TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance
Department of Economics and Finance

Ganenthra Ravindran

MARKET EFFICIENCY OF BITCOIN: EVIDENCE FROM THE EFFICIENT MARKET HYPOTHESIS (EMH)

Bachelor's thesis

Programme International Business Administration, specialisation Finance

Supervisor: Simona Ferraro, PhD

Co-supervisor: Richard Dreyling, MA

Tallinn 2022

I hereby declare that I have compiled the thesis independently
and all works, critical standpoints, and data by other authors
have been properly referenced and the same paper
has not been previously presented for grading.
The document length is 9,900 words from the introduction to the end of the conclusion.

Ganenthra Ravindran ……………………………
(signature, date)
Student code: 194465TVTB
Student e-mail address: ganent@ttu.ee

Supervisor: Simona Ferraro, PhD:
The paper conforms to requirements in force

……………………………………………
(signature, date)

Co-supervisor: Richard Dreyling, MA:
The paper conforms to requirements in force
……………………………………………
(signature, date)

Chairman of the Defence Committee:
Permitted to the defence
…………………………………
(name, signature, date)

TABLE OF CONTENTS

ABSTRACT

The number of Bitcoin users is exponentially growing. This growth is fuelling concerns over its potential adoption as an alternative currency and stable medium of exchange due to its price volatility. The primary aim of this thesis is to identify whether the Bitcoin market is efficient by testing the weak-form of the Efficient Market Hypothesis (EMH).

To determine whether Bitcoin is weak-form market efficient, this thesis analyses the 1-minute weighted logarithmic closing price returns of Bitcoin over a full sample period and two additional sub-sample periods between the 1st of January 2014 -1st of January 2020 with data that is extracted from Kaggle. Five statistical tests are conducted to determine whether Bitcoin returns follow a sequence formed by a random walk. The tests are the following: the Ljung-Box test; the Runs test; the Cox-Stuart test; the Hurst test and finally the Variance Ratio test.

The empirical findings suggest that the Bitcoin market at the 1-minute frequency is weak-form inefficient over the full sample period and two additional sub-sample periods but could be heading towards a path of efficiency based on the results of the Cox-Stuart and Hurst test.

Keywords: Bitcoin, efficiency, efficient market hypothesis, cryptocurrency, blockchain, randomness

INTRODUCTION

Digital currencies, the most popular of which is Bitcoin[1] are seeing immense growth in users from relative obscurity over a decade ago. Global figures of cryptocurrency users exceed 200 million as of June 2021, and out of that figure, 114 million are Bitcoin owners (Wang 2021). Powered by blockchain, Bitcoin is transparent due to its distributed decentralised network, fast due to speedier processing and cheaper due to lower processing fees. The potential for Bitcoin to democratise and revolutionise Finance as an alternative currency for the unbanked masses, a keen interest in digital currencies and data science are primary motivators for choosing the topic of this thesis.

A significant cause of concern about the usage of Bitcoin as an alternative currency is its price volatility. The volatility in Bitcoin prices questions its market efficiency because if it is an actual store of value, it would not exhibit such sharp movements in prices which is an indication that it might be in a speculative bubble (Urquhart 2016). Therefore, the main aim of this thesis is to explore whether the Bitcoin market is in the bubble phase. This is done by examining Bitcoin returns through the weak-form of the Efficient Market Hypothesis (EMH) developed by Eugene F. Fama (1970). A market is deemed efficient when prices fully reflect all information (Fama 1970). According to the EMH, historical asset prices do not predict future asset prices in weak-form efficient markets as prices follow a random walk.

This thesis will attempt to answer the following research question:

- Are Bitcoin returns weak-form efficient according to the Efficient Market Hypothesis (EMH)?

Based on this research question, the following null hypothesis is developed:

- Bitcoin price return follows a sequence formed by a random walk.

[1] Research by Urquhart (2016), Eross *et al.* (2019) and Nadarajah and Chu (2016) refer to Bitcoin with the uppercase "B", referring to the currency and not the payment protocol. In line with prior research, the author of this thesis also references Bitcoin with the uppercase "B" and refers to the currency instead of the payment protocol.

The logarithmic 1-minute weighted closing price returns of Bitcoin are studied over a full sample period and two additional sub-sample periods between the 1st of January 2014 to the 1st of January 2020, using a dataset titled "Bitcoin Historical Data", that has been extracted from Kaggle. The novelty of this thesis lies in the size of the observations within the dataset and the robustness of the statistical tests that are applied to determine randomness. Numerous research has been conducted on the EMH of Bitcoin. However, to the author's knowledge, there has been no research conducted thus far that has evaluated the efficiency or inefficiency of Bitcoin through the EMH, by analysing the returns at the 1-minute frequency using the same statistical tests. Data at this frequency provides more observations and a lower standard error which ensures a higher degree of accuracy

Five different tests are used to answer the research question. With a self-developed code, the tests are conducted in data studio with the programming language R and are the following:

1. Runs test/Wald-Wolfowitz Runs test: A non-parametric statistical test designed to detect if a data set follows a random process and is mutually independent (Wald, Wolfowitz 1940).
2. Ljung-Box test: A statistical test that determines whether any given data frames have correlations different from zero (Ljung, Box 1978).
3. R/S Hurst: A test used to measure long-term memory of time series (Hurst 1951).
4. Cox Stuart Test: A test for identifying trends within a data series. (Cox, Stuart 1955).
5. Variance ratio test (Automatic Portmanteau): Performed to identify whether a data set follows a random walk (Lo, Mackinlay 1989).

This Bachelor's thesis consists of three chapters. In the first chapter, the author provides an in-depth theoretical understanding of Bitcoin and blockchain technology as well as examines the concept of efficiency through the EMH. The second chapter discusses the data and methodology applied, where the data used shall be presented along with the efficiency tests utilised to determine randomness. Lastly, the third chapter presents the results of the efficiency tests and provides further discussions and limitations of the findings before concluding.

1. BITCOIN AND EFFICIENT MARKET HYPOTHESIS (EMH)

In this chapter, the author provides an overview of Bitcoin, the EMH and a review of studies on the efficiency of Bitcoin. The first sub-chapter focuses on Bitcoin, where a brief history of digital currencies shall be provided and a review of its functional and operational principles. In the second sub-chapter, the author summarizes the theoretical and empirical background of the Efficient Market Hypothesis (EMH), explains the different forms of efficiency within the EMH and an alternative theory for evaluating market efficiency. Finally, in the third and final sub-chapter, the author reviews the relevant empirical studies on the efficiency of Bitcoin from which the hypothesis of this thesis is drawn.

1.1. Bitcoin

Bitcoin was first introduced by Satoshi Nakomoto (2008) in his whitepaper “Bitcoin: A Peer-to-Peer Electronic Cash System” and is a form of cryptocurrency or digital currency based on cryptographic proof instead of trust. The concept of digital currency far predates the existence of Bitcoin. Various research has paved the road toward creating Bitcoin and has been cited by Nakamoto (2008) as the source of technical inspiration. This includes Ralph Merkle (1980) through his work on public-key cryptography and the development of the Merkle Tree. Through the Merkle Tree, extensive dataset verification is enabled efficiently and secured via the mathematical method known as a Merkle root. Cited for their research on developing numerous elements of a timestamp verification system, Stuart Haber and Scott Stornetta (1991) are other researchers cited by Nakamoto (2008).

However, an agreed consensus is that cryptocurrency has its primary roots in the dissertation of David Chaum (1982) on decentralised networks. Through his vault system, Chaum (1982) introduced the concept of having mutually suspicious individuals trusting each other without the need of an equally dubious third-party or middleman. This is done through secure physical vaults or encrypted servers that participate in constant exchanges that sign, record and finally broadcast each transaction it processes. In 1989 Chaum founded DigiCash, which is considered the world's first digital currency that held the properties of physical cash (Pitta 1999). Conceptually, DigiCash and Bitcoin share similarities in their encryption. The author shall explain the technology that powers Bitcoin and how it functions in the subsequent sub-chapters.

1.1.1. Blockchain

The decentralisation of Bitcoin is enabled through blockchain, a decentralised distributed public ledger or database (Vigna, Casey 2016). This database is distributed to all participants through a peer-to-peer (P2P) network and is a large scale operational version of David Chaum's (1982) vault system. The decision-making process in a centralised, decentralised and distributed system differs. As the name suggests, in a centralised system, decisions are made by a single authority that has control over the entire decision-making process (Vergne 2020). In a decentralised system, however, the decision-making process is executed at multiple levels. On the other hand, in a distributed decentralised system such as blockchain, as shown in Figure 1 all participants own the same database and have an updated version of the database, much like a google sheet (Mehta *et al.* 2021, 36).

