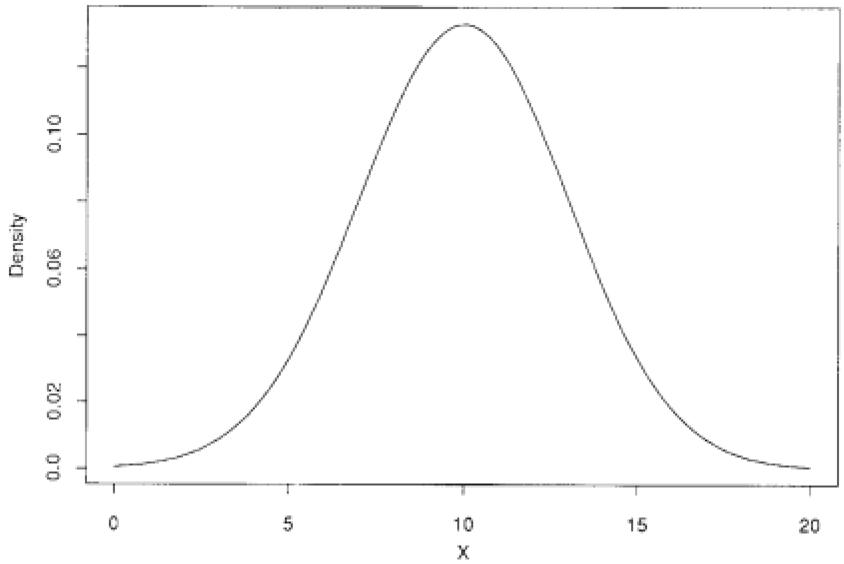
Alternatively, a probability distribution is a mathematical formula of the probability of each value of a variable. This distribution specifies the variable's probability of being under the curve within some interval in a continuous random variable. Sometimes term probability density is used. Several data distributions can be observed (Skrondal, Everitt, 2010).

1.5.3.1 Normal distribution

Also known as Gaussian distribution is one of the data distribution types. It is a continuous probability distribution function with a bell-shaped curve. For random variable X, the probability distribution $F(x)$ following normal distribution is given by:

Equation 11: Normal distribution function
(Source: Processed according to Skrondal, Everitt, 2010, p. 305)

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$



Picture 4: **Normal distribution**

(Source: Skrondal, Everitt, 2010, p. 306)

To fit normal distribution through data set can be z-score used, also called standard scores. It calculates transformed variables' values with zero mean and unit variance (Skrondal, Everitt, 2010, p. 410).

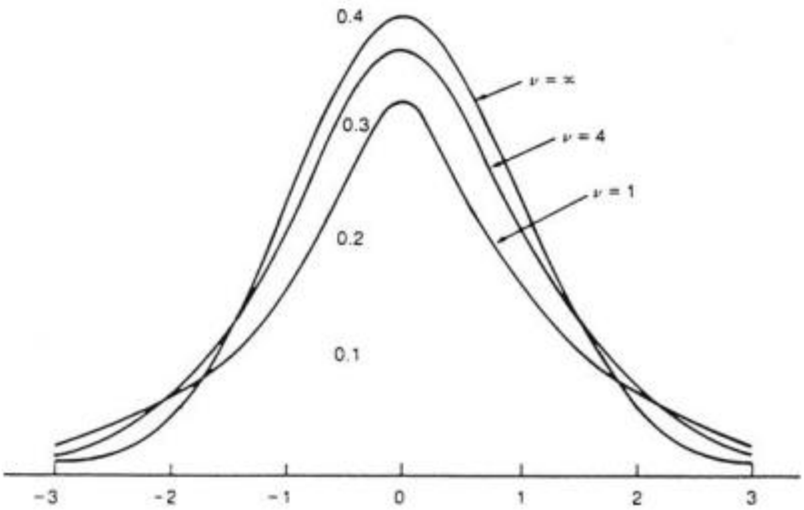
1.5.3.2 Student´s t-distribution

This distribution is often used in conditional probability. Because the shape of this function has thicker tails, it provides a multivariate distribution. This is used in Bayesian inference, a statistical technique in which Bayes´ theorem is used for actualizing the probability of hypotheses with each new evidence (Skrondal, Everitt, 2010).

Equation 12: **Random variable distribution under student´s t-distribution.**

(Source: Processed according to Skrondal, Everitt, 2010, p. 419)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$



Picture 5: **Student´s t-distribution**
(Source: Skrondal, Everitt, 2010, p. 419)

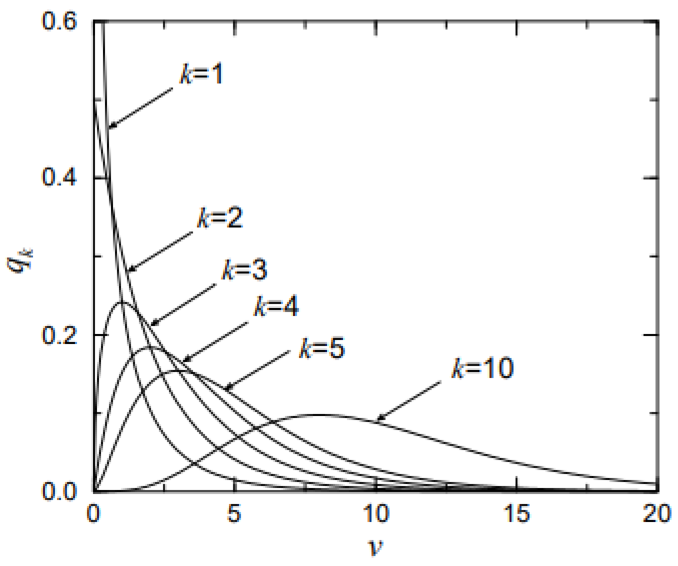
1.5.3.3 Chi-squared distribution

This distribution is used in many areas of statistics, mainly in evaluating the likelihood of models with changing number of parameters. For random variable with probability distribution function $f(x)$ it is given by the sum of squares of a number of independent variables (Skrondal, Everitt, 2010).

Equation 13: **Chi-squared distribution function**
(Source: Processed according to Snyder, 2003, p. 19)

$$f(x) = \frac{x^{\frac{(k-2)}{2}} e^{\frac{-\nu}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)}$$

Where k is the number of degrees of freedom. The mean of this distribution is k and variance $2k$ for $\nu > 0$ (Skrondal, Everitt, 2010).



Picture 6: Chi-squared distribution for different degrees of freedom

(Source: Snyder, 2003, p. 19)

1.5.4 Variance and standard deviation

Variance is the average of the squared differences from the mean. The higher variance means more deviated from the mean of data points. The variance can be only positive by squaring the difference, but units of variance are also squared. For example, when calculating the weight variance of the data set nominated in kilograms, variance gives us back kilograms squared. That interprets this statistic challenging, so we use standard deviation when describing the properties of data sets. Still, we need variance calculation for many different equations (Boyle, 2020).

Population variance of random variable X with mean μ_X and observations at time $t \in T$:

Equation 14: Population variance

(Source: Processed according to Boyle, 2020, p. 117)

$$Var(X) = \sigma_X^2 = E[(X - \mu_X)^2] = \frac{\sum_{t=1}^{n}(x_t - \mu_X)^2}{n}$$

Sample variance for the variable X is similar, but we are using Bessel's correction to correct bias in the estimation of population variance:

Equation 15: Sample variance

(Source: Processed according to Skrondal, Everitt, 2010, p. 445)

$$s_X^2 = \frac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n - 1}$$

As stated before, variance is not suitable for understanding data volatility or comparing two data sets. That is why standard deviation is commonly used for this task.

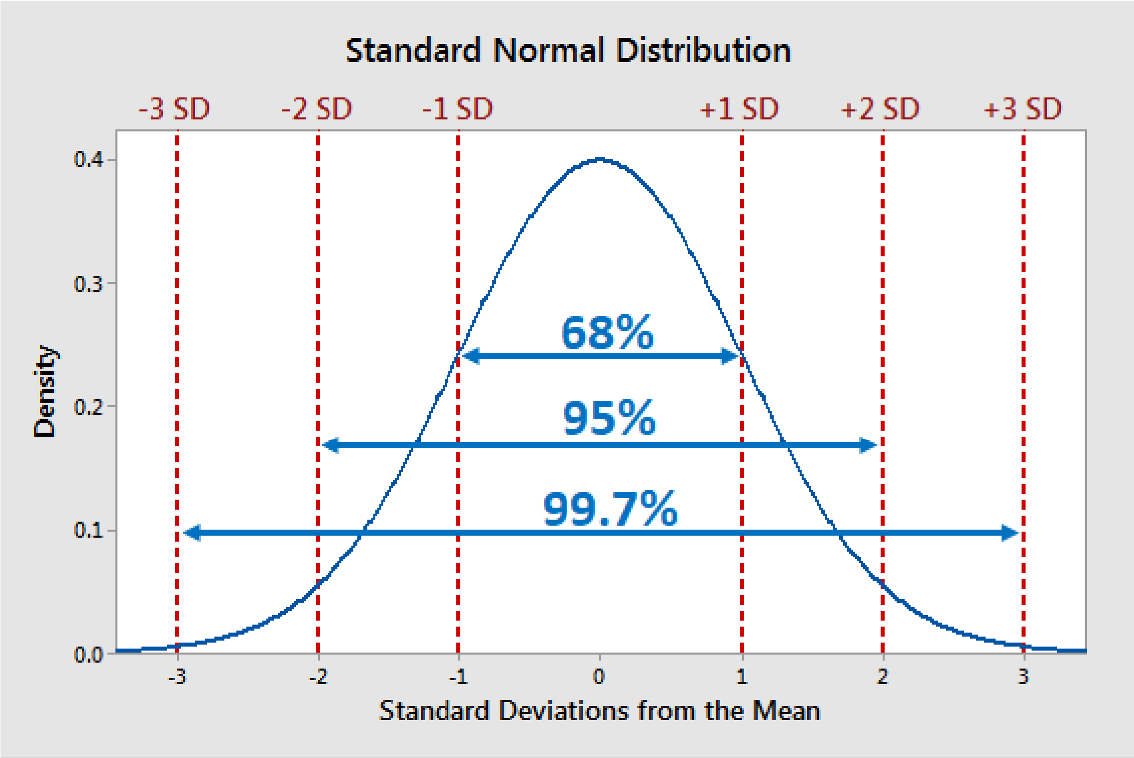
Equation 16: Population and sample standard deviation

(Source: Processed according to Boyle, 2020, p. 119)

$$\sigma_X = \sqrt{\sigma_X^2}$$

$$s_X = \sqrt{s_X^2}$$

The most exciting standard deviation property is the so-called empirical or 68-95-99.7 rule. Under normal distribution, 68% of data lies within one standard deviation from the mean, 95% within two and 99.7% within three (Boyle, 2020).



Picture 7: Empirical rule

(Source: Frost, © 2022)

The standard deviation of a sampling distribution is called standard error (Skrondal, Everitt, 2010).

Equation 17: Standard error

(Source: Processed according to Skrondal, Everitt, 2010, p. 409)

$$SE = \frac{\sigma}{\sqrt{n}}$$

1.5.5 Covariance

Covariance is a descriptive measurement of the joint variability of two random variables. A positive value indicates an increasing linear relationship, and a negative value indicates a decreasing relationship. For the most part, it shows the direction of the linear association between two variables, not its strength (Boyle, 2020).

Equation 18: Population covariance

(Source: Processed according to Skrondal, Everitt, 2010, p. 110)

$$cov(X,Y) = \frac{\sum_{t=1}^{n}(X_t - \mu_X)(Y_t - \mu_Y)}{n} = E[(X - E[X])(Y - E[Y])]$$

Equation 19: Sample covariance

(Source: Processed according to Skrondal, Everitt, 2010, p. 110)

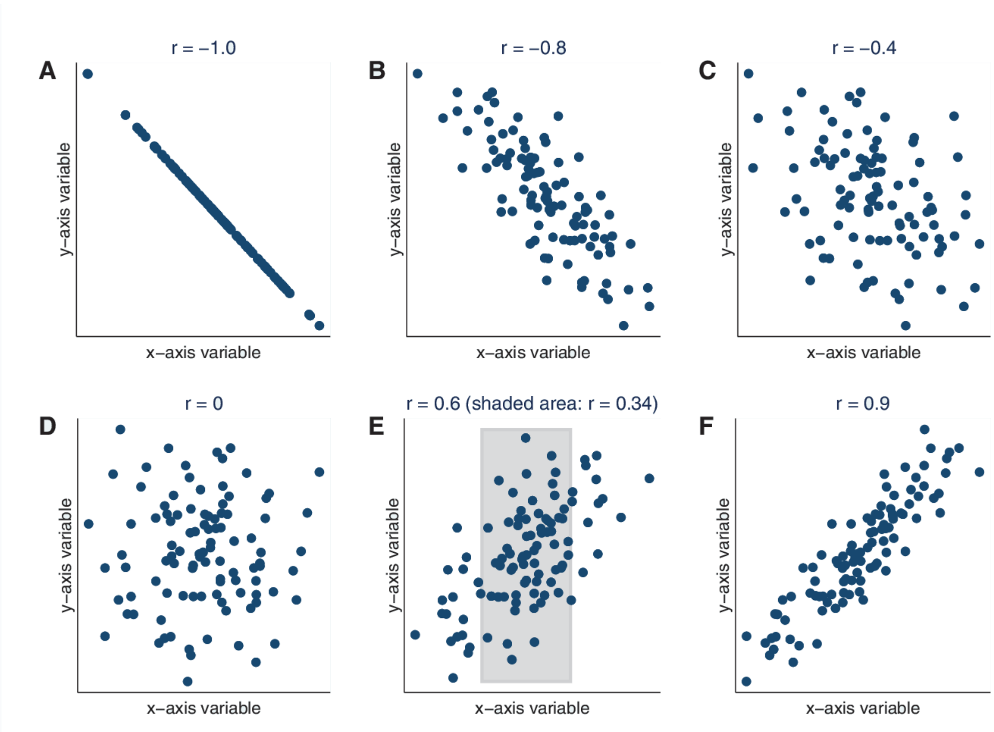
$$c_{xy} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{n - 1}$$

1.5.5.1 Autocovariance

Autocovariance is the covariance measurement of one variable against itself at a different time. The autocovariance measures the linear dependence between two points on the same series observed at different times (Skrondal, Everitt, 2010).

1.5.6 Correlation

The correlation coefficient shows us a linear relationship between two variables, whether causal or not. It measures the strength of this relationship which always lies between -1 and +1. The further the values are from zero, the stronger the relationship between the two variables. Also, a positive correlation coefficient indicates a positive relationship, and a negative one indicates a negative one (Boyle, 2020).



