**CHARLES UNIVERSITY**
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

# Exploring the relationship between Bitcoin price and the network's hashrate

Bachelor thesis

Author: Jan Kubal
Study program: Economics and Finance
Supervisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

Year of defense: 2021

## Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, July 27, 2021

Jan Kubal

# Abstract

The goal of the thesis is to identify factors that drive the price of Bitcoin and the hashrate of the Bitcoin network, which represents the total computing power dedicated to Bitcoin mining, and to explore the relationship between these two variables. In the Bitcoin system, four variables were assumed to be endogenous, thus for each of them, an equation was constructed. This was the case of the Bitcoin price, the hashrate of the Bitcoin network, the total transaction fees paid, and the search volume for the term "bitcoin". The system of four equations was then simultaneously estimated, utilizing the method of Two-stage least squares.

Results revealed several statistically significant explanatory variables of the price and the hashrate, including the money supply of the United States dollar or the number of unique active addresses on the Bitcoin network. The hashrate was shown to drive the price positively, however, the estimated effect of the price on the hashrate was statistically insignificant. It was argued that it might have been caused by exogenous shocks affecting multiple variables, that could not be accounted for in the data. In addition, the factors affecting the hashrate were assessed from the environmental point of view, as the high environmental impact is one of the main points in the criticism of Bitcoin.



**Keywords** Bitcoin, Bitcoin mining, Bitcoin price, hashrate, cryptocurrency


**Title** Exploring the relationship between Bitcoin price and the network's hashrate


# Abstrakt

Cílem této práce je identifikovat faktory ovlivňující cenu Bitcoinu a hashrate Bitcoinové sítě, který reprezentuje celkovou výpočetní sílu využitou na těžbu Bitcoinu, a prozkoumat vztah těchto dvou proměnných. Čtyři proměnné byly považovány za endogenní, proto pro každou z nich byla sestavena jedna rovnice. Šlo o cenu Bitcoinu, celkový hashrate Bitcoinové sítě, celkový objem transakčních poplatků a objem vyhledávaní termínu "bitcoin". Systém rovnic byl poté jednotně odhadnut pomocí Dvoustupňové metody nejmenších čtverců.

Výsledky odhalily několik statisticky signifikantních proměnných vysvětlujících cenu a hashrate, včetně peněžní zásoby Amerického dolaru nebo počtu unikátních aktivních adres v Bitcoinové síti. Ukázalo se, že hashrate pozitivně ovlivňuje cenu, nicméně odhadnutý efekt ceny na hashrate byl statisticky nesignifikantní. To mohlo být způsobeno exogenními šoky působícími na vícero proměnných, které nemohly být zohledněny v datech. V závěru byly posouzeny faktory ovlivňující hashrate z environmentálního hlediska, protože negativní dopad na životní prostředí je jedním z hlavních bodů v kritice Bitcoinu.

**Klíčová slova** Bitcoin, těžba Bitcoinu, cena Bitcoinu, hashrate, kryptoaktiva

**Název práce** Zkoumání vztahu mezi cenou Bitcoinu a hashratem sítě

## Acknowledgments

The author is thankful especially to prof. PhDr. Ladislav Krištoufek, Ph.D.for his patience and guidance during the writing of the thesis, and above all for his invaluable help with the methodological parts of the work.

Additionally, a great deal of thanks belongs to the author's family and friends, namely to Kamila Kyzlíková and Vojtěch Švandelík, for their support.

Typeset in FSV LaTeX template with gratitude to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

**Bibliographic Record**

Kubal, Jan: *Exploring the relationship between Bitcoin price and the network's hashrate.* Bachelor thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2021, pages 74. Advisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Bitcoin is a decentralized payment network that emerged in 2009. It allows users to send and receive transactions with a high level of security. The unit of currency used in the network, also called Bitcoin, has substantially risen in price since the inception of the network, as the popularity of Bitcoin experienced several booms. However, the price of Bitcoin remains extremely volatile compared to traditional assets.

The process of creation of new Bitcoin units, called mining, is done by computer hardware. The hardware has to solve complex mathematical problems in order to mine Bitcoin, and in return, the person running the hardware has a chance to receive a reward of a predetermined number of Bitcoins. The more computing power a person dedicates to mining, the higher chance of receiving the reward they have. Therefore, the growth of the Bitcoin price led to the development of a very competitive mining industry.