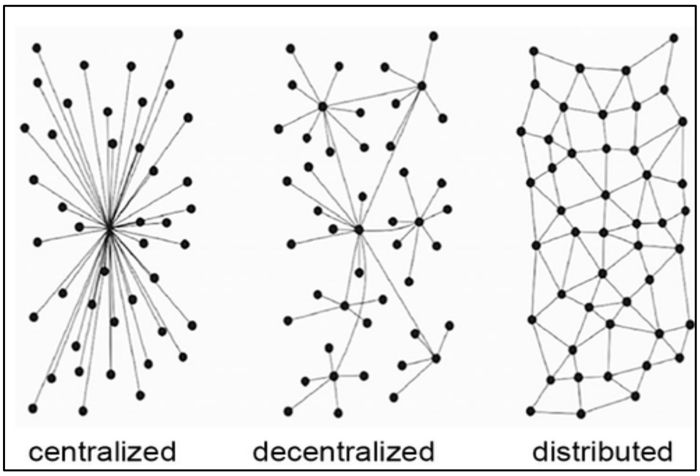

Figure 1. Baran's typology of communication networks
Source: Vergne (2020, 4)

Blocks containing information groups are continuously updated and synchronised throughout this distributed decentralised network. This enables every individual to access the same information, such as when new transactions have taken place, in complete anonymity through randomised usernames (Mehta *et al.* 2021, 56). Each block is stored in a linear chain that mathematically references one block to another and will continue to grow indefinitely (Vigna, Casey 2016, 194).

This mathematical referencing technique is “hashing”, a cryptographic process that feeds input data of any size, such as words and numbers, into an algorithm that generates unique strings of a fixed length that act as "fingerprints" (Mehta *et al.* 2021, 39). In the Bitcoin protocol or set of rules, the hash function that is extensively used is SHA-256, which enables new transactions to be written (Vigna, Casey 2016, 202). Bitcoin batches as many transactions into blocks, with each block having a unique hash that is partly connected to the block's hash before it, thus resulting in an ordered chain-like sequence (Vigna, Casey 2016).

Besides being a means of verification, blockchain is also vital for Bitcoin and other cryptocurrencies as it solves the double-spending problem (Nakamoto 2008). The double-spending problem is whereby a single token could potentially be spent more than once. This problem, for instance, is not a problem for physical currency. This is because the exact Dollar or Euro note spent on a cup of coffee cannot be spent on something else.

Blockchain has become synonymous with Bitcoin with the rise in its usage. Nevertheless, it should be noted that blockchain has its uses far beyond cryptocurrencies. While blockchain is mainly used for money transfers, its application has extended to securing medical information, supply chain management through smart contracts, and even delivering benefits to refugees in recent years through the building blocks program facilitated by the United Nations (Mehta *et al.* 202, 15).

The use of blockchain is vital for the functioning of Bitcoin as discussed in this section. In the subsequent section, the author will provide an overview of how Bitcoin works by using blockchain technology.

1.1.2. How Bitcoin works

A transaction refers to the exchange of Bitcoin value on the blockchain between two parties. Unlike fiat currency, which depends on a bank or payment processor to process the transfer of funds, Bitcoin's method of exchanging funds is vastly superior because transaction costs are either lower or eliminated (Nakamoto 2008). It is also safer because the entire transaction process can be conducted anonymously. Individuals, in essence, act as their banks by facilitating transactions through a few simple clicks without the need to share their private information with any intermediary. The transaction process of Bitcoin involves four separate steps: transaction creation and signing, broadcasting, propagation and verification, and finally, validation (Nakamoto 2008).

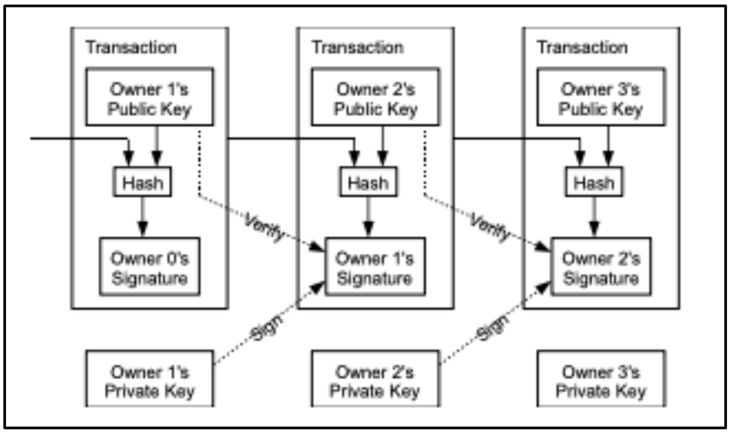

Figure 2. Transaction process
Source: Nakamoto (2008, 2)

Transaction creation and signing via Bitcoin are done through a two-key authentication system and represent the largest component of the entire process. Each transaction consists of three main parts: input, an amount, and output (Nakamoto 2008). The input refers to the Bitcoin address of the sender, the amount relates to the amount that the sender aims to send, and the output relates to the address that will receive the Bitcoin. Individuals who wish to exchange funds can do so anonymously through the blockchain protocol that sends and receives payment information by directly linking individuals through a wallet, a public key as an identity and a private key, as shown in Figure 2.

Creating a Bitcoin wallet can be done with ease, and numerous exists depending on the needs and level of authentication required by the user. Unlike physical wallets that hold fiat currency, these wallets do not contain Bitcoin. Instead, they have Bitcoin addresses (Mehta *et al.* 2021, 109). Each wallet contains a public key as well as a private key. Public keys are addresses that act as an account number that can be shared that enables a person to receive Bitcoins, and there are no limits on the number of public keys that can be owned (Mehta *et al.* 2021).

On the other hand, a private key is like a password with a string of numbers and letters. Private keys act as digital signatures and proof of ownership of a blockchain address, which enables a person to send and eventually spend Bitcoins (Mehta *et al.* 2021, 39). Private keys are unique and are not shared. It is only known to the owner of the wallet. Losing the private key of a Bitcoin wallet would mean losing all the funds in it. After several attempts to determine the correct key, the wallet automatically seizes and encrypts its content forever without a recovery method. It is estimated that around 20 per cent of the total supply of Bitcoin under circulation is lost (Popper 2021).

The second step in the transaction process is broadcasting. Once an individual enters the private key and signs a particular transaction, this transaction is sent for validation or broadcasted to the closest node on the Bitcoin network (Antonopoulos 2015). The third step is the verification and propagation process. Upon arriving at the nearest node, information is propagated into the network and verified by matching the private key with the corresponding public key (Decker, Wattenhofer 2013). If this validation process is a success, it enters a memory pool and waits for miners, a loosely organised network of incentivised participants, to add it to subsequent blocks.

The final step is validating a transaction. After a transaction is in the memory pool, miners pick up a transaction and group them into blocks. Each block has a limit to it. A limited number of transactions can be entered into a particular block. Depending on the exchange and the amount, three confirmations will be required before the data is added onto the blockchain and the Bitcoin can be spent (Antonopoulos 2015). Mining is essentially a process of confirming transaction requests by a decentralised consensus system. It is called mining because resources such as vast amounts of energy and computers with advanced processing power need to be expended.

The consensus type used for Bitcoin and other cryptocurrencies such as Ethereum is the Proof of Work (POW), shown in Figure 3. This is where miners compete to solve complex mathematical problems/puzzles to create new blocks of transactions that exponentially increase in difficulty (Antonopoulos 2015).

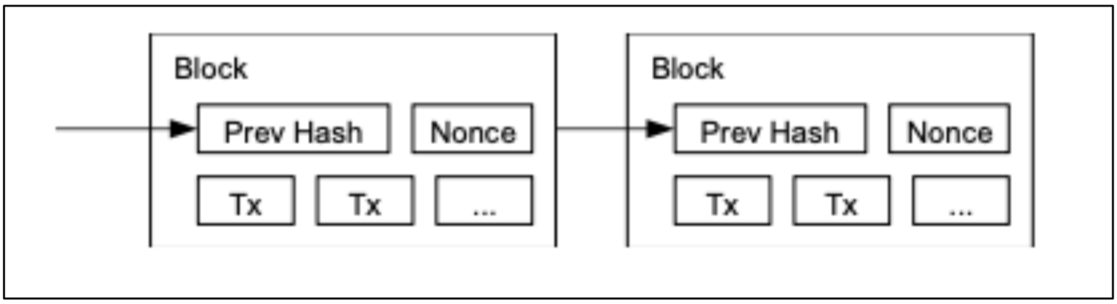

Figure 3: Proof-of-Work Process

Source: Nakamoto (2008, 3)

As previously mentioned, mining is resource-intensive as it requires immense computing power and electricity. The Bitcoin system incentivises miners in two distinctly separate ways for expending their resources. The first way miners are incentivised is by generating the POW, which creates new blocks of transactions. By generating the POW, miners are rewarded with Bitcoins (Sompolinsky, Zohar 2017). The second way in which miners are rewarded is by facilitating transactions. As the number of transactions that can be inputted into a block is pre-defined, senders can add a transaction fee on top of the amount they are sending as an incentive for early inclusion in a block when there is a backlog of transactions waiting to be validated (Antonopoulos 2015, 256). The limit in the supply of Bitcoin will mean that in the future, Bitcoin miners will only be rewarded by facilitating transactions as there will only be 21 million Bitcoins in supply (Ciaian *et al.* 2016). Known as a hard cap, the supply of Bitcoin has been limited through an algorithm encoded into its source code to restrict its supply to ensure its value growth. Known as Bitcoin halving, it has been suggested that this occurs as a way to make it a deflationary currency, which makes individual coins far more valuable (Nakamoto 2008).

This section provided an overview of the functional and operational principles of Bitcoin. In the subsequent section, an overview of the progress of the adoption of Bitcoin as an alternative currency shall be discussed.

1.1.3. Bitcoin as a potential alternative currency

This section serves as a bridge for a smooth transition to the subsequent sub-chapter and aims to briefly provide an overview of the progress made in the adoption of Bitcoin as an alternative currency.