Picture 8: Correlation examples
(Source: Schober, Boer, Schwarte, 2018)

The most common linear correlation coefficient is the Pearson correlation coefficient ρ.

Equation 20: Correlation coefficient
(Source: Processed according to Boyle, 2020, p. 172)

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E[X])(Y - E[Y])]}{\sigma_X \sigma_Y}$$

Population correlation is defined for $\rho_{X,Y}$ between variables X and Y with expected values μ_X and μ_Y and standard deviation $\sigma_X, \sigma_X \in (0, \infty)$ and $\sigma_Y, \sigma_Y \in (0, \infty)$. Correlation has a symmetry property where $\rho_{X,Y} = \rho_{Y,X}$ (Godfrey, 1980).

1.5.6.1 Autocorrelation

It is a function of the time lag between observations in a time series. It is defined at lag $k, \gamma(k)$ as:

Equation 21: Population autocorrelation
(Source: Processed according to Skrondal, Everitt, 2010, p. 25)

$$\gamma(k) = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{E(X_t - \mu)^2}$$

A plot of the autocorrelation values against lag is known as the autocorrelation function (Skrondal, Everitt, 2010).

1.5.6.2 ACF

The autocorrelation function (ACF) takes a sequence of shifts between the autocorrelation values in time n and n-1. For example, at shift zero, the AC is one and with each shift, it decays down. For stationary time series, ACF drops to 0 quickly, while in non-stationary series, it decreases slowly (Skrondal, Everitt, 2010).

Equation 22: Autocorrelation function
(Source: Processed according to Skrondal, Everitt, 2010, p. 25)

$$ACF(k) = \frac{\sum_{i=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{i=1}^{n-k}(x_t - \bar{x})^2}$$

The critical value for individual autocorrelation of lag of order k with a 5% significance level under the null hypothesis is:

Equation 23: ACF critical values
(Source: Processed according to Lütkepohl, 2006, p. 161)

$$ACF(k) > \pm \frac{2}{\sqrt{n-k}}$$

1.5.7 Stationarity

If values of a time series do not depend on the observation time, we can call this time series stationary, like white noise. Therefore, if we observe a trend or seasonality in the data set, it is not stationary. So, in the long run, stationary time series have no predictable patterns, and their variance is constant.

We can spot stationary series visually. There are some main properties such time series has:

- The mean of the series is constant
- Variance σ is constant
- No seasonality
- Constant autocovariance

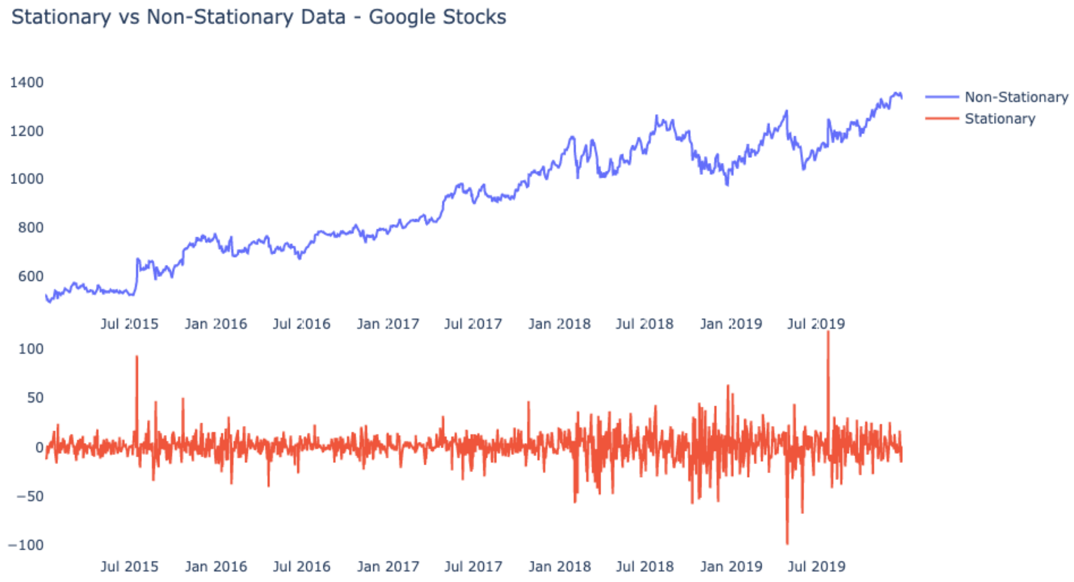
(Park, 2018)

1.5.7.1 Differencing

Differencing is a process of removing trends in time series. The first difference in time series is defined as:

Equation 24: First-order differencing
(Source: Processed according to Skrondal, Everitt, 2010, p. 134)

$$D_{d_t} = d_t - d_{t-1}$$

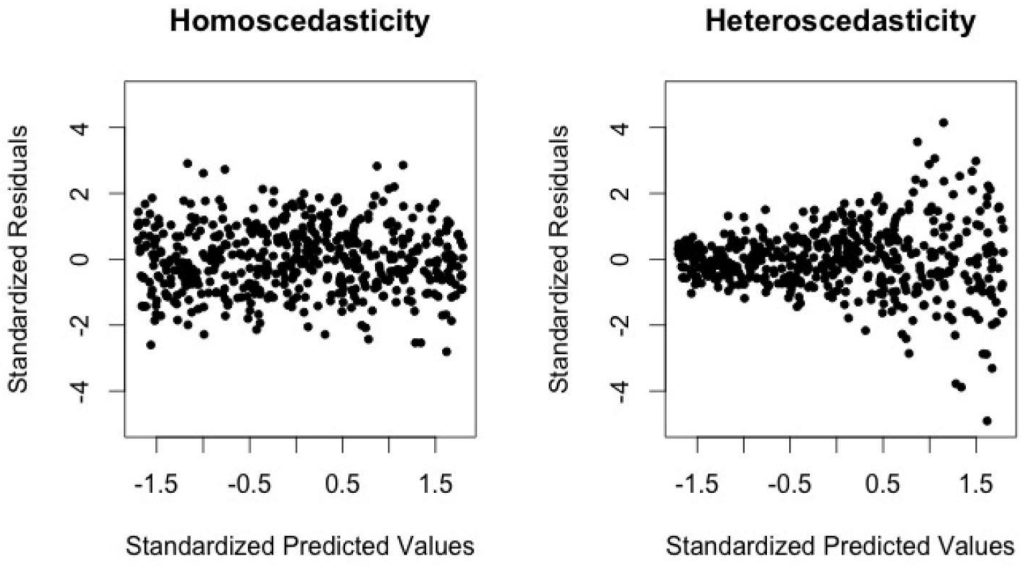


Picture 9: Example of differencing
(Source: Wise, 2020)

After differencing a non-stationary time series, the mean and variance become constant (Boyle, 2020).

1.5.8 Heteroskedasticity

Heteroskedasticity in statistics means the variance's dependence on a parameter, so if the variability of the time series is different to a vector of random variables. Thus, there is no homoscedasticity, where the variance of all variables is constant. The existence of heteroskedasticity can impact the significance of our regression analysis tests because standard tests assume that the modelling errors have the same variance (Boyle, 2020).



Picture 10: Examples of Homoscedasticity and Heteroscedasticity
(Source: Python in Plain English, 2021)

1.6 Null hypothesis

The null hypothesis is a standard statistical theory suggesting that the difference between two possibilities is caused by change alone. Therefore, testing for the null hypothesis can give us the likelihood that the null hypothesis is true (Skrondal, Everitt, 2010).

Testing requires us to create a statistical model or test results of what the data would look like if any phenomenon responsible for the result was pure chance or a random stochastic process in descriptive statistics. Obtained results are then compared to expected results, and thereby the null hypothesis is either rejected or confirmed.

To test the null hypothesis, we can use a p-value or t-stat (Boyle, 2020).

1.6.1 P-value

In testing the null hypothesis, the p-value gives the probability of obtaining results as extreme as the sample result if the null hypothesis is true. It does not tell the probability of the result being true or false but the probability of obtaining such a result if the hypothesis is true (Stigler, 2008). The lower the p-value, the lower the probability of obtaining the result, thus more robust evidence to reject the null hypothesis.

Because the p-value gives the probability in extreme cases, it calculates the value under the upper or lower tail of the distribution (Boyle, 2020).

1.6.2 T-test

Alternatively, a student's t-test tests a statistical hypothesis in which test results follow a student's distribution under the null hypothesis. It is used to decide if the means of two data sets are significantly different.

The T-test can only be used for one or two groups of data, and as a parametric test, our data must fulfil assumptions:

- The data points are independent
- The data approximately follow a normal distribution

(Boyle, 2020)

There are three types of t-test:

- One sample t-test to check if the mean of a population is equal to a hypothesized value
- Two sample t-test to check the means of the population for who independently chosen samples to hypothesized value and test if they are different from each other
- Paired t-test to test the change in the mean of a population before and after intervention

(Skrondal, Everitt, 2010)

1.6.3 F-statistic

To decide if an additional variable should be added to regression or not, F-statistic can be used. It is the change in the squared sum of residuals between two regressions (Skrondal, Everitt, 2010).

Equation 25: F-statistics
(Source: Processed according to Skrondal, Everitt, 2010, p. 286)

$$F = \frac{RSS_m - RSS_{m+1}}{RSS_{m+1}/(n-m-2)}$$

Where RSS is the residual sum of squares with *m* explanatory variables. P-value can be calculated as an F probability distribution for two data sets (Skrondal, Everitt, 2010).

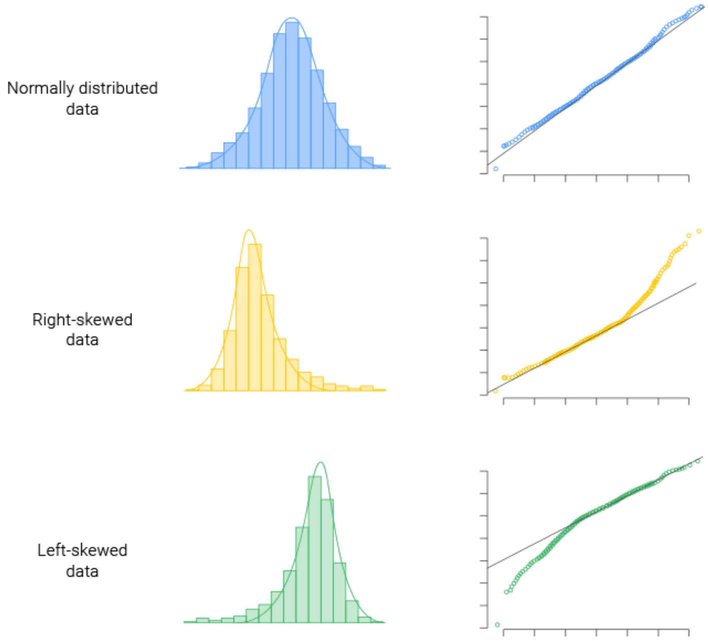
1.7 Normality test

The normality test determines if the data set follows a normal distribution and how likely a random variable from the data set is usually distributed. (Ghasemi, Zahediasl, 2012) The assumption of normality is essential when constructing reference intervals for variables in our prediction models for time series. However, with a significant enough sample size, violation of normal distribution in the data set should not cause significant problems with our models according to the central limit theorem by Pierre-Simon Laplace. The central limit theorem states that a random sample of size n, from data set with mean μ and variance σ^2, distributes normally with mean μ and variance σ^2/n (Kwak, Kim, 2017). Parametric tests developed with the central limit theorem in mind can produce more accurate estimates.

We can test normality by analysing graphs such as Q-Q plot and frequency distribution or by calculating normality tests like Kolmogorov-Smirnov.

1.7.1 Q-Q plot

Q-Q plots visualize the distribution of the given data set and help us decide if it follows normal distributions by plotting quantiles of a variable's distribution against quantiles of normal distribution. Any data points lying away from the line of the normal distribution are outliners that do not follow the normal distribution (Öztuna, Elhan, Tüccar, 2007). It is a visual comparison of theoretical and empirical z-score (Skrondal, Everitt, 2010).



Picture 11: Q-Q plot examples
(Source: LearnByExample, © 2019)

1.7.2 Frequency table

A frequency table displays data by splitting them into groups by value. In statistics, we can visualise these tables to determine if the data set follows normal distribution compared to the graph of a theoretical normal distribution with the same number of observations (Skrondal, Everitt, 2010).

1.7.3 Kolmogorov-Smirnov test

It is a nonparametric test for comparing the cumulative distribution function of the data set to the empirical cumulative distribution function by quantifying their distance. The null hypothesis is that the cumulative distribution of the data follows the null distribution and is based on the p-value of the most significant deviation. It was first proposed by A. N. Kolmogorov and later developed by N. Smirnov (Öztuna, Elhan, Tüccar, 2007).

The empirical distribution function $F(n)$ of normal distribution for n independent and identically distributed ordered observations X_j is defined as:

Equation 26: Empirical distribution function
(Source: Processed according to Mahmoud, 2000, p. 106)

$$F_n(t) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_j \leq t\}}$$

Then the Kolmogorov-Smirnov statistic for the given cumulative distribution function $F(t)$ is:

Equation 27: Kolmogorov-Smirnov statistics
(Source: Processed according to Magg, Dicaire, 1971, p. 653)

$$KS\ statistics = n * \sqrt{sup|F(t) - G(x)|}$$

Where:

sup_x is the supremum of the set of distances

$F(t)$ is the empirical cumulative distribution function

$G(x)$ is a theoretical distribution function

The theoretical distribution function can be obtained, for example, through a theoretical z-score (Skrondal, Everitt, 2010).

1.8 Durbin-Watson test

Durbin-Watson is probably the most used test for autoregression in statistics. The null hypothesis tested in the DW test is that the residuals from linear regression are not correlated. The alternative is that the residuals follow a first-order autoregression process.

Equation 28: Durbin-Watson test
(Source: Processed according to Skrondal, Everitt, 2010, p. 145)

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

DW statistics gives a number from interval <0;4>, where numbers closer to zero indicates positive autocorrelation and numbers closer to 4 negative autocorrelations. Generally, values around two are considered to prove the null hypothesis of no autocorrelation (Skrondal, Everitt, 2010).

1.9 Unit root testing

Unit roots are one cause of non-stationarity in time series. An Autoregression model of order 1, or AR (1) model, is stationary if it has constant covariance. That is achieved if the lag coefficient is less than 1 (Boyle, 2020).

Equation 29: **Stationary process**
(Source: Processed according to Skrondal, Everitt, 2010, p. 442)

$$y_i = \mu + \theta y_{t-1} + \epsilon_i$$

The null hypothesis is that $\theta = 1$, ϵ_i is an independent zero-mean error term (Skrondal, Everitt, 2010). This coefficient is called unit root. If it is significantly different from one, the null hypothesis is not true. Such time series will not return to its original value, so the mean is not constant (Boyle, 2020).