An alternative currency is a medium of exchange that could be used in parallel with fiat currency (Appukuttan 2019). At the time of authoring this thesis, no major developed country has legally adopted Bitcoin or any other cryptocurrency as an alternative currency. In 2015, a decision by the European Court of Justice declared that Bitcoin transactions are exempt from Value Added Tax (VAT) under the provision concerning transactions relating to currency, banknotes and coins used as a legal tender (ECJ judgement C-264/14 point 6). While this judgment does not explicitly define Bitcoin as a legal currency, the court has categorised Bitcoin as a medium of exchange along with banknotes, and coins, thus exempting it from VAT. In 2021, El Salvador became the first sovereign nation to adopt Bitcoin as a legal tender along with the USD, which it had adopted as the de-facto currency in 2001 (Arslanian 2021). Another sovereign nation, the Central African Republic, has announced that it would approve Bitcoin as an alternative legal tender (Browne 2022).

As society is becoming more embracive of Bitcoin and other digital currencies, the case for examining the efficiency of Bitcoin prices becomes more vital to deepen an understanding on the possible existence of anomalies within the cryptocurrency market. This enables individuals, legislators, and economists to take precautionary actions if the Bitcoin market operates in a speculative bubble. In the following sub-chapter, the author shall provide an overview of the Efficient Market Hypothesis.

1.2. Efficient Market Hypothesis (EMH)

The Efficient Market Hypothesis (EMH) is a foundational principle of Finance. A market is efficient when it "fully reflects" all available information (Fama 1970). Research into the EMH can be split into three separate phases. The first phase consists of its conceptual construction in the 1960s, mainly by Paul Samuelson (1965) and Eugene Fama (1965); the second is its empirical testability by Fama (1970). Finally, the third phase is where it becomes increasingly challenged as a theory in the 1980s and beyond. The third phase can be reflected more recently through Andrew Lo's (2004) work on the Adaptive Market Hypothesis (AMH), which provides an updated understanding of market efficiency, briefly touched upon in a section within this sub-chapter.

1.2.1. Importance of the EMH

In Finance, asset pricing determines how much individual investors decide to pay for a particular stock. Rational investors determine the price or market value of stock fairly or unfairly based on future earnings and the discount rate. The riskier the future earnings, the lower the market value and vice versa (Keshari *et al*. 2020). The rationality of evaluating investments in this manner brings us to market efficiency, which suggests that it is impossible to "beat the market" on a risk-adjusted basis.

A market that reflects all public information means that by the time an individual has an insight about a company, so do all other market participants. As a result, the price has already been adjusted, making it impossible for an individual to buy undervalued assets or sell them above their fair value (Chu *et al*. 2019, 222). According to Fama (1970), informational efficiency in a market is due to competition, low transaction costs and readily available information. As new information is incorporated into the market, prices change quickly and randomly. Grossman and Stiglitz (1980) have shown that excessive returns will still, however, exist if costs are attributed to gathering and processing information, which acts as a barrier.

An efficient market ensures that asset prices are as fair as possible because it reflects the collective expectations of all participants in the market. Fama (1970), through his research, has identified three different subsets through the EMH. These are the weak-form, the semi-strong form and the strong form of efficiency. Each form of efficiency emphasises certain types of information instead of others. In the next section, the author shall explain the three different forms of efficiency in much more detail.

1.2.2. The three forms of the EMH

Different forms of efficiency have different types of information. There are three different types of information: unpublished private information, published public information and historical information (Kang *et al.* 2022).

In weak-form efficient markets, publicly available historical information on the movements in prices, volume and earnings data cannot be used to predict its future direction through technical analysis (Malkiel 2003, 59). This is because the movements of future stock prices are random as it follows a random walk. As such, publicly available historical data does not provide any predictive power to an individual investor (Malkiel 2003, 59). Whether a market is weak-form efficient is dependent on whether returns are independently and identically distributed (Smith 2012). Most efficiency studies on Bitcoin concern the weak-form of efficiency, which shall also be the case in this thesis.

The semi-strong form of efficiency considers published past and current publicly available information such as annual reports, dividend announcements, stock splits and others (Kang *et al.* 2022). This form of efficiency assumes that current stock prices adjust and adapt to the release of new public information. The adjustment occurs because the market anticipates it (Fama 1970). Fundamental analysis measures a stock's intrinsic value and can include many elements such as revenue, earnings, profit margins and other data that affect a stock's value. In addition to technical analysis, this form of efficiency suggests that fundamental analysis cannot predict the future movement of prices (Fama 1970).

The final form of efficiency that shall be discussed is the strong-form of efficiency. This form of efficiency incorporates unpublished private, published public and historical information (Degutis, Novickyte 2014). Fama (1970) asserts that even by insider trading or monopolistic access to information, individual investors cannot beat the market or realise an excess return in a strong-form efficient market. Mutual funds are professionally managed investment funds. Fama (1970, 412) studied mutual funds to determine if the performance of professionally managed funds could exceed the market line and found that in 89 out of 115 cases, the risk-adjusted returns of mutual funds over ten years were lower than the market line.

Different forms of efficiency were discussed in this section. A core component of the weak-form of efficiency is the random-walk model which has its basis in the fair game model. Both these models shall be discussed in the next section.

1.2.3. The random walk and fair game model

A cornerstone of weak-form efficiency is that a random walk process forms prices. This process asserts that the ideation of a path is followed by a succession of random steps that cannot be predicted (Fama 1970). According to the random walk model, history does not repeat itself as future stock price projections, for instance, are uncorrelated with past data.

The theoretical underpinnings of the random walk model span the academic disciplines of physics, mathematics, and Finance. Thermal collision with liquid molecules causes the erratic motion of pollen grains, known as the "Brownian motion" after Robert Brown (1828). The observation by Brown significantly furthered the inquiry of researchers into the random walk. In Finance, this stochastic process can trace its immediate roots in research on the movements of sound waves through heterogeneous materials by Lord Rayleigh (1880).

After that, Karl Pearson (1905) introduced a formal concept of the random walk in a letter to the scientific journal Nature. Pearson sought to investigate the distribution of mosquitos. However, random walk as a basis for modern quantitative Finance can be traced back to mathematician Louis Bachelier (1900), who proposed it as a fundamental model for financial time series through his doctoral thesis “La Theorie de la Speculation”. More empirical work in the 1950s and 1960s suggested that prices move randomly.

There are two fundamental assumptions of the random walk model. The first assumption is that successive price changes or returns are independent because a particular observation within a sequence is assumed to have no impact on subsequent observations. The second assumption is that price changes are identically distributed (Fama 1970). Both these assumptions are captured with the Equation 1 (Philips 1988, 244):

$$f\left(r_{jt+1} | \Phi_t \right) = f\left(r_{jt+1} \right) \tag{1}$$

where
f – probability function,
r_{jt+1} – one-period percentage return,
t – time,
Φ – the general symbol for the information set at time t.

To investigate whether a time series follows a random walk, methods such as the R/S Hurt analysis, amongst others, have been used by the author to measure the long-term memory within a time series, i.e., the amount by which the series deviates from the random walk.

The fair game model serves as the basis of the random walk model. The fair game model was developed by Paul Samuelson (1965), and it asserts that in a competitive market, prices reflect investor expectations and are expected to adapt to new information. Both Fama (1965) and Samuelson (1965) assert that randomness in stock market prices is due to the rationality of individuals that participate in it. Nevertheless, both researchers deviate in explaining randomness from the probabilistic models that they use. Samuelson introduces the Martingale model, whereas Fama uses the already well-established random walk model (Delcey 2019).

Fama (1970) begins his study by stating that the EMH theory on efficiency is too broad, so it needs to be narrowed down to make the model testable. To make it testable, Fama states that the process of price formation needs to be specified. He suggests using the Capital Asset Pricing Model (CAPM) to compute the market equilibrium of prices. Differing theories have different methods of defining risk. Nevertheless, Fama (1970, 384) asserts that the formation of prices can be defined through the Equation 2:

$$E\left(\tilde{P}_{j,t+1} | \Phi_t\right) = \left[1 + E\left(\left(r_{j,t+1} | \Phi_t\right)\right)\right] P_{jt} \tag{2}$$

where
E – expected value operator,
$\tilde{P}_{\mathrm{j,t+1}}$ – price of security j at $\mathrm{t}+1$,
$\mathrm{P_{jt}}$ – price of security j at time t,
$\mathrm{r_{j,t+1}}$ – one period percentage return at $\mathrm{t}+1$,
Φ_t – a general symbol for the information set is assumed to be "fully-reflected" at time t.

Fama states that two assumptions can be made from Equation 2 thus far. The first assumption is that market equilibrium is described as an expected return. The second assumption is that the expected return fully reflects the relevant information set, illustrated in the formation of the price $\mathrm{p_{jt}}$ (Fama 1970, 384). Combined, these assumptions lead to the statement that capital allocation is "fair game" and that it is impossible to beat the market.

The EMH is the dominant theory in determining efficiency. However, there are alternate theories that examine efficiency. In the subsequent section, the author shall discuss the Adaptive Market Hypothesis (AMH), which combines numerous schools of thoughts.

1.2.4. Alternative theory of efficiency

The basic idea in Economics and Finance is that financial market prices are efficient and that people are rational actors. In the last decades, increasing numbers of research have challenged the EMH by asserting that markets are inefficient, and people are irrational, creating an opposing school of thought in the process.

The Adaptive Market Hypothesis (AMH) by Andrew Lo (2004) provides a dynamic and evolutionary perspective of efficiency by bringing together principles from various disciplines such as ecology, neuroscience, and biology, amongst others that reconcile the two schools of thought, the school which supports EMH and the school which opposes EMH. According to the AMH, rational and irrational behaviour exists in the market. Most of the time, people are rational; however, on some occasions, people can also be irrational, which can lead to temporary inefficiencies in the markets.

As opposed to EMH, which examines market efficiency from a static perspective, the AMH asserts that market efficiency varies at different periods due to the evolution of markets. Lo (2004), through his research, asserts that markets are sometimes rational and sometimes irrational. A core principle of the EMH is that expected returns depend on the level of risk that an individual takes.

This notion is significantly contrasted with the AMH, which asserts that consistent expected returns are expected by individuals who adapt to the evolving nature of market conditions (Chu *et al.* 2019). Conclusively, Lo (2017) argues that, for the most part, markets are efficient as it is hard to generate excess returns; nonetheless, on occasions, markets do reflect individual sentiment.

1.3. Studies on the efficiency of Bitcoin

The EMH has been researched extensively on various assets and commodities. Although the efficiency of Bitcoin and other cryptocurrencies is becoming more prevalent due to its infancy, there are still areas that need further examination. The aim of this sub-chapter is to provide a detailed overview of the empirical studies on the EMH of Bitcoin along with the tests used to determine randomness. The findings of several researchers on the EMH of the Bitcoin market are discussed in the following paragraphs and summarised in Table 1.

Urquhart (2016) is the first to examine weak-form efficiency in the Bitcoin market and serves as the foundation for subsequent studies. The tests used by Urquhart include the following five: Ljung-Box test, Bartels test, Runs test, Brock-Deckhert-Scheinkman (BDS) test, automatic variance ratio (AVR) test and the R/S Hurst analysis. The entire sample period of this research study by Urquhart (2016, 82) was from the 1st of August 2010 to the 31st of July 2016. The first sub-sample period was from the 1st of August 2010 to the 31st of July 2013 and the second sub-sample period was from the 31st of August 2013 to the 31st of July 2016. The research was split into further sub-sample periods to analyse if efficiency varies over time. The Ljung-Box hypothesis is rejected over the entire sample period and the first subsample period, thus indicating serial correlation. The Runs and Bartels test is subsequently rejected for all periods, suggesting serial correlation and therefore a lack of randomness.

The AVR test also determines whether the return data follows a random walk. As the p-values were below the 0.05 significance level for the entire sample period and first sub-sample period, the null hypothesis is rejected, suggesting the non-randomness of successive values. The BDS determines the independence and distribution of a data sequence, and this test also rejects the null hypothesis for all sample periods. The final adopted test was the R/S Hurst test, which showed a pattern of negative correlation. Urquhart concludes that Bitcoin is becoming more efficient, as suggested by the results in the second subsample of the study (Urquhart 2016).

Table 1: Summary of literature review

Author(s) (Year)	Sample Size (N)	Method	Sample Period	Main Findings
Urquhart (2016)	2183	Ljung and Box, Runs, Bartels, AVR, BDS, R/S Hurst analysis	2010-2016	Mainly inefficient but moving towards efficiency.
Nadarajah and Chu (2016)	2191	Ljung and Box, Runs, Bartels, AVR, BDS, R/S Hurst analysis, Spectral Shape, Robustified Portmanteau, Generalized Spectral	2010-2016	Weak-form efficient
Zhang *et al.* (2018)	1712	Same as Urquhart (2016), along with automatic portmanteau test, rank and sign variance ratio test, turning point test and wild-bootstrapped AVR	2013-2018	Mainly inefficient.
Bariviera (2017)	1434	R/S Hurst, DFA	2011-2017	Mainly inefficient, but signs of moving towards efficiency
Khuntia and Pattanyak (2018)	2714	MDH	2010-2017	Utilized the AMH. Market efficiency evolves with time and validates AMH

Source: Urquhart (2016), Nadarajah and Chu (2016), Zhang et.al (2018) and Bariviera (2017)

Using Urquhart's (2016) research as a foundation, Nadarajah and Chu (2016) also studied the efficiency of Bitcoin for the same period and the same form of market efficiency as well as the same frequency. However, unlike Urquhart, the tests were conducted using power transformation of the log returns by applying an odd integer power, asserting no loss of information in the process

(Nadarajah, Chu 2016). In addition to the tests by Urquhart (2016), Nadarajah and Chu (2016) used the spectral shape test, portmanteau test, and generalised spectral test for this study. Through the power transformation method of calculating returns, Nadarajah and Chu (2016) conclude that a weak form of efficiency is shown. The most comprehensive study and the largest of its kind has been conducted by Zhang *et al*. (2018), where a battery of efficiency tests from econometrics as well as econophysics have been employed to test for the inefficiency of not only Bitcoin but also other cryptocurrencies such as Ripple, Ethereum, NEM, Stellar, Litecoin, Dash, Monero and Verge at the daily frequency from the 28th of April 2013 to the 4th of January 2018. Besides some of the tests employed by Urquhart (2016) to test for randomness, this study employed the automatic portmanteau test, the Cox-Stuart test, the Mann-Kendall Rank test, Rank Score variance ratio test, Turning Point test, Variance Ratio test as well as the Wild-bootstrapped AVR test. The conclusion made by the authors of this research paper based on their findings is that every single cryptocurrency operates in inefficient markets.

Bariviera (2017) examined the long memory of returns of Bitcoin at the daily frequency from the 18th of August 2011 to the 15th of February 2017 with 1435 observations. Tests used in this study include the R/S Hurst analysis, using the sliding window methodology which expands or shrinks a subset and the Detrended-Fluctuation-Analysis (DFA) method. Between 2011 and 2014, Bariviera found signs of persistence in the time series of the study, meaning that one value of the data is closely related to the previous value. After 2014, however, Bariviera concludes that the market shows signs of being weak-form efficient, whereby the behaviour of the dataset seems to follow a white noise meaning zero autocorrelation (Bariviera 2017). Khuntia and Pattanayak (2018) were the first to study the evolution of the efficiency of Bitcoin through the Adaptive Market Hypothesis (AMH) which states that market efficiency evolves with time. This study offers contrasting results on the efficiency of Bitcoin. The study used the daily closing returns from July 18th 2010 to December 21st 2017. A total of 2714 observations were analyzed mainly using the Martingale Difference Hypothesis (MDH) using the rolling window framework. Additionally, the Ljung-Box test is employed. Through the AMH, Khuntia and Pattanayak (2017) assert that speculators can exploit excess returns during periods of inefficiency in the Bitcoin market but this is not always the case, thus validating the AMH in the Bitcoin market.

An overview of prior research using contrasting methods and theoretical frameworks to test for randomness has been discussed in this section. In the following chapter, the author provides a description of the dataset and research methodology used to answer the research question.

2. DATA AND METHODOLOGY

In this chapter, the author provides a description of the dataset used and presents the various tests adopted to test for randomness and independence of Bitcoin returns for the weak-form of EMH.

2.1. Data

The dataset titled "Bitcoin Historical Data"[2] utilised for this thesis has been extracted from Kaggle, an open-source online community for data scientists. This thesis examines the 1-minute weighted logarithmic closing price returns of Bitcoin over a full sample period and two additional sub-sample periods between the 1st of January 2014 and the 1st of January 2020. The first sub-sample ranges from the 1st of January 2014 to the 1st of January 2017, and the second sub-sample ranges from the 1st of January 2017 to the 1st of January 2020. To calculate the weighted logarithmic price returns, Equation 3 is used:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) * 100 \tag{3}$$

where
R_t – Bitcoin compound return in period t,
P_t – Bitcoin price in period t,
P_{t-1}– Bitcoin price in period t-1.

The dataset consists of almost every minute between the 1st of January 2014 and the 1st of January 2020. Nevertheless, the author of this study found that specific timestamps are missing through cleaning the data. There are jumps possibly due to the exchange down or having other technical difficulties. The price of Bitcoin and other cryptocurrencies varies across platforms, limiting the

[2] The dataset has been compiled by Dr.Zielinski with the handle @Zielak on Kaggle, a data scientist with a background in neuroscience by scraping various exchanges. The dataset has the highest usability rating at 10. The usability rating considers several important factors, such the availability of references. More on the usability rating of Kaggle data sources can be found here https://www.kaggle.com/product-feedback/93922

ability to make a fair comparison. The weighted-closing price has been used instead of the closing price as it reflects prices obtained from numerous platforms; as such, it provides a higher degree of accuracy.

2.1.1. Time series and descriptive statistics

The logarithmically scaled time-series graph illustrated in Figure 4 through Microsoft Power BI shows the weighted average closing price of Bitcoin for the entire sample period at the 1- minute frequency ranging from the 1st of January 2014 to the 1st of January 2020. It has been logarithmically scaled to illustrate a wide range of values compactly to ensure that equal visual weight is given to equally relative changes. Between 2014 to early 2015, the weighted average Bitcoin price declined, reaching an all-time low for the sample period at $156.655 on the 14th of January 2015.

Figure 4. Logarithmic Scaled time-series graph of weighted Bitcoin prices
Source: Author's visualisation based on data from Appendix 3.