1.9.1 Dickey-Fuller test

The Dickey-Fuller test was the first test in statistics to test the null hypothesis of the presence of a unit root in an autoregressive model of a time series. The test is similar to a simple lag-1 autoregressive model, an extension of a simple unit root test (Boyle, 2020).

Equation 30: **Dickey-Fuller test**
(Source: Processed according to Dolando, Gonzalo, Mayoral, 2002, p. 1969)

$$y_t = \alpha + \theta y_{t-1} + \delta t + \epsilon_t$$

By putting restraints on α and δ we can control if we also test for drift.

The test is carried out under the null hypothesis Y=0 against the alternative hypothesis $Y < 0$.

Equation 31: **Dickey-Fuller statistics**
(Source: Processed according to Dolando, Gonzalo, Mayoral, 2002, p. 1969)

$$DF_\tau = \frac{\hat{Y}}{SE(\hat{Y})}$$

The test returns negative numbers, the more negative it is than the critical value, the stronger evidence for rejecting the null hypothesis.

The critical values table was published in the original paper in 1981 alongside with formula for calculating them.

Equation 32: **Critical values formula**
(Source: Processed according to Dickey, Fuller, 1979)

$$crit = t + \frac{u}{N} + \frac{v}{N^2} + \frac{w}{N^3}$$

Values t, u, v and w are defined for Model 0, Model 1 and Model 2 with over 500 observations and a 1% significance level (Dickey, Fuller, 1979).

Table 1: **Dickey-Fuller test values for critical value calculation**
(Source: Processed according to Dickey, Fuller, 1979)

	t	u	v	w
Model 0	-2.56574	-2.2358	-3.627	0
Model 1	-3.43035	-6.5393	-16.786	-79.433
Model 2	-3.95877	-9.0531	-28.428	-134.155

1.10 White´s homoscedasticity test

White´s test assesses if residuals in regression analysis have constant variance. If the null hypothesis is true, homoscedasticity is present in the data set term. These residuals can also be evaluated graphically (Skrondal, Everitt, 2010).

1.11 Moving average models

Moving average (MA) models use past forecast errors to predict future values and should not be confused with average smoothing or different methods used to conclude the means of a data set. Unlike MA models, these help only to understand the nature of past values (Boyle, 2020).

Equation 33: **MA(q) model**
(Source: Processed according to Boyle, 2020, p. 258)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

This is MA(q) model or moving average model of order q, where:

μ is the mean of the series

the $\theta_1, \dots, \theta_q$ are parameters of the model

the $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ are white noise error terms

the value of q is called the order of the MA model.

In every period of the model y_t is a moving average of several ε-terms, while the mean of the moving average is equal to c. Using past errors in the model, a moving average time series with a trend or seasonality will be first differentiated into time series with stationary covariance. Also, in a MA(q) model value of θ coefficients may differ, giving different weights to each value in the time series. We can use the ACF of lagged residuals to decide what order of q to use in our MA model (Hyndman, Athanasopoulos, 2018).

1.12 Autoregressive model

While regression analysis regresses the variable against one or more independent variables, autoregression plots the variable against itself lagged in time. It forecasts the values using a linear combination of past values. The autoregressive model of a variable can be written as the total of past values of the variable and white noise (Boyle, 2020).

Equation 34: AR(p) model
(Source: Processed according to Boyle, 2020, p. 258)

$$y_t = \mu + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + +b_q y_{t-p} + \varepsilon_t$$

This is the AR(p) model or autoregressive model of order p, where:

μ is the mean of the series

the $b_1, \ldots, b_q$ are parameters of the model

the ε_t is white noise error term

the value of q is called the order of the MA model.

The autoregressive process is stochastic, which means its probability distribution is random. While they can be helpful, valid statistical conclusions may be made from them only if the time series has stationary covariance. This means the expected value and variance of the time series must be constant and finite in all periods, and the covariance of the time series must be constant and finite in all periods with itself over a fixed number of periods.

If the time series we are trying to predict is stationary, autoregressive models can be flexible thanks to the ability to change the parameters $b_1, \ldots, b_q$. Because we use them for stationary time series, we use some constraints for the parameters when estimating them.

For AR(1) model:

When $b_1 = 0, y_t$ is white noise

When $b_1 = 0 \wedge \mu = 0, y_t$ is random walk

When $b_1 = 0 \wedge \mu \neq 0, y_t$ is a random walk with drift

When $b_1 < 0, y_t$ oscillate around mean

Constraints are then:

For AR(1) model: $-1 < b_1 < 1$

For AR(2) model: : $-1 < b_2 < 1 \wedge b_1 + b_2 < 1$

(Hyndman, Athanasopoulos, 2018)

1.13 ARCH model

Autoregressive conditionally heteroscedastic models are used for forecasting volatility with changing variance over time. ARCH models were made for forecasting financial assets, where it was observed that volatility tends to cluster, and variance increases and decreases over time. Hence the parameter used is the proportion of gained or lost variance since the last observation. The ARCH process captures volatility clustering by taking time-varying conditional variance α. So, we can say the ARCH model is modelling the volatility of variance and its time variety (Boyle, 2020).

Based on several studies, overtime rates of return are uncorrelated and characterized by quiet and volatile periods. To capture this dependence conditional mean is used as a constant (Bollerslev, 1987).

Equation 35: Conditional mean
(Source: Processed according to Bollerslev, 1987, p. 543)

$$c_{t|t-1} = \mu + \varepsilon_t$$

ARCH (q) model is where the variance at time *t* is conditional on observations of the previous q times. The output variable of the model is linearly dependable on its previous values and a stochastic term, an error. It works best on high-frequency data.

Equation 36: ARCH (p) conditional variance
(Source: Processed according to Bollerslev, 1987, p. 542-543)

$$h_{t|t-1} = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

Where ω is unconditional variance and α is the ARCH coefficient. So, ARCH (1) conditional variance would be:

Equation 37: ARCH (1) conditional volatility
(Source: Processed according to Bollerslev, 1987, p. 542-543)

$$h_t = \omega + \alpha_1 \epsilon_{t-i}^2$$

When a big shock occurs in period t-1, it is more likely that the value of ε_t in absolute terms will also be more significant. But ARCH constant α_1 must be $\alpha_1 < 1$, otherwise h_t would be explosive and continue to increase over time (Boyle, 2020).

In the long-run, volatility is expected to revert to calmness. This is called volatility means reversion.

Equation 38: ARCH long-run variance

(Source: Processed according to Bollerslev, 1987, p. 542-543)

$$\frac{\omega}{1-\sum \alpha_q}$$

1.14 GARCH model

For the generalized autoregressive conditionally heteroscedastic model, is the variance at time t conditional on values of the past squared observations and past variances to the original model at time t. It is an updated version of the ARCH model (Boyle, 2020).

GARCH (p, q) is a model with q lagged terms of the squared error term and p terms of the lagged conditional variance.

Equation 39: GARCH (p, q)

(Source: Processed according to Bollerslev, 1987, p. 543)

$$h_{t|t-1} = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j|t-1-j}$$

Where ω is unconditional variance and conditional variance h at time t depends on lagged squared error ε_{t-i}^2 and past value of itself $h_{t-j|t-1-j}$.

A conditional mean is used for the calculation of residuals like in the ARCH model (Bollerslev, 1987). GARCH (1,1) model is equal to the ARCH (2) model, as GARCH (p, q) is equivalent to ARCH (p+q) (Boyle, 2020).

For GARCH (1,1) model, the long-run volatility is computed as

Equation 40: GARCH long-run volatility

(Source: Processed according to Bollerslev, 1987, p. 542-543)

$$\frac{\omega}{1-\sum \alpha_q - \sum \beta_p}$$

1.15 Evaluation of models

To compare and pick the best model, each model must be evaluated. Although mathematical modelling aims to prevent overfitting by creating the simplest model, our evaluation methods must consider the number of variables and their contribution to their accuracy.

1.15.1 Coefficient of determination R^2

Coefficient R^2 shows how significant a proportion of variation can our independent variable predict in the dependent variable. It provides a straightforward measurement of how well-observed outcomes are replicated by our model, thus testing the accuracy of our hypothesis. It is the squared value of the correlation between two variables (Skrondal, Everitt, 2010). It usually gives a value between zero and one, where one is a perfect prediction of data using a given regressor, and zero shows no correlation (Barrett, 2012).

Equation 41: Coefficient of determination
(Source: Processed according to Barett, 2012, p. 19)

$$R^2 = 1 - \frac{\sum_i (d_i - y_i)^2}{\sum_i (d_i - \bar{d})^2}$$

Where:

d_i is a vector of observed data points $d_i = [d_1, d_2, \dots, d_n]$

$\bar{d}$ is the mean of a sample

y_i is a vector of predicted data points $y_i = [y_1, y_2, \dots, y_n]$

In numerator vector of residuals is defined as $d_i - y_i = e_i$.

Equation 42: Coefficient of determination
(Source: Processed according to Barett, 2012, p. 19)

$$R^2 = 1 - \frac{Sum\ of\ squared\ residuals}{Total\ sum\ of\ squares}$$

1.15.2 Maximum Likelihood

The likelihood function is the joint probability of observed data as a function of the parameters of our model.

Equation 43: Likelihood function
(Source: Processed according to Skrondal, Everitt, 2010, p. 251)

$$L = \prod_{i=1}^{n} f(x_i, \theta)$$

The likelihood function for each parameter θ is a probability of observed data set X of n observations $x_1, x_2, \dots x_n$ given parameter θ, or $p(X|\theta)$ (Skrondal, Everitt, 2010).

Maximum likelihood estimates a maximum value of the likelihood function given the test parameters. The best estimate of parameter occurs at:

Equation 44: Joint probability distribution of likelihood function for a single observation
(Source: Processed according to Robinson, 2016, p. 12)

$$f_i(\varepsilon_i, \sigma_i, \theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left[-\frac{1}{2}\frac{(\varepsilon_i - \theta)^2}{\sigma_i^2}\right]$$

The log-likelihood is a differentiation of the likelihood function generally used because it is easier to deal with (Skrondal, Everitt, 2010).

Equation 45: Log-likelihood function
(Source: Processed according to Robinson, 2016, p. 13)

$$\hat{L} = \sum_{i=1}^{n} ln\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - \theta)^2}{\sigma_i^2}$$

1.15.2.1 Likelihood ratio

Likelihood ratios can be used to decide if H_0 or H_1 is rejected. Under H_0 the statistic is given by λ.

Equation 46: Likelihood ratio
(Source: Processed according to Skrondal, Everitt, 2010, p. 251)

$$\lambda = 2 - \ln\frac{L_{H_0}}{L_{H_1}}$$

Likelihood ratio statistic λ has an approximately chi-squared distribution. The difference in parameters between the hypothesis is the number of degrees of freedom (Skrondal, Everitt, 2010).

1.15.3 Akaike information criterion

Akaike information criterion is used as an index number to compare several models with a different number of parameters. It is the estimator of prediction error suited for relative comparison of models.

Equation 47: AIC
(Source: Processed according to Skrondal, Everitt, 2010, p. 10)

$$AIC = -2\hat{L} + 2m$$

Where $\hat{L}$ is maximized log-likelihood and m number of parameters in a model. By imposing a penalty for an increasing number of parameters, this index considers the goodness of fit achieved with a particular number of parameters. Lower values indicate a better model (Skrondal, Everitt, 2010).

1.16 Evaluation of model prediction

The final prediction made by the model must always be evaluated based on its accuracy of prediction.

The easiest one is Mean Forecast Error. Mean Forecast Error expresses error as the average of the sum of the difference between observed and predicted values. Positive and negative error terms may influence the result and thus easily distorted.

Mean Absolute Error solves this problem by using the absolute value of the difference:

Equation 48: Mean Absolute Error
(Source: Processed according to Bratu, 2012, p. 24)

$$MAE = \frac{1}{n} \times \sum_{t=1}^{n} |d_t - y_t|$$

While Mean Square error solves the influence of positive and negative error terms by using the square value of the difference:

Equation 49: Mean Absolute Error
(Source: Processed according to Bratu, 2012, p. 23)

$$MSE = \frac{1}{n} \times \sum_{t=1}^{n} (d_t - y_t)^2$$

Because of using square values, a big swing in a data set can influence the value of Mean Square Error. This can be solved by using the Root Mean Square Error

Equation 50: Mean Absolute Error
(Source: Processed according to Bratu, 2012, p. 23)

$$RMSE = \sqrt{MSE}$$

By using square root, we can minimize the impact of big swings in the data set while maintaining the advantage of eliminating the influence of positive and negative values on test results. However, by showing us result in absolute values, it can be hard to compare results of RMSE across tests on different data sets. Mean Absolute Percentage Error goes around this problem by returning the result in percentage (Gustriansyah, Sensuse, Ramadhan, 2017).

Equation 51: MAPE

(Source: Processed according to Gustriansyah, Sensuse, Ramadhan, 2017, p. 716)

$$MAPE = \frac{1}{n} \times \sum_{t=1}^{n} \left| \frac{d_t - y_t}{d_t} \right|$$

Because MAPE returns the value in percentage, the result can be compared across different data sets and critical values sets (Gustriansyah, Sensuse, Ramadhan, 2017).

Table 2: MAPE values for prediction evaluation

(Source: Processed according to Gustriansyah, Sensuse, Ramadhan, 2017, p. 716)

MAPE value	Prediction accuracy
MAPE ≤ 10%	High
10% < MAPE ≤ 20%	Good
20% < MAPE ≤ 50%	Reasonable
MAPE > 50%	Low

2 ANALYTICAL PART

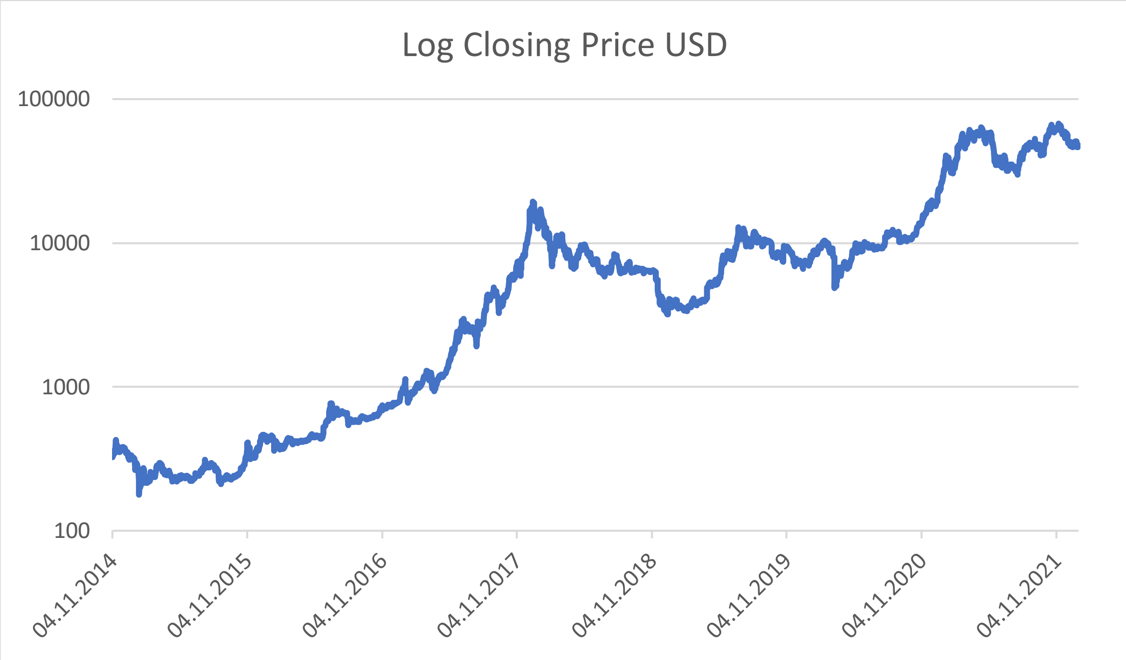
In this section of the thesis are econometrics models constructed and evaluated. First, Bitcoin´s daily return time series is tested for autocorrelation, normality and heteroscedasticity. Then, based on these tests are constraints of models proposed. The critical component of econometrics modelling is verification. The whole modelling process was done in software Microsoft Excel with the help of statistical software Gretl.

2.1.1 Data

We are using daily data from the database of coinase.com from 04. 11. 2014 to 02. 01. 2022. Because Bitcoin is exchanged on deregulated markets, it is traded nonstop, 24 hours a day. To change this continuous time series into discrete time series, we use the closing price at 0:00 GMT+1 time denominated in US dollars.

Graph 1: Bitcoin/USD linear chart

(Source: Own processing in software Microsoft Excel)



Graph 2: Bitcoin/USD logarithmic chart
(Source: Own processing in software Microsoft Excel)

As stated in the theoretical part of the thesis, we can spot stationarity visually by the existence of a trend. This time series has a trend, and its mean is not constant. For the successful construction of chosen models, we need stationary time series.

2.1.2 First differencing

Because of visible trend and cycle in this time series, we will not be using closing price, but its daily change expressed in percentage.

Equation 52: First differencing in Excel
(Source: Processed according to the equation n. 24)

$$d_t = \frac{P_t}{P_{t-1}} - 1$$

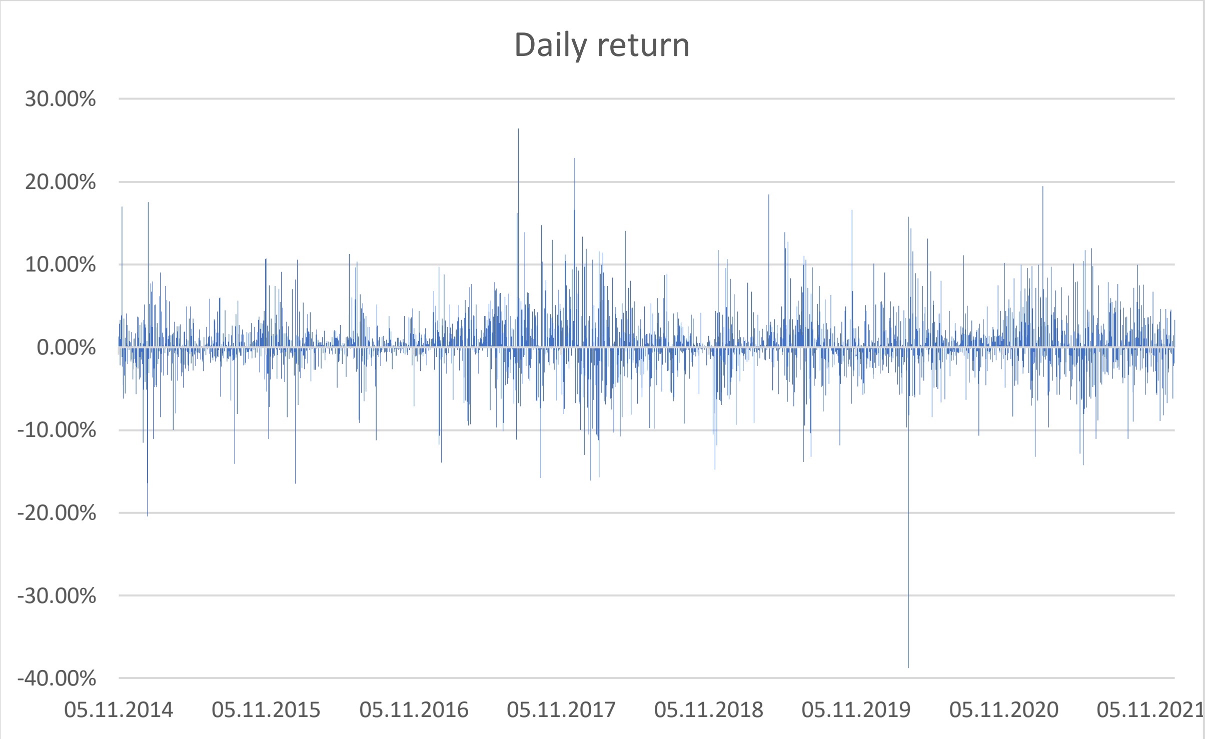
Where:

P_t is the closing price on day one

P_{t-1} the closing price of the previous day

d_t our day-to-day change

We include -1 and change the format of cells to percentage to calculate day to day percentages return.



Graph 3: Daily return of BTC/USD

(Source: Own processing in software Microsoft Excel)

After first differencing, we have 2616 observations. This table shows descriptive statistics compared to the index S&P 500 that tracks the performance of the 500 biggest firms on the US market.

Table 3: Descriptive statistics for Bitcoin and S&P 500

(Source: Own processing in software Gretl)

	Bitcoin	S&P 500
Mean	0.27%	0.05%
Median	0.20%	0.07%
Minimum	-38.76%	-11.98%
Maximum	26.42%	9.38%
Standard deviation	0.039495	0.011244
Variance coefficient	14.64	0.575
Skewness	-0.17268	0.73115
Standard skewness	7.3783	0.567
5% percentile	-0.06106	0.016243
95% percentile	0.065689	0.014844
Interquartile range	0.032678	0.008436
Number of valid observations	2616	1802
Missing observations	0	66

Bitcoin's mean daily return of 0.27% is enormous compared to 0.05% of S&P500, which is considered a market benchmark. Also, with three times bigger maximum, minimum and standard deviation, Bitcoin returns have great volatility compared to traditional markets. This can influence the accuracy of our models.

2.2 Dickey-Fuller test

First tested is stationarity to ensure the data set is stationary after first differencing. Dickey-Fuller test is used for that. The null hypothesis of the presence of unit root is tested.

For the Dickey-Fuller test, we calculate the difference e_t of each data point d in time t and time $t-1$. This difference e_t is the dependent variable. Model 0 of the Dickey-Fuller test without drift is calculated by using lagged d_t as the independent variable in simple linear regression. Linear regression in Microsoft Excel is calculated with function =LINEST() (Microsoft Excel 365, 2022). Including 0 for no constant and 1 for additional regression statistics.

Equation 53: Model 0 of the Dickey-Fuller test in Excel
(Source: Processed according to equation n. 30)

$$\{= \text{LINEST}\big(\, e_{t_3}\!: e_{t_{2613}}; d_{t_3}\!: d_{t_{2613}}; 0; 1\big)\}$$

The same function is used for Model 1 with drift, setting =LINEST() function to 1 after the independent variable allows for constant to be calculated.

Equation 54: Model 1 of the Dickey-Fuller test in Excel
(Source: Processed according to equation n. 30)

$$\{= \text{LINEST}\big(\, z_{t_3}\!: z_{t_{2613}}; d_{t_3}\!: d_{t_{2613}}; 1; 1\big)\}$$

Model 2 is with drift and time trend. Each lagged d_t value is ranked from 1 to 2615 using Microsoft Excel =RANK() function (Microsoft Excel 365, 2022). Rank is used in linear regression as an additional regressor.

Equation 55: Model 2 of the Dickey-Fuller test in Excel
(Source: Processed according to equation n. 30)

$$\{= \text{LINEST}\big(\, z_{t_3}\!: z_{t_{2613}}; d_{t_3}\!: d_{t_{2613}}; Rank_{t_3}\!: Rank_{t_{2613}}; 1; 1\big)\}$$

Table 4: **Dickey-Fuller test results**
(Source: Own processing in software Microsoft Excel)

			Constant	Trend
Model 0	Coefficient	-1.04412		
	SE	0.019542		
	t-stat	-53.42917		
Model 1	Coefficient	-1.04897	0.0028254	
	SE	0.01954	0.0007735	
	t-stat	-53.67961		
Model 2	Coefficient	-1.04900	0.0000003	0.0024030
	SE	0.01954	0.0000010	0.0015448
	t-stat	-53.6713		

Coefficients and standard errors are calculated by the =LINEST() function, and t-stat is calculated according to equation 31.

Equation 56: **Dickey-Fuller test**
(Source: Processed according to the equation 31)

$$t - stat = \frac{coefficient}{standart\ error}$$

With the number of observations of *n*=2616, critical values of t-stat with a 1% significance level are calculated according to equation number 32 and table number 1.

Comparing results with the critical values, the null hypothesis is rejected because the more negative values our test gives us, the more substantial evidence there is for rejecting the presence of unit root.

Table 5: **Dickey-Fuller test results compared to critical values**
(Source: Own processing in software Microsoft Excel)

	Critical value	Test value
Model 0	-2.56660	-53.42917
Model 1	-3.43285	-53.67961
Model 2	-3.96223	-53.67128

Based on this test, it can be safely assumed that Bitcoin's daily return is a stationary time series.

2.3 Normality

For proper function of most models, the data set should follow normal distribution at least approximately. If normality is not proven for Bitcoin´s daily return time series, chosen models may not be accurate. Data outliners, or data points deviating from observations, are often the cause of non-normality.

2.3.1 Q-Q plot

To create a Q-Q plot, we rank each data point from 1 to 2616 using the Excel function =RANK.AVG() (Microsoft Excel 365, 2022). Then percentile for each data point can be calculated.

Equation 57: Percentile calculation
(Source: Processed according to Microsoft Excel 365)

$$Percentile_t = \frac{Rank_t - 0.5}{COUNTIF(Rank_1 : Rank_{2616}; >0)}$$

A theoretical z-score, given by equation 11, is needed to compare data. Excel function =NORM.S.INV() is used to return inverse function to standard distribution with mean 0 and unit variance (Microsoft Excel 365, 2022).

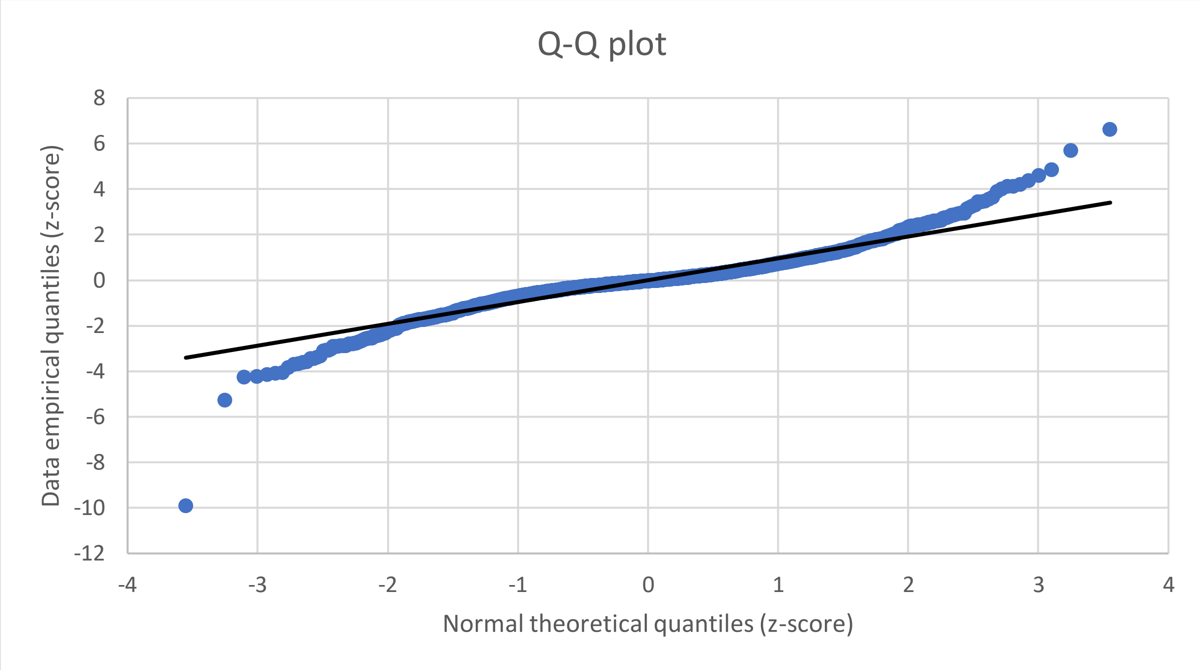
Equation 58: Theoretical z-score
(Source: Processed according to Skrondal, Everitt, 2010, p. 410)

$$theoretical\ z - score_t = NORM.S.INV(Percentile_t)$$

If done correctly, this theoretical z-score should appear as a straight line in the plot. Normality can be proven or rejected by comparing this line to z-scores of observed data points. To calculate empirical z-score function =STANDARDIZE() is used, which gives back a normalized value with a normal distribution defined by the average and standard deviation of the sample. Function =AVARAGEA() calculates sample mean, =STDEVA() sample standard deviation. (Microsoft Excel 365, 2022).

Equation 59: Empirical z-score
(Source: Processed according to Skrondal, Everitt, 2010, p. 150)

$$emp.z - score_t = STANDARDIZE(d_t; AVERAGEA(d_1 : d_{2616}); STDEVA(d_1 : d_{2616}))$$



Graph 4: Q-Q plot of Bitcoin daily volatility

(Source: Own processing in software Microsoft Excel 365)

There is a deviation in the data set compared to the straight line that represents the normal distribution, proving visually normal distribution is not precisely followed in Bitcoin´s daily return time series.

2.3.2 Frequency chart

The most extreme daily returns in our time series are +26.42% and -38.76%, so 21 bins from -20% to 20% were chosen to sort our data into a frequency table. To get our frequency matrix, we use =FREQUENCY() Excel function (Microsoft Excel, 2022).