The weighted prices have been steadily growing, reaching an all-time high of $19,663.299 for the entire sample period on the 17th of December 2017 before rapidly declining to $5,949.997 on the 6th of February 2018, representing an almost 70% decline in value from its previous high. Figure 4 illustrates sharp increases and decline in Bitcoin price. When prices dramatically increase, individuals, stand to immensely gain. Nevertheless, there is a big chance that individuals could lose immense amounts of money, which questions whether Bitcoin could be a viable alternative currency that could serve as a store of value.

The descriptive statistics of the weighted logarithmic price returns of Bitcoin at the 1- minute frequency over the entire sample period and the two sub-sample periods, are reported in Table 2.

Table 2. Descriptive statistics

Sample Period	N	Mean	Median	SD	Min	Max	Skewness	Kurtosis
01/01/2014 – 01/01/2020	2621848	0.000087	0.000000	0.149	-9.787	11.768	-0.232	65.148
01/01/2014 – 01/01/2017	1123849	0.000028	0.000000	0.175	-7.722	6.391	-0.387	38.239
01/01/2017 – 01/01/2020	1498838	0.000134	0.000213	0.126	-9.787	11.768	0.102	114.828

Source: Author's calculation based on data from Appendix 3

The entire sample period and the two sub-sample periods have standard deviations higher than the mean, suggestive of significant dispersions in the observations. A higher standard deviation than the mean is also shown in other studies examining Bitcoin's weighted closing price returns daily and at 5-minute intervals. For example, Urquhart (2016, 81) examined the daily closing price of Bitcoin between 2010 to 2016, and over his sample period, as well as two subsample periods, the standard deviation from Urquhart's study is much higher than the mean values. In a separate study that examined the returns of the 5-minute closing price of Bitcoin from 2014 to 2017, the standard deviation is once again much higher than the mean even when compared to the daily closing price (Eross *et al.* 2019, 76).

Skewness indicates whether a dataset is concentrated on one side or the other. The values for skewness in Table 2 for the entire sample period and the first sub-sample period indicate that the data is slightly positively skewed as the mean is more than the median. On the contrary, the data is negatively skewed for the second sub-sample period as the median is more than the mean. Overall, as all the skewness values are between -0.5 and 0.5, the data is relatively symmetrical and within a similar range to the values indicated by Urquhart (2016,81) and Zhang *et al.* (2018, 661).

The kurtosis values indicated in Table 2 for the entire sample period as well as the two sub-sample periods suggest that the distribution of the dataset is leptokurtic as it has a positive kurtosis value above 3, thus resulting in a greater chance of positive or negative events. The distribution for all three periods has a higher peak with fatter and heavier tails than a normal distribution. This means that individuals that have bought Bitcoin are likely to experience occasional extreme positive or negative returns. The highest kurtosis is exhibited in the second sub-sample from 2017 to 2020. Based on the results of Table 2, it can be said that there are no anomalies for the full-sample period and the two sub-sample periods of returns when compared to prior research by Urquhart (2016, 81), Zhang et al. (2018, 661) at the daily level and (Eross et al. 2019, 76) at the 5-minute frequency level.

2.2. Methodology

In total, over 2.6 million observations were used to determine whether the hypothesis that has been defined holds true or not, over an entire sample period and two additional subsample periods to analyze whether the efficiency level varies over time. Table 3 provides a summary of the hypotheses for the efficiency tests conducted.

Table 3. Summary of hypotheses for the efficiency test[3]

Test Name	Null Hypothesis	Alternative hypothesis
Ljung-Box test	No serial correlation.	Serial correlation.
Runs test	Data follows a sequence of randomness.	Data is not following a sequence of randomness.
Cox-Stuart test	Randomness against an upward and downward trend.	Non-randomness.
Variance ratio (Automatic Portmanteau)	Martingale (sequence of randomness).	Non-Martingale (sequence of non-randomness).

Source: Authors adaptation of Zhang *et al*. (2018, 662)

A central tenet of the EMH is that prices are not predictable and are therefore random. The tests used in this study to test for the randomness mainly replicate Urquhart (2016) and Nadarajah and

[3] The Hurst test is denoted with the Hurst exponent. Values are categorised into one of three separate categories. A further explanation is provided in the paragraph discussing the Hurst exponent within this sub-chapter.

Chu (2016), where the Ljung and Box test (1978), Runs test (1940), and R/S Hurst (1951) analysis have been applied. In addition to that, the Cox-Stuart test and Variance Ratio test have been employed in this thesis, replicating Zhang *et al.* (2018).

Autocorrelation can more formally be defined as the degree of similarity between observations in a time series. When analysing historical data, it is vital to test for autocorrelation to ensure that one observation is not correlated with another observation. To test for the autocorrelation of returns in this study, the Ljung-Box test has been employed. The Ljung-Box test is an improved version of the Box-Pierce test that statistically determines whether any given data frames have lag/delay different from zero.

The null and alternative hypotheses can be stated as follows:
H0: no serial correlation
Ha: serial correlation

Using the simplified equation by (Ljung, Box 1978), this test is conducted using the following test statistic as in Equation 4:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{p}_k^2}{n-k} \tag{4}$$

where
$n -$ is the number of observations,
$k -$ is the lag,
$\hat{p}_k^2 -$ sample autocorrelation at lag k,
$h -$ number of lags being tested.

Rejecting the null hypothesis is indicative of serial correlation and linear dependence in a time-series and as such violates one of the two principles of the random walk model and EMH.

The second test performed is the Runs test, also known as the Wald and Wolfowitz runs test, the earliest and the most straightforward test for randomness. The runs test is non-parametric; the parameters are unknown and not fixed. Another attribute of non-parametric tests is that they do not need to meet certain assumptions or parameters, unlike many other hypothesis tests, which heavily rely on the assumption that the population follows a normal distribution. It is another test designed to detect non-randomness in a data series (Wald, Wolfowitz 1940).

The null and alternative hypotheses can be stated as follows:

H0: Sequence of data is random

Ha: Sequence of data is not random

A rejection of the null hypothesis is indicative of non-randomness. The test statistic, as stated by (Wald, Wolfowitz 1940), is as follows in Equation 5:

$$Z = \frac{R - \mathbb{E}[R]}{\sigma[R]} \tag{5}$$

where

R – represents the number of runs,

$\mathbb{E}[R]$ – is the expected number of runs R,

$\sigma[R]$ – is the standard deviation of the number of runs R.

A "run" is a series of observations moving in the same direction. In a dataset where the values are random, the probability that a specific observation is higher or lower than the previous follows a binomial distribution. The sample sequence is given the symbol + if the return is above the median and – below the median. If it equals the median, the observation is discarded. Each run is a sequence with the same sign, and a run end when the sign changes.

The Hurst test is another test for market efficiency. This test has been first proposed in hydrodynamics, particularly to determine the optimum dam size for the Nile River to regulate the annual discharge of rain by analysing over 800 years of data (Hurst 1951).

The Hurst exponent values range from 0 to 1. A time series can be categorised into one of three distinct categories (Kroha, Skoula 2018, 374):

- A value of H $0 \leq 0.5$ suggests anti persistence or negative autocorrelation, meaning that another decrease shall follow a decrease between values.
- A value of H = 0.5 shows an actual random walk, meaning that, for example, the return of Bitcoin at a particular time is likely to increase or decrease with no memory of values that has preceded it in a time series.
- A value of H $0.5 \geq 1$ indicates a time series with long-term positive autocorrelation, meaning that another high value will probably follow a high value.

The Hurst exponent is defined as a rescaled range (R/S analysis), a technique to determine the nature and magnitude of variability in a time series. The Hurst test can be defined and summarised with Equation 6 (Qian, Rasheed 2005, 2):

$$\mathbb{E}\left[\frac{R(n)}{S(n)}\right] = Cn^H \tag{6}$$

where

$\mathbb{E}$ – represents the expected value,
n – data points in a time series,
$R(n)$ – range of n standard deviations from the mean,
$S(n)$ – sum of the n standard deviations,
C – constant,
H – known as the Hurst exponent.

The Hurst exponent is the "index of dependence" or "index of long-range dependence" for time series. It quantifies the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction (Hurst 1951).

The Cox Stuart test is a test for identifying monotonic trends within a data series. Monotonic trends are increasing and decreasing trends. The basic principle of this method is that an upward trend is exhibited in a series of observations if the magnitudes of the later observations tend to be greater than the earlier observations (Cox, Stuart 1955).

The null and alternative hypotheses can be stated as follows (Zhang *et al*. 2018):
H0: Randomness against an upward and downward trend.
Ha: Non-randomness.

The test statistic can be denoted with the following abbreviated example by Yong (1991, 153) with Equation 7:

$$Z = \frac{(k \pm 0.5) - 0.5n}{\sqrt{0.5n}} \tag{7}$$

where

k – refers to the number of observations of either negative or positive less observed signs,
n – numbers of pairs constructed,
0.5 – continuity adjustment factor, whereby the sign is plus if k is less than n/2 and conversely minus if otherwise.

The assumption of normality in the data distribution is not a prerequisite for this test, which adds to the robustness of the test itself and makes it much more reliable. The method is based on the binomial distribution and differences in pairs of data series.

Lastly, the Variance Ratio test introduced by Lo and Mackinlay (1989) is performed to identify whether the data series follows a random walk and is commonly used to determine whether a particular stock price exhibits autocorrelation. This test essentially tests for statistical differences between group means. The variance ratio test used in this study is the Automatic Portmanteau test for the Martingale difference hypothesis.

The conventional variance ratio posits that the variance of increments of a random walk Xt is linear; therefore, the variance of $\left(X_t - X_{t-q}\right)$ is q times the variance of $(X_t - X_{t-1})$ as stated by Chen (2008).

The null and alternative hypotheses can be stated as follows:
H0: martingale
Ha: non-martingale

The variance ratio can be defined with the following simplified Equation 8 by Chen (2011,98):

$$VR(q) = \frac{\sigma^2(q)}{\sigma^2(1)} \tag{8}$$

where
$\sigma^2(q)$ – refers to 1/q times the variance of $\left(X_t - X_{t-q}\right)$,
$\sigma^2(1)$ – refers to the variance of $(X_t - X_{t-1})$.

If a particular security price does follow a random walk, then it is stated that the ratio of the variances for the conventional variance ratio test should be equal to one. A higher than one figure would suggest a positive autocorrelation, and a figure below one would be suggestive of negative autocorrelation.

3. RESULTS AND LIMITATIONS

In this chapter, the author examines and discusses the results of the various tests conducted to determine the randomness of the logarithmic returns of Bitcoin. In addition to that, the limitations of the findings in this thesis shall also be discussed within this chapter.

3.1. Results

Results of tests used to determine weak-form market efficiency for the full sample period and two sub-sample periods are reported in Table 4.

Table 4. Test results of weak-form efficiency tests

Data frame	Ljung-Box	Runs	Cox-Stuart	R/S Hurst	Variance Ratio
01/01/2014-01/01/2020	0.000	0.000	0.065	0.531	0.000
01/01/2014-01/01/2017	0.000	0.000	0.001	0.503	0.000
01/01/2017-01/01/2020	0.000	0.000	0.000	0.538	0.000

Source: Author's calculations based on the dataset in Appendix 3

The author tests the results for the Ljung-Box test against a significance value of 0.01 and rejects the null hypothesis since the p-values are less than the significance value. This advocates some degree of autocorrelation in the data series at the given lag of 1. For the full sample period and the first-subsample period, Urquhart (2016) reports similar results at the significance value of 0.05. Nevertheless, the Ljung-Box test in Urquhart (2016) fails to reject the null hypothesis for the second sub-sample period, with a p-value of 0.35, suggesting no autocorrelation. Zhang *et al.* (2018) also fails to reject the null hypothesis for the Ljung-Box test in a similar period for Bitcoin and Litecoin. On the other hand, Nadarajah and Chu (2016) found no autocorrelation using the simple power transformation method to calculate returns even for lags 1 to 10 for the exact timeframe as Urquhart (2016).

The runs test is the second test that is performed. For all three data frames, the p-value is much lower than the significance value of 0.01, and thus the null hypothesis of randomness is rejected. This suggests that the assumptions of randomness and no autocorrelation do not hold for this data series. The test results for the Runs test in Urquhart (2016) also report the same values. Nevertheless, both Zhang *et al.* (2018) and Nadarajah and Chu (2016) fail to reject the null hypothesis for all the periods as reported in Urquhart (2016). Based on prior research, the results of this test offer contradictory results due to different periods and different methods of calculating returns.

The third test conducted is the Cox-Stuart test. The significance value that has been chosen is 0.01 (a confidence level of 99%). The p-value of the full sample period is higher than 0.01, which is not statistically significant and indicates strong evidence for the null hypothesis, consequently rejecting the alternative hypothesis. The results of the first sub-sample period in Table 4 are important as the period in Table 4 overlaps with Zhang *et al.* (2018, 663), which also fails to reject the null hypothesis for this test with a p-value of 0.08 at a significance value of 0.05. On the contrary, for both the sub-samples, the p-values are statistically significant as it is less than 0.01. A significant trend is present in the data series for both the sub-sample periods and, as such, would suggest that the data is not random; therefore, the null hypothesis is rejected.

The values of the Hurst exponent for this thesis are 0.531 for the entire sample period, 0.503 for the first sub-sample period and 0.538 for the second sub-sample period at the 0.01 significance level. The results of this test on the full sample period and the second sub-sample period suggest a long-term autocorrelation in the dataset and persistence, which violates a fundamental assumption of the random walk theory, which is that observations in a time series should not be closely related to a previous value. However, data for the first sub-sample is noteworthy as it indicates an entirely uncorrelated data series, thus failing to reject the null hypothesis.

The R/S Hurst result in Table 4 is also important as it shows similarities with the results reported by Urquhart (2016) as both periods overlap. The R/S Hurst exponent values reported by Urquhart (2016, 82) show strong levels of anti-persistence. However, the second sub-sample period in Urquhart (2016, 82), from the 1st of August 2013 to the 31st of July 2016, reported a H value of 0.406, very close to the value reported in the second-subsample period Table 4 by the author. Using the sliding window methodology, Bariviera (2017) reports contrasting results which suggest positive autocorrelation with stronger evidence of persistence, as all values of the R/S Hurst exponent are above 0.6

Lastly, the variance ratio is performed to identify whether the data series follows a random walk. Martingales refer to random variables with unpredictable future variations given the current information set (Mandelbrot 1966). The p-value for closing prices is less than 0.01, identifying a high degree of autocorrelation in the data series. Therefore, the null hypothesis that is suggestive of a martingale is rejected. The results in Table 4 significantly contrast that of Zhang *et al.* (2018, 663), where there is strong evidence for the null hypothesis for Bitcoin and all the other cryptocurrencies examined, the only exemption being the cryptocurrency Verge.

Based on the results of the findings reported in Table 4, it can conclusively be stated that the Bitcoin market is currently not weak-form efficient but could be heading towards a path of efficiency as suggested by the results of the R/S Hurst tests, and the Cox-Stuart test. There are two main reasons that could justify the inefficiency of Bitcoin. The first reason is the relative infancy of Bitcoin, and the second reason is the Herding behavior that is caused by existing and new market participants that are influenced by emotion rather than independent analysis. The Bitcoin market can be classified as an emerging market due to its number of users and market capitalization. Prior research has shown that in large emerging markets such as Brazil, Russia, India, China, and South Africa (BRICS), there are prominent signs of inefficiency (Majumder 2012). Utilizing industry-wise data for BRIC markets, and examining market efficiency with the Hurst analysis, Majumder (2012) reports that equity market returns are serially correlated as values of the Hurst exponent are either above or below 0.5. Herding in Finance can broadly be defined as the imitation of an investor's actions of others and could also explain the inefficiency of the Bitcoin market (Merli, Roger 2013). Using cross-sectional absolute deviation of returns (CSAD), Tomas *et al.* (2019) reports evidence of herding in the cryptocurrency market during market downturns which signals inefficiency.

In this sub-chapter, the results of the tests were reported and discussed. In the subsequent sub-chapter, the author shall discuss the limitations that have to be considered when interpreting the results of the findings.

3.2. Limitations

This thesis mainly follows the methodology utilised by Urquhart (2016) and expands the study period to cover the years until 01/01/2020. An ideal test procedure for the author would have been to extract data at the 1-minute frequency from bitcoinaverage.com. The reason for doing so is because it is the same exchange used by (Urquhart 2016). However, access to the dataset used by Urquhart (2016) and other commonly used datasets from coinmarketcap.com, such as Zhang et al. (2018), is very costly[4]. The accuracy of the results could be further improved if the same dataset was used.

The cryptocurrency market is in its infancy. The inclusion of additional cryptocurrencies such as Ethereum, Binance and others could have also provided a deeper analysis of the findings of the thesis. The main reason for not doing so for this thesis is the lack of available historical data for the selected frequency and time frame as newer cryptocurrencies such as Ethereum and Binance were introduced later.

Another limitation to the results reported is that Urquhart (2016) and Zhang *et al.* (2018) do not provide much detail in the descriptions of the test procedures. Different tests require different parameters and specifications, such as determining increments for the R/S Hurst analysis. Tests were conducted based on the author's knowledge of R, which may be calibrated differently than the test procedures used by Urquhart (2016) and other research. Consequently, the potential for inconsistency in the test procedures may cause differing results relative to Urquhart (2016) as well as other similar research.

A theoretical limitation of the EMH is that it considers market efficiency as a static phenomenon. This significantly contrasts the AMH, which asserts that markets continuously evolve due to structural changes (Lo 2004). Unlike the stock market, which operates for a defined period, the Bitcoin Market lacks supervision and works round the clock; as such, it is evolving exponentially faster. A lack of prior research on the efficiency of Bitcoin from the perspective of the AMH prompted the author to use the more established and dominant EMH as a theoretical framework.

[4] An API for 12 months of historical data on coinmarketcap.com costs $699 per month with limited functionality and $56 on bitcoinaverage.com for a single file download when authoring this thesis.

CONCLUSION

This thesis examined whether the Bitcoin market is efficient by testing the weak-form of the Efficient Market Hypothesis (EMH). The data used in this thesis was extracted from Kaggle and mainly replicated the work of Urquhart (2016) and expanded to include additional tests as used in Zhang *et al.* (2018) with a larger dataset at a higher frequency. The 1-minute weighted logarithmic closing price returns of Bitcoin were examined over a full sample period and two additional sub-sample periods between the 1st of January 2014 and the 1st of January 2020.

To reiterate, the main research question of this thesis was:

- Are Bitcoin returns weak-form efficient according to the Efficient Market Hypothesis?