Equation 60: Frequency matrix

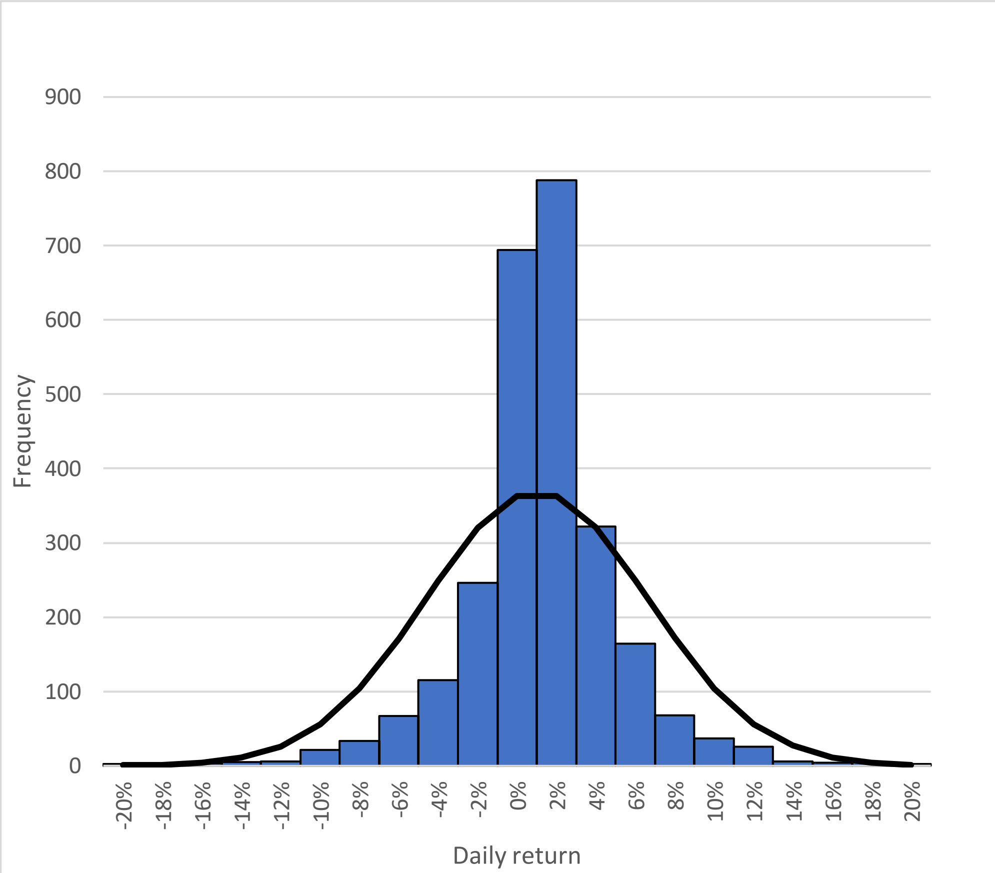
(Source: Processed according to Microsoft Excel 365)

$$\{= FREQUENCY(d_1 : d_{2616}; Bin_1 : Bin_{21})\}$$

Table 6: Frequency table of Bitcoin daily volatility

(Source: Own processing in software Microsoft Excel 365)

Bins	Frequency
-20%	2
-18%	0
-16%	3
-14%	5
-12%	6
-10%	21
-8%	33
-6%	67
-4%	115
-2%	246
0%	694
2%	788
4%	322
6%	164
8%	68
10%	37
12%	26
14%	6
16%	4
18%	5
20%	2



Graph 5: Frequency table of Bitcoin daily volatility

(Source: Own processing in software Microsoft Excel 365)

Visualization of the normal distribution is based on a theoretical z-score according to equation 58. By comparing results to the black line that represents normal distribution with the same number of data points, findings assumed from Q-Q plot are proven.

2.3.3 Kolmogorov-Smirnov test

To calculate K-S test in Microsoft Excel 365, returns from lowest to highest must be ranked 1 to maximum observation using =RANK.AVG() function like in Q-Q Plot.

The empirical distribution is based on our ranks and is calculated as:

Equation 61: Empirical distribution

(Source: Processed according to equation 26)

$$Empirical\ distribution_k = \frac{Rank_k}{number\ of\ observation}$$

Function =NORM.DIST() can be used to calculate theoretical CDF if set to *TRUE* (Microsoft Excel 365, 2020).

Equation 62: Theoretical CDF

(Sourced: Processed according to the equation 11)

$$Theoretical\ CDF_k = NORM.DIST(d_k; AVERAGEA(D_K); STDEVA(D_K); TRUE)$$

Then the absolute difference in empirical and theoretical distribution is calculated using the =ABS() function. This Excel function returns absolute value (Microsoft Excel 365, 2020).

Equation 63: KS test

(Source: Processed according to the equation 27)

$$Supremum = ABS(empirical\ CDF_k - theoretical\ CDF_k)$$

To find supremum, maximum deviation in difference has to be found using =MAX() function, which returns the maximum value (Microsoft Excel 365, 2020).

Equation 64: KS statistics

(Source: Processed according to the equation 27)

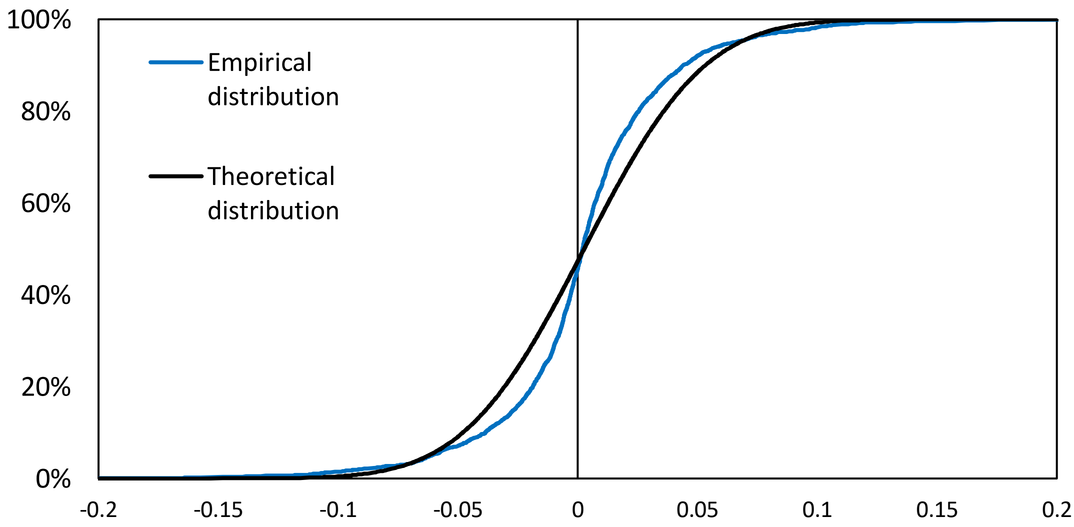
$$K - S = number\ of\ observation * \sqrt{supremum}$$

Table 7: Kolmogor-Smornov test results

(Source: Own processing in software Microsoft Excel 365)

Supremum	9.60%
Kolmogorov-Smirnov statistic	4.912
p-value	0.00

We can reject the null hypothesis of Bitcoin daily returns following normal distribution based on p-value < 0.05. This can affect the accurency of models used, but they can still be calculated.



Graph 6: Kolmogorov-Smirnov test

(Source: Own processing in software Microsoft Excel 365)

2.4 Autocorrelation tests

High autocorrelation of the data set can influence the estimation of coefficients in models' calculations. Especially for a model with a low number of coefficients, the distribution of residuals might not be captured correctly.

2.4.1 Durbin-Watson test

Residuals *e* of daily returns can be calculated by deducting the average return. The difference is then calculated as residual *e* in time *t* minus residual in time *t-1*.

To calculates the sum of squared residuals, excel function =SUMSQ() is used (Microsoft Excel 365, 2022).

Equation 65: Sum squared difference
(Source: Processed according to equation 28)

$$\sum_{t=2}^{T} (e_t - e_{t-1})^2 = SUMSQ(Difference_1 : Difference_{2615})$$

Equation 66: Sum squared residuals
(Source: Processed according to equation 28)

$$\sum_{t=1}^{T} e_t^2 = SUMSQ(e_1 : e_{2616})$$

Durbin-Wattson test is then:

Equation 67: Durbin-Watson test
(Source: Processed according to equation 28)

$$= \frac{sum\ squared\ differences}{sum\ squared\ residuals}$$

Table 8: Durbin-Watson test results
(Source: Own processing in software Microsoft Excel 365)

sum squared difference	8.56
sum squared residual	4.08
Durbin-Wattson statistic	2.10

Durbin-Watson tests give a number from interval <0;4> where values around 2 are considered weak or no autocorrelation. Hence autocorrelation should not have an effect on the calculation of parameters in used models.

2.4.2 ACF

From the Durbin-Watson test, it can be suspected that the daily return of Bitcoin price does not have autocorrelation. The autocorrelation function is used to visualize how quickly autocorrelation drops with each lag. Based on ACF number of significant lags for models are chosen.

Microsoft Excel function =SUMPRODUCT() calculates scalar sum between more variables (Microsoft Excel 365, 2022).

Equation 68: Numerator of ACF
(Source: Processed according to the equation 22)

$$\sum_{i=1}^{n-k}(x_t-\bar{x})(x_{t+k}-\bar{x})$$

The numerator in ACF calculation is equal to

Equation 69: Numerator of ACF
(Source: Processed according to the equation 22)

$$= SUMPRODUCT(d_t{:}\,d_{max}-AVERAGE(D_T);d_{t+k}{:}\,d_{max}-AVERAGE(D_T))$$

Excel function =DEVSQ() returns the sum of squares of residuals (Microsoft Excel 365, 2022). That means

Equation 70: Denominator of ACF
(Source: Processed according to the equation 22)

$$\sum_{i=1}^{n-k}(x_t-\bar{x})^2 = DEVSQ(d_1{:}\,d_{2616})$$

Equation 71: ACF
(Source: Processed according to the equation 22)

$$ACF = \frac{SUMPRODUCT(d_t{:}\,d_{max}-AVERAGE(D_T);d_{t+k}{:}\,d_{max}-AVERAGE(D_T))}{DEVSQ(d_1{:}\,d_{2616})}$$

Critical values are calculated as:

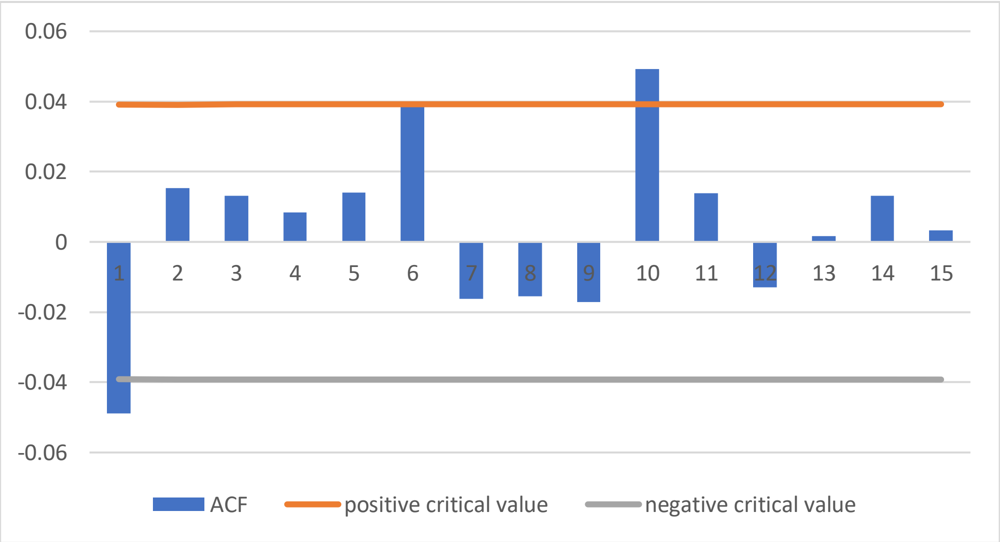
Equation 72: ACF critical values
(Source: Processed according to the equation 23)

$$\pm critical\ value = \pm\frac{2}{\sqrt{number\ of\ observation-number\ of\ lags}}$$

Table 9: **ACF table of Bitcoin daily volatility for 15 lags**

(Source: Own processing in software Microsoft Excel 365)

Lag	ACF	positive critical value	negative critical value
1	-0.049	0.039	-0.039
2	0.015	0.039	-0.039
3	0.013	0.039	-0.039
4	0.008	0.039	-0.039
5	0.014	0.039	-0.039
6	0.039	0.039	-0.039
7	-0.016	0.039	-0.039
8	-0.015	0.039	-0.039
9	-0.017	0.039	-0.039
10	0.049	0.039	-0.039
11	0.014	0.039	-0.039
12	-0.013	0.039	-0.039
13	0.002	0.039	-0.039
14	0.013	0.039	-0.039
15	0.003	0.039	-0.039



Graph 7: **ACF for 15 lags**

(Source: Own processing in software Microsoft Excel 365)

Lag 1, 6 and 10 show slight statistical significance. Based on that, up to 10 lags are included in all the models.

2.5 White´s homoscedasticity test

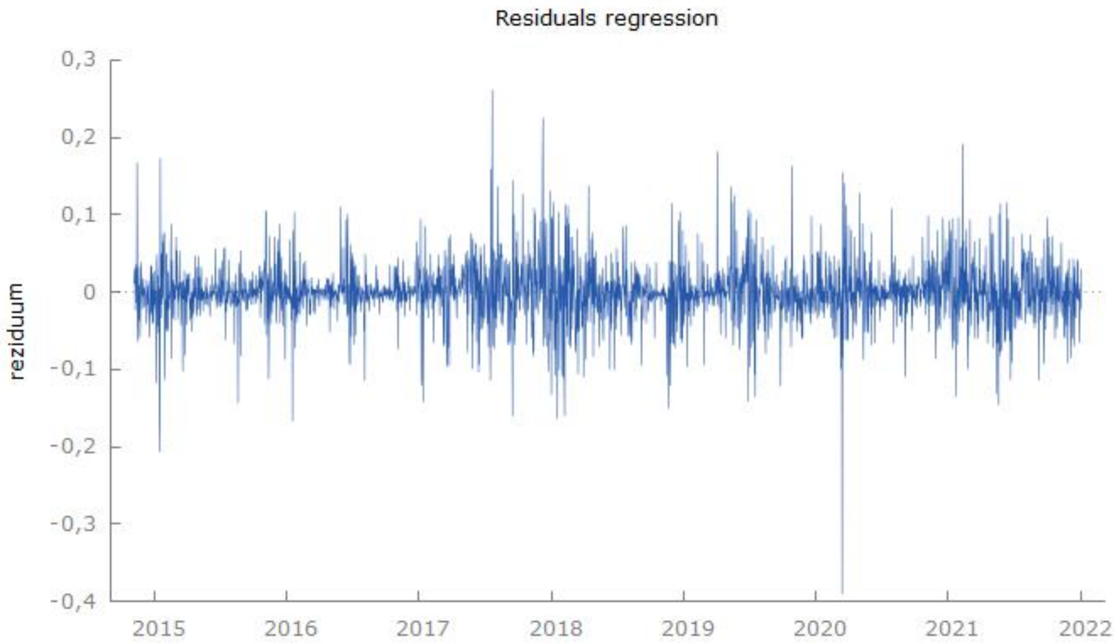
The null hypothesis for this test is that residuals in regression have constant variance. This test is calculated in software Gretl by using Bitcoin´s daily returns as dependent and time trend as an independent.

Table 10: White´s test results

(Source: Own processing in software Gretl and Microsoft Excel 365)

	coefficient	standard error	p-value
constant	0.0009	0.0003	0.0120
time	0.1121	0.0494	0.0235
sq_time	0.0000	0.0000	0.0514
coefficient of determination	0.0023		
p-value	0.0473		

P-value is barely under 0.05, the null hypothesis is rejected, and heteroscedasticity is present in Bitcoin´s daily returns. Graphical analysis confirms changing variance in residuals over time.



Picture 12: Development of residuals in time

(Source: Own processing in software Gretl)

This limits the accuracy of chosen models, especially the MA and AR models, because ARCH and GARCH are designed to predict heteroscedastic time series.

2.6 Moving average model

As the MA model uses past forecast errors for prediction MA (0) model is the first one calculated, which would be simply the mean of the Bitcoin time series.

That way, the error term is $\epsilon_{MA(0)t} = d_t - \mu$. Coefficients and standard errors for MA (1) model can be calculated using =LINEST() excel function (Microsoft Excel 365, 2022).

Including 1 after dependent matrix to calculate constant, 1 for additional regression statistics, in our case, standard error.

Equation 73: MA (1) initial parameters calculation

(Source: Processed according to the equation 33)

$$\{= LINEST(d_2{:}\, d_{2616}; \epsilon_{MA(0)1}{:}\, \epsilon_{MA(0)2615}; 1; 1)\}$$

Result is constant c_1 and parameter θ_1. MA (1) error term is based on MA (0) model.

Equation 74: MA (1) error term

(Source: Processed according to the equation 33)

$$\epsilon_{MA(1)t} = d_t - (c_1 + \theta_1 * \epsilon_{MA(0)t-1})$$

Equation 75: MA (1) parameters calculation

(Source: Processed according to the equation 33)

$$\{= LINEST(d_2{:}\, d_{2616}; \epsilon_{MA(1)1}{:}\, \epsilon_{MA(1)2615}; 1; 1)\}$$

The concept of iteration is used for this process until the model converges and the constant and parameter will no longer change at five decimal places. Each iteration calculates the error term based on parameters from the previous iteration.

Table 11: Iteration table of MA (1) model

(Source: Own processing in software Microsoft Excel 365)

	Iteration 2		Iteration 3		Iteration 4	
	θ_1	c_1	θ_1	c_1	θ_1	c_1
Coefficient	-0.04833	0.00269	-0.04830	0.00269	-0.04830	0.00269
Standard error	0.01957	0.00077	0.01956	0.00077	0.01956	0.00077
t-stat	-2.4701	3.4903	-2.4689	3.4903	-2.4689	3.4903
p-value	1.36%	0.05%	1.36%	0.05%	1.36%	0.05%

Model converges on iteration 3, and coefficients do not change on five decimal places anymore. Student´s t-distribution is used when evaluating the accuracy of parameters because of Bayesian inference.

Equation 76: MA (q) parameters p-value

(Source: Processed according to the equation 12)

$$p - value = T.DIST.2T(ABS(t - stat); number\ of\ observations - number\ of\ coefficients)$$

Excel function =T.DIST.2T() function calculates two-tailed Student's t distribution (Microsoft Excel 365, 2022).

For MA(2):

Equation 77: MA (2) initial parameters calculation

(Source: Processed according to the equation 36)

$$\{= LINEST(d_3: d_{2616}; \epsilon_{MA(0)2}: \epsilon_{MA(0)2615}: \epsilon_{MA(0)1}: \epsilon_{MA(0)2614}; 1; 1)\}$$

Equation 78: MA (2) error term

(Source: Processed according to the equation 36)

$$\epsilon_{MA(2)t} = d_t - \left(c_2 + \theta_1 * \epsilon_{MA(0)t-1} + \theta_2 * \epsilon_{MA(0)t-2}\right)$$

Equation 79: MA (2) parameters calculation

(Source: Processed according to the equation 36)

$$\{= LINEST(d_3: d_{2616}; \epsilon_{MA(2)2}: \epsilon_{MA(2)2615}: \epsilon_{MA(2)1}: \epsilon_{MA(2)2614} \; ; 1; 1)\}$$

Table 12: Iteration 4 of MA (2) model

(Source: Own processing in software Microsoft Excel 365)

	Iteration 4		
	θ_2	θ_1	c_2
Coefficient	0.01597	-0.04860	0.00268
Standard error	0.01957	0.01957	0.00077
t-stat	0.8162	-2.4833	3.4766
p-value	41.45%	1.31%	0.05%

For MA (3):

Equation 80: MA (3) initial parameters calculation

(Source: Processed according to the equation 36)

$$\{LINEST(d_4: d_{2616}; \epsilon_{MA(0)3}: \epsilon_{MA(0)2615}: \epsilon_{MA(0)2}: \epsilon_{MA(0)2614}: \epsilon_{MA(0)1}: \epsilon_{MA(0)2613}; 1; 1)\}$$

Equation 81: MA (3) error term

(Source: Processed according to the equation 36)

$$\epsilon_{MA(3)t} = d_t - \left(c_3 + \theta_1 * \epsilon_{MA(0)t-1} + \theta_2 * \epsilon_{MA(0)t-2} + \theta_3 * \epsilon_{MA(0)t-3}\right)$$

Equation 82: MA (3) parameters calculation

(Source: Processed according to the equation 36)

$$\{LINEST(d_4: d_{2616}; \epsilon_{MA(3)3}: \epsilon_{MA(2)2615}: \epsilon_{MA(3)2}: \epsilon_{MA(2)2614}: \epsilon_{MA(3)1}: \epsilon_{MA(3)2613}; 1; 1)\}$$

Table 13: **Iteration 4 of MA (3) model**

(Source: Own processing in software Microsoft Excel 365)

	Iteration 4			
	θ_3	θ_2	θ_1	c_3
Coefficient	0.01280	0.01567	-0.04892	0.002672
Standard error	0.01958	0.01958	0.01958	0.000772
t-stat	0.6537	0.8004	-2.4987	3.460026
p-value	51.34%	42.36%	1.25%	0.05%

All orders can be done like this, but it can take up to 6 iterations before stabilizing coefficients. A table of all coefficients and p-values can be found in Annex I.

To evaluate if log-likelihood is calculated using more lags in the MA model. Excel function =LN() calculates natural logarithm (Microsoft Excel 365, 2022).

Equation 83: **Log-likelihood for MA (q) models**

(Source: Processed according to the equation 45)

$$\hat{L} = SUM(LN(NORM.DIST(\epsilon_{MA(q)1}:\epsilon_{MA(q)2616};0;\sigma_{\epsilon_{MA(q)T}};0)))$$

Akaike information criterion is calculated according to equation 47, likelihood ratio based on equation 46. P-value is calculated using Excel function =CHISQ.DIST.RT() that returns a right tail probability of chi-squared distribution with degrees of freedom equal to extra coefficient used in comparison to the previous model (Microsoft Excel 365, 2022).

Equation 84: **P-value for log-likelihood**

(Source: Processed according to the equation 13)

$$p - value = \text{CHISQ.DIST.RT}(\hat{L}\text{ ratio; degrees of freedom})$$

Table 14: **Evaluation of MA (1-10) models**

(Sourced: Own processing in software Microsoft Excel 365)

	$\hat{L}$	AIC	λ	p-value
MA (1)	4745.43	-9486.86		
MA (2)	4745.79	-9485.58	0.72	39.70%
MA (3)	4746.01	-9484.01	0.44	50.92%
MA (4)	4746.12	-9482.25	0.24	62.68%
MA (5)	4746.56	-9481.12	0.88	34.95%
MA (6)	4748.46	-9482.92	3.79	5.14%
MA (7)	4748.91	-9481.82	0.91	34.08%
MA (8)	4749.40	-9480.80	0.97	32.38%
MA (9)	4749.51	-9479.01	0.22	64.19%
MA (10)	4753.11	-9484.21	7.20	0.73%

Because AIC estimates the amount of information lost, the less information model loses, the better. MA (1) model looks most efficient based on AIC. We can see a steady decline in accuracy for more orders only to increase for MA (10) model. Loglikelihood is the biggest for MA (10) model. The low p-value of the MA (6) model is caused by the terribleness of the previous MA (5) model.

Now all coefficients can be applied to calculate all MA(p) models based on equation 33. The accuracy of the final models is evaluated using MAE, MSE, and RMSE based on the equations 48, 49 and 50. Because these evaluations are used only for comparison, it tells which of the compared model is best, not if they are accurate. Values of all these evaluations across all the models are similar. For this reason, MAPE is calculated according to equation 51, and the results are compared to critical values in table 2. As predicted based on log-likelihood and AIC, MA (1) model is best with a MAPE of 138.619%. But because this is still way above the MAPE threshold of 50%, this model accuracy is low.

Table 15: MA (q) models prediction evaluation

(Source: Own processing in software Microsoft Excel 365)

	MAE	MSE	RMSE	MAPE
MA (1)	2.601%	0.155%	2.601%	138.619%
MA (2)	2.603%	0.155%	2.603%	140.882%
MA (3)	2.602%	0.155%	2.602%	140.450%
MA (4)	2.602%	0.155%	2.602%	141.593%
MA (5)	2.603%	0.155%	2.603%	141.564%
MA (6)	2.602%	0.155%	2.602%	143.469%
MA (7)	2.604%	0.155%	2.604%	143.213%
MA (8)	2.605%	0.154%	2.605%	143.120%
MA (9)	2.605%	0.154%	2.605%	142.992%
MA (10)	2.610%	0.154%	2.610%	147.520%

2.7 Autoregression model

Unlike MA (p) model, AR (q) model uses past values for prediction. Because the sample size is large enough, the first ten observations are not included because autoregression is done for ten lags. So, the data set has the same size for all the model variations.

Simple linear regression is used to obtain coefficient and parameters, regressing data points against their lagged values.

Equation 85: Linear regression for AR (1) coefficients
(Source: Processed according to the equation 3)

$$= \{LINEST(d_{10}: d_{2616}; d_9: d_{2615}; 1; 1)\}$$

For AR (2) an additional lag is used as a regressor.

Equation 86: Linear regression for AR (2) coefficients
(Source: Processed according to the equation 8)

$$= \{LINEST(d_{10}: d_{2616}; d_9: d_{2615}: d_8: d_{2614}\ ; 1; 1)\}$$

For AR (3)

Equation 87: Linear regression for AR (3) coefficients
(Source: Processed according to the equation 8)

$$= \{LINEST(d_{10}: d_{2616}; d_9: d_{2615}: d_8: d_{2614}: d_7: d_{2613}\ \ ; 1; 1)\}$$

And so on for all orders.

With all of the parameters calculated, all AR (p) can be evaluated. The coefficient of determination is calculated with the =LINEST() function alongside parameters (Microsoft Excel 365, 2022).

Equation 88: F-stat
(Source: Processed according to the equation 25)

$$F - stat_{AR\ (p)} = \frac{\frac{R^2}{p}}{\frac{(1 - R^2)}{n - m - p}}$$

Where m is the number of degrees of freedom. Excel function =F.DIST.RT() calculates the p-value of the F-stat (Microsoft Excel 365, 2022).

Equation 89: F-stat p-value
(Source: Processed according to Skrondal, Everitt, 2010, p. 286)

$$p - value_{AR\ (p)} = \mathrm{F.DIST.RT(F - stat; p; n - m - p)}$$

Log-likelihood, AIC, log-likelihood ratio, and p-value calculation are the same as MA (q) models. A table of all coefficients can be found in Annex II.

Table 16: **Evaluation of AR (1-10) models**

(Source: Own processing in software Microsoft Excel 365)

	R^2	F-stat	p-value	$\hat{L}$	AIC	λ	p-value
AR (1)	0.0023	6.00	1.44%	10171.83	-20339.67		
AR (2)	0.0024	3.18	4.16%	10250.74	-20495.47	157.81	0.00
AR (3)	0.0026	2.28	7.79%	10314.65	-20621.30	127.83	0.00
AR (4)	0.0027	1.76	13.50%	10339.54	-20669.07	49.77	0.00
AR (5)	0.0029	1.53	17.71%	10402.83	-20793.67	126.59	0.00
AR (6)	0.0045	1.94	7.04%	10507.02	-21000.03	208.37	0.00
AR (7)	0.0047	1.74	9.56%	10504.60	-20993.19	-4.84	N/D
AR (8)	0.0050	1.64	10.88%	10495.18	-20972.36	-18.84	N/D
AR (9)	0.0054	1.57	11.82%	10479.13	-20938.26	-32.10	N/D
AR (10)	0.0077	2.01	2.87%	10333.11	-20644.22	-292.03	N/D

Looking at R^2 table, no order shows robust regression. However, given the big sample size and nature of our time series, it is still a good comparison between models. Biggest R^2 of 0.0077 has an order of 10 with a significant p-value of 2.87%. AR (1) p-value of 1.44% shows significance as well. On the other hand, log-likelihood has a maximum at AR (6) and then decreases. AIC values confirm that AR (6) has the least amount of information lost. Adding more lags as an explanatory variable might offer a better coefficient of determination but not a better model.

Now prediction for each AR(p) model can be calculated based on equation 34.

Table 17: **AR (p) models prediction evaluation**

(Source: Own processing in software Microsoft Excel 365)

	MAE	MSE	RMSE	MAPE
AR (1)	2.601%	0.155%	2.601%	139.255%
AR (2)	2.617%	0.156%	2.617%	146.446%
AR (3)	2.625%	0.156%	2.625%	152.316%
AR (4)	2.620%	0.156%	2.620%	145.731%
AR (5)	2.618%	0.156%	2.618%	148.346%
AR (6)	2.601%	0.155%	2.601%	139.753%
AR (7)	2.603%	0.155%	2.603%	140.433%
AR (8)	2.604%	0.154%	2.604%	141.138%
AR (9)	2.604%	0.154%	2.604%	144.260%
AR (10)	2.609%	0.154%	2.609%	145.982%

According to MAE, MSE, RMSE, and MAPE AR (1) model offers the best prediction, followed up by AR (6) model. Non the less, the accuracy prediction is still low, with MAPE for AR (1) of 139.225% exceeding the threshold of 50%.