Based on the research question, the following hypothesis was put forth:

- Bitcoin price return follows a sequence formed by a random walk.

For a market to be weak-form efficient, returns must follow a random walk process. Therefore, the Ljung-Box test, Runs test, Cox-Stuart test, R/S Hurst test and Variance ratio test were employed to test if the Bitcoin market meets the two fundamental assumptions of the random-walk model, which are first, whether successive price changes or returns are independent and secondly whether price changes or returns are identically distributed.

Based on the results of the tests that have been employed, Bitcoin returns are shown to be inefficient. Therefore, the null hypothesis is rejected at the 0.01 significance level for the entire sample period, the first sub-sample period and the second sub-sample period for the Runs test, the Ljung-Box test and the Variance ratio test. However, the full sample period of the Cox-Stuart test and the second-subsample R/S Hurst exponent fails to reject the null hypothesis at the 0.01 significance level. Conclusively, it can be stated that the Bitcoin market is currently not weak-form efficient but might be heading towards a path of efficiency based on the results of the Cox-Stuart and Hurst test.

Bitcoin and cryptocurrencies, in general, are still in their infancy. Nevertheless, the popularity of digital currencies such as Bitcoin is soaring from relative obscurity over a decade ago when it was incepted. It is safe to state that Bitcoins usage has entered mainstream society. Whilst no large, developed country has adopted Bitcoin as a legal tender or alternative currency, sovereign nations such as El Salvador and, most recently, the Central African Republic have adopted it as a legal tender.

Results of this thesis as well as prior and future research are vital to understanding the price movements of the Bitcoin market as they could potentially assist legislators and economists at a macro level in determining the viability of Bitcoin and perhaps other digital currencies as alternative currencies. The results of this thesis are also crucial to individuals who are keen on utilizing Bitcoin either as a store of value or a medium of exchange.

In a couple of years, as the usage and number of Bitcoin users continue to grow, results on the efficiency of Bitcoin may be different. Researchers must continue the debate on the efficiency of Bitcoin and, in general, cryptocurrencies to uncover possible anomalies within the cryptocurrency market. A proposal for further research would be to include a more extensive testing period that will enable further comparisons with other well-established cryptocurrencies. As more data becomes available, another proposal for future research would be to analyse the efficiency of Bitcoin from the perspective of the AMH.

LIST OF REFERENCES

Appukuttan Nair, D. (2019). The bitcoin innovation, cryptocurrencies and the Leviathan. *Innovation and Development*, 9(1), 85–103.

Arslanian, H., Donovan.R., Blumenfeld, M., Zamore, A. (2021). *El Salvador's law: a meaningful test for Bitcoin.* Retrieved from https://www.pwc.com/gx/en/financial-services/pdf/el-salvadors-law-a-meaningful-test-for-bitcoin.pdf, the 10th of April 2022.

Antonopoulos, A. (2015). *Mastering Bitcoin: unlocking digital cryptocurrencies.* (1st ed.) California, USA: O'Reilly Media Inc.

Browne, R. (2022). *Central African Republic becomes second country to adopt bitcoin as legal tender*. Retrieved from https://www.cnbc.com/2022/04/28/central-african-republic-adopts-bitcoin-as-legal-tender.html, the 29th of April 2022.

Brown, R. (1828). A Brief Account of Microscopical Observations Made in the Months of June, July and August, 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies. *Philosophical Magazine*, 4, 161-173.

Bartels, Robert. (1982). The Rank Version of von Neumann's Ratio Test for Randomness. *Journal of the American Statistical Association*, 77 (377), 40–46.

Bariviera, Aurelio F. (2017). The Inefficiency of Bitcoin Revisited: A Dynamic Approach. *Economics Letters*, 161, 1–4.

Bachelier, L. (1900). Théorie de La Spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 17, 21–86.

Chaum, D. L. (1982). Computer systems established, maintained and trusted by mutually suspicious groups (Doctoral dissertation, University of California, Berkeley).

Chen, J.-H. (2011). Variance ratio tests of random walk hypothesis of the euro exchange rate. *International Business & Economics Research Journal (IBER)*, *7*(12). 97-106.

Chu, J., Zhang, Y., Chan, S. (2019). The adaptive market hypothesis in the high frequency cryptocurrency market. *International Review of Financial Analysis*, 64, 221–231.

Ciaian, P., Rajcaniova, M., Kancs, d'Artis. (2016). The economics of BitCoin price formation. *Applied Economics*, 48(19), 1799–1815.

Cox, D. R., Stuart, A. (1955). Some Quick Sign Tests for Trend in Location and Dispersion. *Biometrika*, 42 (1/2). 80-95.

Delcey, Thomas. (2019). Samuelson vs Fama on the Efficient Market Hypothesis: The Point of View of Expertise. *OEconomia*, 9 (1), 37–58.

Degutis, A., Novickytė, L. (2014). The efficient market hypothesis: A critical review of literature and methodology. *Ekonomika*, *93*(2), 7–23.

Decker, C., Wattenhofer, R. (2013). Information propagation in the Bitcoin network. *IEEE P2P 2013 Proceedings*, pp. 1-10.

Eross, A., McGroarty, F., Urquhart, A., Wolfe, S. (2019). The intraday dynamics of bitcoin. *Research in International Business and Finance*, 49, 71–81.

Fama, Eugene F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21 (5), 55–59.

Fama, Eugene F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25, 383-416.

Grossman, Sanford J., Stiglitz, Joseph E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70 (3), 393–408.

Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, 116 (1), 770–99.

Haber, S., Stornetta, W.S. (1991). How to timestamp a digital document. *Journal of Cryptology*, 3 (2), 99-111.

Judgement of 22 October 2015, Skatteverket v David Hedqvist, C-264/14, EU:C:2015:718, paragraph 6.

Kang, H.-J., Lee, S.-G., & Park, S.-Y. (2022). Information efficiency in the cryptocurrency market:the efficient-market hypothesis. *Journal of Computer Information Systems*, *62*(3), 622–631.

Keshari Jena, S., Tiwari, A. K., Doğan, B., & Hammoudeh, S. (2020). Are the top six cryptocurrencies efficient? Evidence from time-varying long memory. *International Journal of Finance & Economics*, 1-11.

Kroha, P., & Škoula, M. (2018). Hurst exponent and trading signals derived from market time series: *Proceedings of the 20th International Conference on Enterprise Information Systems*, 371–378.

Khuntia, S., & Pattanayak, J. K. (2018). Adaptive market hypothesis and evolving predictability of bitcoin. Economics Letters, 167, 26–28.

Ljung, G. M., G. E. P. Box. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65 (2), 297–303.

Lo, Andrew W., MacKinlay A. Craig. (1989). The Size and Power of the Variance Ratio Test in Finite Samples. *Journal of Econometrics*, 40 (2), 203–38.

Lo, Andrew W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, 30 (5), 15–29.

Lo, Andrew W. (2017). Adaptive Markets: Financial Evolution at the Speed of Thought. (1st ed.) New Jersey, USA: Princeton University Press.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Mandelbrot, B. (1966). Forecasts of future prices, unbiased markets, and "martingale" models. *The Journal of Business*, *39*(1), 242.

Merkle, Ralph C. (1980). Protocols for Public Key Cryptosystems. IEEE Symposium on Security and Privacy, *IEEE*,122–122.

Merli, M., & Roger, T. (2013). What drives the herding behavior of individual investors?. *French Finance Association*, Vol. 34(3), 67–104.

Mehta, N., Agashe, A., & Detroja, P. (2021). *Bubble or revolution: The present and future of blockchain and cryptocurrencies*. (2nd ed.) USA: Paravene Ventures.

Majumder, D. (2012). When the market becomes inefficient: Comparing BRIC markets with markets in the USA. *International Review of Financial Analysis*, 24, 84–92.

Nadarajah,S., Chu J. (2017). On the Inefficiency of Bitcoin. *Economics Letters*, 150, 6–9.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. 1-8.

Pearson, K. (1905). The Problem of the Random Walk. *Nature*, 72, 294.

Pitta, J. (1999). *Requiem for a Bright Idea*. Retrieved from https://www.forbes.com/forbes/1999/1101/6411390a.html, the 10th of April 2022.

Phlips, L. (1988). *The economics of imperfect information*. (1st ed.) New York, USA: Cambridge University Press.

Popper, N. (2021). *Lost Passwords Lock Millionaires Out of Their Bitcoin Fortunes*. Retrieved from https://www.nytimes.com/2021/01/12/technology/bitcoin-passwords-wallets-fortunes.html, the 10th of April 2022.

Qian B., Rasheed K. (2004). Hurst Exponent and Financial Market Predictability. Proceedings of the IASTED conference on Financial Engineering and Applications, 203–209.

Rayleigh, L. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine*, 10, 73-78.

Samuelson, P. (1965). Rational Theory of Warrant Pricing. *Industrial Management Review*, 6, 13-31.

Smith, G. (2012). The changing and relative efficiency of European emerging stock markets. *The European Journal of Finance*, *18*(8), 689–708.

Sompolinsky, Y., Zohar, A. (2017). Bitcoin's Underlying Incentives: The unseen economic forces that govern the Bitcoin protocol. *Queue*, 15(5), 29–52.