2.8 ARCH

Unconditional variance ω needed as the coefficient for the ARCH (p) model is first calculated as a variance of Bitcoin daily volatility with =VAR.S() excel function (Microsoft Excel 365, 2022).

Residuals are calculated with conditional mean c.

Equation 90: ARCH residuals
(Sourced: Processed according to the equation 35)

$$\varepsilon_t = d_t - c$$

Then conditional variance for the first ten observed values is equal to the long-run variance because prediction based on previous values is not yet available.

Equation 91: ARCH long-run volatility
(Sourced: Processed according to the equation 38)

$$h_{1-10}^2 = Long\ run\ varinace = \frac{\omega}{1 - SUM(\alpha_1 : \alpha_{10})}$$

Conditional variance for the rest of the observation from the 11^{th} observation:

Equation 92: ARCH conditional variance
(Sourced: Processed according to the equation 36)

$$h_t^2 = \omega + SUMPRODUCT(\alpha_1 : \alpha_{10}; \varepsilon_t^2 : \varepsilon_{t-10}^2)$$

Including all α coefficients is used in order to use the same number of observations for all ARCH(p) models and their log-likelihood function.

Equation 93: ARCH (p) log-likelihood for a single observation
(Sourced: Processed according to the equation 45)

$$L_t(c, \omega, \alpha) = LN\left(\frac{1}{\sqrt{2\pi h_t^2}} exp\left[-\frac{\varepsilon_t^2}{2h_t^2}\right]\right)$$

Equation 94: ARCH (p) log-likelihood function
(Sourced: Processed according to the equation 45)

$$\hat{L} = SUM(L_{10} : L_{2616})$$

Because the first ten observations have conditional variance equal to long-run variance, they are not included in maximizing the log-likelihood function.

ARCH (0) has $\alpha_1 = 0$, so this model is already created.

Table 18: ARCH (0) coefficients
(Source: Own processing in software Microsoft Excel 365)

c	0.27%
ω	0.001560
Long-run variance	0.001560
$\hat{L}$	4729.99
α_1	0.00

To maximize the log-likelihood function for ARCH (1) Excel solver tool is used. Setting solver to maximize the function using Gradient descent with variables μ, ω, α_1.

Table 19: ARCH (1) coefficients

(Source: Own processing in software Microsoft Excel 365)

c	0.27%
ω	0.001322
Long-run variance	0.001540
$\hat{L}$	4784.62
α_1	0.14

For ARCH (2) solver is set to maximize the function using Gradient descent with variables $\mu, \omega, \alpha_1, \alpha_2$.

Table 20: ARCH (2) coefficients

(Source: Own processing in software Microsoft Excel 365)

c	0.29%
ω	0.001171
Long-run variance	0.001571
$\hat{L}$	4802.58
α_1	0.12
α_2	0.13

For ARCH (3), variables are $\mu, \omega, \alpha_1, \alpha_2, \alpha_3$. All lags up to ARCH (10) are calculated like this. AIC, log-likelihood ratio and the p-value are calculated like for MA (q) and AR (p) models. All coefficients can be found in Annex III.

Table 21: ARCH (q) model evaluation

(Source: Own processing in software Microsoft Excel 365)

	$\hat{L}$	AIC	λ	p-value
ARCH (1)	4784.62	-9563.24	109.26	0.00%
ARCH (2)	4802.58	-9597.15	35.91	0.00%
ARCH (3)	4824.93	-9639.85	44.70	0.00%
ARCH (4)	4888.15	-9764.30	126.44	0.00%
ARCH (5)	4922.75	-9831.50	69.20	0.00%
ARCH (6)	4942.74	-9869.49	39.99	0.00%
ARCH (7)	4946.92	-9875.84	8.35	0.39%
ARCH (8)	4964.54	-9909.08	35.24	0.00%
ARCH (9)	4964.88	-9907.76	0.68	41.00%
ARCH (10)	4969.28	-9914.56	8.80	0.30%

Unlike with moving average and autoregression model, the ARCH model offers higher log-likelihood and lower AIC with each additional parameter added. ARCH (10) has the biggest log-likelihood while maintaining the lowest AIC.

Now prediction is calculated based on equation 36 and evaluated.

Table 22: **ARCH (q) models prediction evaluation**

(Source: Own processing in software Microsoft Excel 365)

	MAE	MSE	RMSE	MAPE
ARCH (1)	4.363%	0.290%	4.363%	1073.537%
ARCH (2)	4.346%	0.292%	4.346%	1064.891%
ARCH (3)	4.323%	0.293%	4.323%	1050.389%
ARCH (4)	4.298%	0.299%	4.298%	1002.386%
ARCH (5)	4.283%	0.302%	4.283%	966.962%
ARCH (6)	4.267%	0.302%	4.267%	944.176%
ARCH (7)	4.262%	0.302%	4.262%	937.501%
ARCH (8)	4.265%	0.305%	4.265%	932.330%
ARCH (9)	4.265%	0.305%	4.265%	932.059%
ARCH (10)	4.266%	0.306%	4.266%	929.468%

With MAPE values far exceeding not only the threshold of 50% but also MA and AR models, we can mark it as insufficient for the prediction of Bitcoin volatility.

2.9 GARCH

GARCH(p,q) model is similar to the ARCH(q) model with the addition of β GARCH coefficient, which describes how ARCH coefficient α changes over time.

Equation 95: **GARCH (1,1) conditional variance**

(Source: Processed according to the equation 39)

$$h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2$$

Residuals are calculated with conditional mean *c*.

Equation 96: **GARCH residuals**

(Sourced: Processed according to the equation 35)

$$\varepsilon_t = d_t - c$$

Then conditional for the first ten observed values is equal to long-run variance like in the ARCH(p) model.

Equation 97: **GARCH (p,q) long-run variance**

(Source: Processed according to the equation 40)

$$h_{1-10}^2 = Long\ run\ varinace = \frac{\omega}{1 - SUM(\alpha_1 : \alpha_{10} : \beta_1 : \beta_{10})}$$

Conditional variance for the rest of the observation from the 11th observation.

Equation 98: GARCH (p,q) conditional variance
(Source: Processed according to the equation 42)

$$h_t^2 = \omega + SUMPRODUCT(\alpha_1 : \alpha_{10} : \beta_1 : \beta_{10} ; \varepsilon_{t-1}^2 : \varepsilon_{t-10}^2 : h_{t-1}^2 : h_{t-10}^2)$$

With log-likelihood including β as well.

Equation 99: GARCH (p,q) log-likelihood for a single observation
(Sourced: Processed according to the equation 49)

$$L_t(c, \omega, \alpha, \beta) = LN\left(\frac{1}{\sqrt{2\pi h_t^2}} exp\left[-\frac{\varepsilon_t^2}{2h_t^2}\right]\right)$$

$$\hat{L} = SUM(L_{10} : L_{2616})$$

GARCH(0,0) is again created, with $\alpha_1 = 0$ and $\beta_1 = 0$. So the result is the same as with the ARCH(0) model.

Table 23: GARCH(0,0) coefficients
(Source: Own processing in software Microsoft Excel)

c	0.27%
ω	0.001560
Long-run variance	0.001560
Log Likelihood	4729.99
α_1	0.00
β_1	0.00

To maximize the log-likelihood function for GARCH (1,1) Excel solver tool is used. Setting solver to maximize the function using Gradient descent with variables $c, \omega, \alpha_1, \beta_1$.

Table 24: GARCH(1,1) coefficients
(Source: Own processing in software Microsoft Excel)

c	0.24%
ω	0.000060
Long-run variance	0.002019
Log Likelihood	4949.67
α_1	0.12
β_1	0.85

The exact process is done for all the lags up to GARCH(10,10) model. A list of all coefficients is in Annex IV. And Annex V.

Table 25: **GARCH (p,q) model evaluation**

(Source: Own processing in software Microsoft Excel)

	$\hat{L}$	AIC	λ	p-value
GARCH (1,1)	4949.67	-9891.35	439.37	0.00%
GARCH (2,2)	4950.16	-9888.32	0.97	61.62%
GARCH (3,3)	4961.10	-9906.2	21.89	0.00%
GARCH (4,4)	4965.67	-9911.33	9.13	1.04%
GARCH (5,5)	4970.49	-9916.99	9.66	0.80%
GARCH (6,6)	4978.43	-9928.87	15.88	0.04%
GARCH (7,7)	4982.53	-9933.06	8.19	1.66%
GARCH (8,8)	4982.53	-9929.06	0.00	100.00%
GARCH (9,9)	4982.53	-9925.06	0.00	100.00%
GARCH (10,10)	4982.53	-9921.06	0.00	100.00%

Like in the ARCH model, adding more lags makes GARCH prediction better. GARCH (7,7) shows the highest log-likelihood while maintaining the lowest AIC. Based on the values of GARCH (8,8) and higher, we can assume that the Gradient descent algorithm found the *local* maximum of the log-likelihood function and can not improve by adding more parameters.

Table 26: **GARCH (p,q) prediction evaluation**

(Source: Own processing in software Microsoft Excel)

	MAE	MSE	RMSE	MAPE
GARCH (1,1)	4.244%	0.299%	4.244%	936.377%
GARCH (2,2)	4.246%	0.299%	4.246%	935.906%
GARCH (3,3)	4.251%	0.301%	4.251%	933.314%
GARCH (4,4)	4.256%	0.303%	4.256%	926.975%
GARCH (5,5)	4.252%	0.303%	4.252%	925.470%
GARCH (6,6)	4.237%	0.301%	4.237%	918.926%
GARCH (7,7)	4.237%	0.301%	4.237%	921.052%
GARCH (8,8)	4.237%	0.301%	4.237%	921.052%
GARCH (9,9)	4.237%	0.301%	4.237%	921.052%
GARCH (10,10)	4.237%	0.301%	4.237%	921.052%

With MAPE values far exceeding not only the threshold of 50% we can mark it as insufficient for the prediction of Bitcoin volatility.

3 EVALUATION OF THE PRACTICAL PART

In the practical part time series of Bitcoin price was made stationary through first differentiation. This stationarity was tested with Dickey-Fuller test. Normality, homoscedasticity and autocorrelation were tested both visually and by test. Because normality and homoscedasticity were not fully proven, Bitcoin data set only approaches both conditions. This can make estimation of parameters into models more difficult. While autocorrelation was not proven, because of nature of chosen models it was desired to choose number of lags taken into an account by models. Based on autocorrelation function were 10 lags chosen into all models.

Moving average model which predicts next value based on previous error terms was found inaccurate based on the prediction evaluation methods. The most accurate one was MA (1) with MAPE value of 138%. Values over 50% are considered highly inaccurate. Values over 100% means that prediction errors are bigger than the observed values. Adding more lags into moving average model will not improve them.

Autoregression model makes predictions in regards to previous values themselves. Based on prediction accuracy the AR (1) and AR (6) provide most reliable prediction according to MAPE values. Yet with MAPE of 139% for both of them the are found to be inaccurate. Unlike moving average model adding more lags made the AR (6) model better compared to other orders with higher log likelihood and lower AIC. This makes sense for Bitcoin daily return time series given the difference between MA and AR models.

Because of proven heteroscedasticity ARCH models were showing a promise. Just like autoregressive models, adding more order into ARCH made the log likelihood higher while the AIC values decreased. The most accurate model was ARCH (10) but with MAPE value of staggering 932% it can be said that ARCH models are highly inaccurate. It means that prediction error was nine times bigger then observed values themselves. This can be caused by ARCH inability of predicting values lower then conditional volatility, as seen on the graph comparing test result and realized volatility.

GARCH predicts next observation upon passed error terms like ARCH but adds coefficient based on passed conditional variance as well. As more orders were added the log likelihood function was getting higher and AIC lower. Gaussian descent used for estimating all coefficient probably found local maximum for log likelihood function at

GARCH (7,7) models and adding more orders did not change the result. With MAPE value of 836% this model is highly inaccurate.

Comparison of predictive ability of chosen models came out better for moving average and autoregressive model, especially versions MA (1), AR (1) and AR (6). But no model calculated in this thesis is accurate. Non the less, it was found that Bitcoin is volatile asset and models from traditional financial markets might not be accurate without modification.

4 PROPOSAL

Sample size of 2616 observations is too big for computing econometrics models. Smaller sample size could improve prediction ability of models used in this thesis.

For moving average and autoregression model there are not many modifications available to be made for bigger accuracy of models. Because Bitcoin price is not influenced only by past values they are not suited for predictions of this kind. Multiple linear regression taking into an account external factors should be beneficial for prediction accuracy.

From results of ARCH and GARCH it can be deducted that Microsoft Excel found only local maximum of log likelihood function. For successful calculation of these model better software or calculation technique is required.

All used models use only linear relationship in calculation, using non-linear models might be beneficial as well.

CONCLUSION

Main objective of this thesis was to successfully construct an econometric model predicting Bitcoin volatility on daily data. Total of 40 models were constructed based on previous findings. The result found all used models as highly inaccurate and unable of predicting Bitcoin volatility.

In the first part theory behind econometry, modelling and time series was described. Each model was given detailed description followed by evaluation methods.

The second part construct models from the first theoretical part. For successful construction of the models a number of tests were calculated and evaluated. Findings from the tests defined constraints for the model construction. All models and their prediction were evaluated based on the evaluation techniques from the theoretical part.

Third part was evaluation of the second part.

Fourth part presented suggestions for improvement.

Bitcoin is new type of financial asset with unique properties and high volatility not seen in traditional financial markets. Its time series has non-linear stochastic trend and returns are heteroscedastic. Based on these statistical properties it is hard to model and predict such asset.

It is important to keep in mind, that not all determinants of Bitcoin price can be identified and quantified. Typical example is behaviour of people, which can cause inefficiency in the market. Especially these days with big hype around new digital currencies. Bitcoin is a flag ship of this movement.

All models described were highly inaccurate and unfit for daily predictions. Still their predictions were evaluated, and their inaccuracy explained. This can still help with understanding the nature of Bitcoin daily volatility and with future model creations.

Using smaller sample size, more variables than only past values and picking more robust models might help with accuracy of future models. Also, Microsoft Excel might not be suitable for estimating models with high number of coefficients. For econometric modelling should be a better software or programming language picked.

LIST OF REFERENCES

ALEXANDER, Carol and Daniel F. HECK. Price discovery in Bitcoin: The impact of unregulated markets. The Journal of Portfolio Management [online]. 2020, 50 [cit. 2022-05-07]. Available at: https://www.sciencedirect.com/science/article/pii/S1572308920300759

ALTMAN, Naomi and Martin KRZYWINSKI. Price discovery in Bitcoin: The impact of unregulated markets. Nature Methods [online]. 2015, 12(11), 999-1000 [cit. 2022-05-07]. Available at: https://www.nature.com/articles/nmeth.3627

BOLLERSLEV, Tim. A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. The Review of Economics and Statistics [online]. 1987, 69(3), 542-547 [cit. 2022-05-08]. Dostupné z: https://www.jstor.org/stable/1925546

BOYLE, Patrick. Statistics for the Trading Floor: Data Science for Investing. 1st ed. London: Independently published, 2020, 306 s. ISBN 979-8644826551

BRATU, Mihaela. New Methods of Evaluating the Forecasts Accuracy: A Case Study for USA Inflation. Business and Economic Research [online]. 2013, 3(1), 21-37 [cit. 2022-05-07]. ISSN 2162-4860. Available at: https://www.macrothink.org/journal/index.php/ber/article/view/3103

CoinDesk [online]. Stamford, Connecticut, USA: Digital Currency Group, ©2022 [cit. 2022-03-02]. Available at: https://www.coindesk.com/

Deep Learning 101: What even is a Neural Network?. In: Saltdatalabs.com [online]. New York, USA: SALT DATA LABS, © 2021 [cit. 2022-04-15]. Available at: https://www.saltdatalabs.com/blog/deep-learning-101/what-even-is-a-neural-network

DOLADO, Juan J., Jesús GONZALO and Laura MAYORAL. A Fractional Dickey-Fuller Test for Unit Roots. Econometrica [online]. 2002, 70(5), 1963-2006 [cit. 2022-05-05]. Available at: https://www.jstor.org/stable/3082026

FRANCES, Philip H. and Richard PAAP. Periodic Time Series Models. 1. Oxford: OUP Oxford, 2004. ISBN 978-0-19-152926-9

GHASEMI, Asghar Ghasemi and Saleh ZAHEDIASL. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. Spring [online]. 2012, 10(2), 486-489 [cit. 2022-05-07]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/

GODFREY, K.R. Correlation methods. Automatica [online]. 1980, 16(5), 527-534 [cit. 2022-05-07]. Available at: https://www.sciencedirect.com/science/article/pii/000510988090076X?via%3Dihub

Heteroscedasticity Analysis in Time Series Data. In: Python in Plain English [online]. Nottingham, England: Python in Plain English, 2021 [cit. 2022-04-15]. Available at: https://python.plainenglish.io/heteroscedasticity-analysis-in-time-series-data-fee51503cc0e

HYNDMAN, Rob J and George ATHANASOPOULOS. Forecasting: principles and practice. 2nd ed. Melbourne, Australia: Independently published, 2018, 382 p. ISBN 978-0987507112. Available at: https://otexts.com/fpp2/

KHUNTIA, Sashikanta and J.K. PATTANAYAK. Adaptive market hypothesis and evolving predictability of Bitcoin. Economics Letters [online]. 2018, 167, 26-28 [cit. 2022-05-07]. Available at: https://www.nature.com/articles/nmeth.3627

KIRCHGÄSSNER, Gebhard, Jürgen WOLTERS and Uwe HASSLER. Introduction to Modern Time Series Analysis. 2sc ed. München: Springer, 2012, 332 s. ISBN 978-3642334351

KWAK, Sang Gyu and Jong Hae KIM. Central limit theorem: the cornerstone of modern statistics. Korean J Anesthesiol [online]. 2017, 70(2), 144-156 [cit. 2022-05-07]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370305/#

LO, Andrew W. The Adaptive Markets Hypothesis. The Journal of Portfolio Management [online]. 2004, 30(5), 15-29 [cit. 2022-05-07]. Available at: https://jpm.pm-research.com/content/30/5/15

MAAG, Urs R. and Ghislaine DICAIRE. On Kolmogorov-Smirnov Type One-Sample Statistics. Biometrika [online]. 1971, 58(3), 653-657 [cit. 2022-05-07]. Available at: https://www.jstor.org/stable/2334402

MAHMOUD, Hosam M. Sorting: A Distribution Theory. 1. Hoboken, New Jersey, USA: Wiley-Interscience, 2000. ISBN 978-0471327103.

Microsoft Corporation. Microsoft Excel 365 [software]. ©2022 [přístup 2020-04-02]. Available at: https://www.microsoft.com/cs-cz/microsoft-365/excel

OLSON, David L. and Desheng WU. Predictive Data Mining Models. 1. Singapore: Springer, Singapore, 2017. ISBN 978-981-10-2543-3

ÖZTUNA, Derya, Atilla Halil ELHAN and Ersöz TÜCCAR. Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. Turk J Med Sci [online]. 2006, 36(3), 171-176 [cit. 2022-05-07]. Available at: https://journals.tubitak.gov.tr/medical/abstract.htm?id=8332

PARK, Kun Il. Fundamentals of Probability and Stochastic Processes with Applications to Communications. Softcover reprint of the original 1st ed. 2018. Cham, Switzerland: Springer International Publishing, 2018, 275 s. ISBN 9783319885414.

R Quantile-Quantile (QQ) Plot – Base Graph. In: Learn by example [online]. India: LearnByExample.org, 2019 [cit. 2022-04-19]. Available at: https://www.learnbyexample.org/r-quantile-quantile-qq-plot-base-graph/

SCHÄR, Fabian and Aleksander BERENTSEN. Bitcoin, Blockchain, and Cryptoassets: A Comprehensive Introduction. 1. Cambridge, MA: MIT Press, 2020. ISBN 978-0-262-53916-6.

SCHOBER, Patrick, Christa BOER and Lothar A. SCHWARTE. Correlation Coefficients: Appropriate Use and Interpretation. Anesthesia & Analgesia [online]. 2018, 126(5), 1763-1768 [cit. 2022-04-13]. Available at: https://www.semanticscholar.org/paper/Correlation-Coefficients%3A-Appropriate-Use-and-Schober-Boer/b8d5dbafc507ec21221d523ae869fc14941f61ad

SKRONDAL, Anders a Brian EVERITT. The Cambridge Dictionary of Statistics. 4th ed. Cambridge, England: Cambridge University Press, 2010, 480 s. ISBN 9780521766999

SNYDER, Kenneth A. Condition Assessment of Concrete Nuclear Structures Considered for Entombment. In: National Institute of Standarts and Technology [online]. Gaithersburg, Maryland, USA: United States goverment, 2003 [cit. 2022-05-05]. Available at: https://www.nist.gov/publications/condition-assessment-concrete-nuclear-structures-considered-entombment

STIGLER, Stephen. Fisher and the 5% Level. Automatica [online]. 2013, 21(4), 12 [cit. 2022-05-07]. Available at: https://www.tandfonline.com/doi/abs/10.1080/09332480.2008.10722926

WISE, Bradley. Time Series Analysis with ARIMA: Part 2. In: Cisco Blogs [online]. San Jose, California, USA: Cisco, 2020 [cit. 2022-05-07]. Available at: https://blogs.cisco.com/analytics-automation/arima2

LIST OF EQUATIONS, PICTURES, GRAPHS AND TABLES

List of equations:

List of pictures:

List of graphs:

List of tables:

List of Annex

Annex I: MA (q) coefficients and p-values

(Source: Processed in software Microsoft Excel 365)

	c	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MA (1)	0.0027	-0.0483									
p-value	0.05%	1.36%									
MA (2)	0.0027	-0.0486	0.0160								
p-value	0.05%	1.31%	41.45%								
MA (3)	0.0027	-0.0489	0.0157	0.0128							
p-value	0.05%	1.25%	42.36%	51.34%							
MA (4)	0.0027	-0.0489	0.0155	0.0130	0.0089						
p-value	0.05%	1.26%	42.96%	50.64%	67.78%						
MA (5)	0.0027	-0.0495	0.0157	0.0133	0.0092	0.0160					
p-value	0.05%	1.16%	42.20%	49.79%	64.04%	41.59%					
MA (6)	0.0027	-0.0489	0.0167	0.0135	0.0069	0.0155	0.0392				
p-value	0.06%	1.27%	39.36%	49.25%	72.44%	42.93%	4.59%				
MA (7)	0.0026	-0.0495	0.0165	0.0145	0.0068	0.0151	0.0391	-0.0174			
p-value	0.06%	1.16%	40.11%	45.87%	72.91%	44.25%	4.63%	37.53%			
MA (8)	0.0027	-0.0496	0.0175	0.0147	0.0064	0.0152	0.0394	-0.0173	-0.0175		
p-value	0.06%	1.15%	37.22%	45.38%	74.32%	43.87%	4.49%	37.78%	37.16%		
MA (9)	0.0026	-0.0483	0.0161	0.0133	0.0063	0.0162	0.0381	-0.0181	-0.0172	-0.0142	
p-value	0.08%	1.37%	41.09%	49.80%	74.62%	40.90%	5.19%	35.57%	37.86%	46.88%	
MA (10)	0.0026	-0.0483	0.0168	0.0135	0.0061	0.0161	0.0377	-0.0204	-0.0179	-0.0155	0.0505
p-value	0.08%	1.39%	39.20%	48.98%	75.55%	40.99%	5.44%	29.79%	36.18%	42.87%	1.00%

Annex II: AR (p) coefficients and p-values

(Source: Processed in software Microsoft Excel 365)

	c	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
AR (1)	0.0027	-0.0480									
p-value	0.04%	1.44%									
AR (2)	0.0027	0.0118	-0.0474								
p-value	0.05%	54.45%	1.57%								
AR (3)	0.0027	0.0133	0.0125	-0.0475							
p-value	0.06%	49.74%	52.34%	1.54%							
AR (4)	0.0026	0.0087	0.0137	0.0124	-0.0477						
p-value	0.07%	65.55%	48.43%	52.75%	1.52%						
AR (5)	0.0026	0.0155	0.0095	0.0135	0.0121	-0.0478					
p-value	0.09%	42.91%	62.83%	49.12%	53.52%	1.49%					
AR (6)	0.0025	0.0391	0.0174	0.0089	0.0129	0.0117	-0.0484				
p-value	0.15%	4.55%	37.43%	64.75%	50.96%	54.84%	1.36%				
AR (7)	0.0025	-0.0140	0.0384	0.0176	0.0091	0.0130	0.0120	-0.0479			
p-value	0.13%	47.48%	4.97%	36.94%	64.04%	50.53%	54.07%	1.48%			
AR (8)	0.0026	-0.0189	-0.0149	0.0387	0.0178	0.0093	0.0133	0.0127	-0.0481		
p-value	0.11%	33.33%	44.64%	4.83%	36.23%	63.35%	49.51%	51.60%	1.43%		
AR (9)	0.0026	-0.0199	-0.0199	-0.0146	0.0390	0.0180	0.0097	0.0141	0.0124	-0.0485	
p-value	0.09%	31.00%	30.94%	45.48%	4.67%	35.68%	62.13%	47.03%	52.52%	1.35%	
AR (10)	0.0025	0.0477	-0.0175	-0.0206	-0.0154	0.0384	0.0172	0.0077	0.0148	0.0134	-0.0475
p-value	0.16%	1.49%	37.14%	29.34%	43.26%	4.94%	37.98%	69.24%	44.97%	49.45%	1.54%

Annex III: ARCH (q) coefficients

(Source: Processed in software Microsoft Excel 365)

ARCH(p)	0	1	2	3	4	5	6	7	8	9	10
	α	α	α	α	α	α	α	α	α	α	α
1	0.00	0.14	0.13	0.13	0.14	0.15	0.13	0.13	0.14	0.14	0.15
2			0.12	0.09	0.08	0.08	0.06	0.05	0.06	0.06	0.06
3				0.11	0.11	0.07	0.05	0.05	0.03	0.03	0.02
4					0.21	0.20	0.19	0.18	0.18	0.18	0.18
5						0.16	0.14	0.14	0.13	0.13	0.12
6							0.12	0.11	0.10	0.10	0.09
7								0.05	0.03	0.03	0.03
8									0.12	0.12	0.13
9										0.01	0.00
10											0.05
c	0.27%	0.27%	0.29%	0.28%	0.25%	0.26%	0.25%	0.24%	0.22%	0.22%	0.22%
ω	0.00156	0.00132	0.00117	0.00106	0.00078	0.00065	0.00058	0.00056	0.00047	0.00046	0.00043
long-run variance	0.00156	0.00154	0.00157	0.00158	0.00174	0.00187	0.00196	0.00195	0.00231	0.00232	0.00254

Annex IV: GARCH (p,q) coefficients

(Source: Processed in software Microsoft Excel 365)

GARCH(p,q)	(0,0)		(1,1)		(2,2)		(3,3)		(4,4)		(5,5)	
	α	β	α	β	α	β	α	β	α	β	α	β
1	0	0	0.12	0.85	0.14	0.70	0.16	0.60	0.13	0.82	0.14	0.22
2					-0.01	0.14	-0.03	-0.56	-0.04	-0.45	0.04	0.25
3							0.10	0.67	0.02	0.56	-0.03	0.06
4									0.09	-0.18	0.14	0.35
5											-0.01	-0.23
6												
7												
8												
9												
10												
c	0.27%		0.24%		0.24%		0.24%		0.23%		0.23%	
ω	0.001560		0.000060		0.000064		0.000115		0.000105		0.000141	
long-run variance	0.001560		0.002019		0.002034		0.002085		0.002179		0.002291	
Log-Likelihood	4729.99		4949.67		4950.16		4961.10		4965.67		4970.49	

Annex V: GARCH (p,q) coefficients

(Source: Processed in software Microsoft Excel 365)

GARCH(p,q)	(6,6)		(7,7)		(8,8)		(9,9)		(10,10)	
	α	β	α	β	α	β	α	β	α	β
1	0.15	0.25	0.17	0.20	0.17	0.20	0.17	0.20	0.17	0.20
2	0.03	-0.76	0.02	-0.71	0.02	-0.71	0.02	-0.71	0.02	-0.71
3	0.10	0.03	0.12	-0.02	0.12	-0.02	0.12	-0.02	0.12	-0.02
4	0.19	0.46	0.19	0.47	0.19	0.47	0.19	0.47	0.19	0.47
5	-0.01	-0.22	-0.03	-0.21	-0.03	-0.21	-0.03	-0.21	-0.03	-0.21
6	0.16	0.50	0.17	0.47	0.17	0.47	0.17	0.47	0.17	0.47
7			-0.02	0.04	-0.02	0.04	-0.02	0.04	-0.02	0.04
8					0.00	0.00	0.00	0.00	0.00	0.00
9							0.00	0.00	0.00	0.00
10									0.00	0.00
c	0.21%		0.21%		0.21%		0.21%		0.21%	
ω	0.000287		0.000289		0.000289		0.000289		0.000289	
long-run variance	0.002181		0.002207		0.002207		0.002207		0.002207	
Log-Likelihood	4978.43		4982.53		4982.53		4982.53		4982.53	