Tomás, D., Ibáñez, A. M., & Farinós, J. E. (2019). Herding in the cryptocurrency market: CSSD and CSAD approaches. Finance Research Letters, 30, 181–186.

Urquhart, Andrew. (2016). The Inefficiency of Bitcoin. *Economics Letters*, 148, 80–82.

Vergne, J. (2020). Decentralised vs. Distributed organisation: Blockchain, machine learning and the future of the digital platform. *Organisation Theory*, *1*(4), 1-22.

Vigna, P., Casey, M. (2016). *The age of cryptocurrency: How bitcoin and the blockchain are challenging the global economic order*, (1st ed.) New York, USA: St.Martin's Press.

Yong, O. (1991). Randomness of stock market movement: A nonparametric approach. *Jurnal Pengurusan*, *10*, 49–58

Wald, A., Wolfowitz, J. (1940). On a Test Whether Two Samples Are from the Same Population. *The Annals of Mathematical Statistics*, 11 (2), 147–62.

Wang, K. (2021). Measuring Global Crypto Users A Study to Measure Market Size Using On-Chain Metrics. Retrieved from https://crypto.com/eea/, 10th April 2022.

Zhang, W., Wang, P., Li, X., & Shen, D. (2018). The inefficiency of cryptocurrency and its cross-correlation with Dow Jones Industrial Average. *Physica A: Statistical Mechanics and Its Applications*, 510, 658–670.

APPENDICES

Appendix 1: Source code for efficiency tests

```
library(tidyverse)
library(patchwork) # To display 2 charts together
library(hrbrthemes)
library(lubridate)
library(snpar)
library(trend)
library(vrtest)
library(tseries)
library(pracma)
library(data.table)
library(tseries)
library(fBasics)
library(reshape2)
library(randtests)
library(trend)

options(scipen = 999)

# Load the data using data.table
price_data <- fread("bitstamp_cleaned.csv")

# Split DateTime variable into 2 new variables, date and time
price_data[, c("date", "time") := tstrsplit(DateTime, " ")]
```

Calculate Returns and Descriptive Statistics

```
price_data[, btc_ret := log(price_data$Weighted_Price / dplyr::lag(price_data$Weighted_Price))*100]
price_data <- price_data[-1,]

basicStats(price_data$btc_ret)

price_data[, date2 := mdy(as.character.Date(date))]
price_data[, time2 := hms(time)]

close_1 <- price_data[date2 >= "2014/1/1" & date2 <= "2020/1/1", "btc_ret"]
close_2 <- price_data[date2 >= "2014/1/1" & date2 <= "2017/1/1", "btc_ret"]
close_3 <- price_data[date2 >= "2017/1/1" & date2 <= "2020/1/1", "btc_ret"]
```

```
lapply(list(close_1$btc_ret, close_2$btc_ret, close_3$btc_ret), basicStats)

price_data[, date_time_ := as.POSIXct(Unix_Timestamp, origin = "1970/01/01")]

price_data[, date_time_ := date_time_ + hours(6)]

setnames(price_data, c("Volume_(BTC)", "Volume_(Currency)"), c("volume_btc", "volume_currency"))

plot_data <- price_data %>% select(date_time_, Close, volume_btc, date2, time2)
```

Ljung-Box Test

```
lapply(list(close_1, close_2, close_3), function(x)
  Box.test(x, lag = 1, type = "Ljung-Box", fitdf = 0))
```

Runs test

```
lapply(list(close_1, close_2, close_3), function(x)
  snpar::runs.test(unlist(x), exact = FALSE))
```

Hurst test

```
lapply(list(close_1, close_2, close_3), function(x)
  hurstexp(x$btc_ret, d = floor(nrow(x)/50000)))
```

Cox-Stuart Test

```
lapply(list(close_1, close_2, close_3), function(x)
  cox.stuart.test(x$btc_ret))
```

Variance Ratio test

```
lapply(list(close_1, close_2, close_3), function(x)
  vrtest::Auto.Q(x$btc_ret))

vrtest::Auto.Q(close_1$btc_ret)
```

Appendix 2: Summary results

Descriptive statistics:

```
[1]]
            X...X.i
nobs     2621848.000000
NAs            0.000000
Minimum       -9.786751
```

Maximum	11.767972
1. Quartile	-0.045125
3. Quartile	0.047473
Mean	0.000087
Median	0.000000
Sum	228.309029
SE Mean	0.000092
LCL Mean	-0.000093
UCL Mean	0.000268
Variance	0.022222
Stdev	0.149069
Skewness	-0.232767
Kurtosis	65.148350

[[2]]

	X...X.i
nobs	1123849.000000
NAs	0.000000
Minimum	-7.722283
Maximum	6.390522
1. Quartile	-0.046669
3. Quartile	0.050244
Mean	0.000028
Median	0.000000
Sum	30.960325
SE Mean	0.000165
LCL Mean	-0.000296
UCL Mean	0.000351
Variance	0.030589
Stdev	0.174898
Skewness	-0.387157
Kurtosis	38.239221

[[3]]

	X...X.i
nobs	1498838.000000
NAs	0.000000
Minimum	-9.786751
Maximum	11.767972
1. Quartile	-0.044415
3. Quartile	0.046175
Mean	0.000134
Median	0.000213
Sum	200.544731
SE Mean	0.000103
LCL Mean	-0.000068
UCL Mean	0.000336
Variance	0.015938
Stdev	0.126247
Skewness	0.102204

Kurtosis 114.828089

Ljung-Box

[[1]]

```
	Box-Ljung test

data:  x
X-squared = 28711, df = 1, p-value < 0.00000000000000022
```

[[2]]

```
	Box-Ljung test

data:  x
X-squared = 65097, df = 1, p-value < 0.00000000000000022
```

[[3]]

```
	Box-Ljung test

data:  x
X-squared = 12439, df = 1, p-value < 0.00000000000000022
```

Runs

[[1]]

```
	Approximate runs rest

data:  unlist(x)
Runs = 1343754, p-value < 0.00000000000000022
alternative hypothesis: two.sided
```

[[2]]

```
	Approximate runs rest

data:  unlist(x)
Runs = 596633, p-value < 0.00000000000000022
alternative hypothesis: two.sided
```

[[3]]

```
	Approximate runs rest

data:  unlist(x)
Runs = 746029, p-value = 0.00000003003
alternative hypothesis: two.sided
```

Cox-Stuart

```
[1]]

	Cox Stuart test

data:  x$btc_ret
statistic = 653992, n = 1310096, p-value = 0.06514
alternative hypothesis: non randomness


[[2]]

	Cox Stuart test

data:  x$btc_ret
statistic = 280673, n = 558785, p-value = 0.0006156
alternative hypothesis: non randomness


[[3]]

	Cox Stuart test

data:  x$btc_ret
statistic = 371030, n = 749246, p-value < 0.00000000000000022
alternative hypothesis: non randomness
```

Hurst

```
Simple R/S Hurst estimation:         0.5308135
Corrected R over S Hurst exponent:   0.5439876
Empirical Hurst exponent:            0.5301213
Corrected empirical Hurst exponent:  0.5156832
Theoretical Hurst exponent:          0.5116384
Simple R/S Hurst estimation:         0.5028798
Corrected R over S Hurst exponent:   0.522745
Empirical Hurst exponent:            0
Corrected empirical Hurst exponent:  0
Theoretical Hurst exponent:          0.5164273
Simple R/S Hurst estimation:         0.5382099
Corrected R over S Hurst exponent:   0.5407297
Empirical Hurst exponent:            0.5297387
```

```
Corrected empirical Hurst exponent:  0.5138916
Theoretical Hurst exponent:          0.5147399

[[1]]
[[1]]$Hs
[1] 0.5308135

[[1]]$Hrs
[1] 0.5439876

[[1]]$He
[1] 0.5301213

[[1]]$Hal
[1] 0.5156832

[[1]]$Ht
[1] 0.5116384


[[2]]
[[2]]$Hs
[1] 0.5028798

[[2]]$Hrs
[1] 0.522745

[[2]]$He
[1] 0

[[2]]$Hal
[1] 0

[[2]]$Ht
[1] 0.5164273
[[3]]
[[3]]$Hs
[1] 0.5382099

[[3]]$Hrs
[1] 0.5407297

[[3]]$He
[1] 0.5297387

[[3]]$Hal
[1] 0.5138916

[[3]]$Ht
[1] 0.5147399
```

Variance Ratio Test

```
+   vrtest::Auto.Q(x$btc_ret))
[[1]]
[[1]]$Stat
[1] 1637.44

[[1]]$Pvalue
[1] 0


[[2]]
[[2]]$Stat
[1] 3590.862

[[2]]$Pvalue
[1] 0


[[3]]
[[3]]$Stat
[1] 917.3935

[[3]]$Pvalue
[1] 0


>
> vrtest::Auto.Q(close_1$btc_ret)
$Stat
[1] 1637.44

$Pvalue
[1] 0
```

Appendix 3: Dataset

The dataset can be downloaded from Kaggle here but would need to be cleaned and prepped to be used in the data studio:
https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data

Appendix 4: Non-exclusive license

A non-exclusive licence for reproduction and publication of a graduation thesis[5]

I Ganenthra Ravindran (*author's name*)

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis

MARKET EFFICIENCY OF BITCOIN: EVIDENCE FROM THE EFFICIENT MARKET HYPOTHESIS (EMH)

supervised by Simona Ferraro PhD

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

12/05/2022 (date)

[5] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*