One attempt of the mining hardware to resolve the mentioned mathematical problem is called a hash. Based on this, the worldwide combined computing power dedicated to Bitcoin mining is called hashrate (*i.e.* how many hashes are being performed per unit of time in total). As the total hashrate grew, Bitcoin has been repeatedly criticized for its high electricity consumption (De Vries (2020); Mora *et al.* (2018)), however, opposing voices also appeared (Bevand, 2017a).

Various authors also analyzed the factors that drive the Bitcoin price. Among others, the hashrate was assumed to be an important technological factor, that might have an effect on the price, however only a weak relationship was found (Kristoufek (2015; 2020)). On the other hand, the opposite relationship was proved, the price affecting hashrate (Fantazzini & Kolodin

(2020); Kristoufek (2020)). Both directions of a causal effect are thus plausible, hashrate weakly affecting the price and the price strongly affecting the hashrate. In fact, the latter can be expected as the revenue of miners depends strongly on the price of Bitcoin. The other factors that were shown to be important drivers of price are *e.g.* the number of active users of the Bitcoin network, the public interest in Bitcoin (represented by the search volume from an online search engine), or the price level of Bitcoin. The ability of Bitcoin to serve as a safe haven was also examined, however, there is no clear consensus in the academic literature.

The objective of this thesis is to identify the main drivers of the Bitcoin price and of the total network hashrate and to find, whether there is a clear relationship between these two variables. Since the hashrate is an important technological factor of the Bitcoin system and results regarding it vary, findings of the proposed analysis might be useful in analyzing the Bitcoin price movements, as well as they might enrich the debate regarding the effect of mining Bitcoin on the environment.

Four equations were created, explaining four variables that were assumed to be endogenous in the system (including the price and the hashrate). Due to endogeneity, the Two-stage least squares estimator was used to regress the model. To address the question of whether Bitcoin can serve as a hedge against inflation, the US dollar money supply, a variable that was not analyzed in the context of Bitcoin before, was included as one of the explanatory elements of the changes in the price. Results revealed several factors being significant drivers of the price and hashrate and the hashrate was found to affect the price positively, according to expectation. However, the price was statistically insignificant while explaining the hashrate changes. This was presumably caused by exogenous shocks affecting some variables, that can not be accounted for in the data. Despite that, the analysis yields useful results in form of identified factors affecting the Bitcoin price and the network hashrate and subsequent discussion of the revealed relationships from the environmental point of view.

The remainder of the thesis is structured as follows. Chapter 2 provides a detailed literature review of papers concerned with a question of the environmental impact of Bitcoin mining and with factors that drive the price of Bitcoin and the hashrate of the network. Chapter 3 provides a motivation for the empirical part of the thesis, proposes a model that shall be used for the analysis, discusses the methodology to be employed, and describes the process of collecting data. Chapter 4 describes various statistical tests that were used so the analysis could be considered reliable and presents the results of the analysis. Chapter 5 provides a discussion of the results and their implications, interprets results through the lens of the environmental impact, and describes the limitations of the thesis, as well as topics for eventual further research. Chapter 6 summarizes the findings of the thesis.

# Chapter 2

# Literature review

Bitcoin, created by Nakamoto (2008), is a decentralized network that allows users to conduct transactions safely over the internet without the need of a third party as a provider of trust. The process that allows the Bitcoin network to run is called mining, as it issues new Bitcoins and releases them into circulation. This chapter provides a brief summary of the mining of Bitcoin, its history, mechanisms, and problems it faces, and further continues with a review of articles concerned with estimating of environmental impacts of bitcoin, whereas these are estimates on electricity demand or tons of $CO_2$ produced. It concludes with a summary of papers analyzing factors that drive the price of bitcoin and the total hashrate of the network.

## 2.1 Summary of the mining process and its evolution

Bitcoin is a type of digital asset. It can be held or traded over the internet, purely in a digital form. Transactions can be sent and received thanks to the decentralized network of nodes. The decentralization originates from the fact that there is no central authority allowing or denying rights, any user with proper hardware can become a node in a network or participate in Bitcoin mining. Transactions are being saved in the form of blocks into the blockchain, an online ledger, that is shared by all nodes. The mining is done by miners[1], who run computers solving mathematical problems as fast as possible, with the hope of being the first to find a solution to a problem, that is specified by

[1]Somewhat confusingly, people owning or operating mining hardware are called miners, but the specialized hardware units used for mining are called miners as well.

the Bitcoin network. Only after this solution is found, a new block containing incoming transactions can be written into the blockchain and a miner who found it (and thus closed the block) is rewarded with a certain amount of Bitcoin. The system is set such that, on average, a new block is added every ten minutes. This is ensured by a variable called Difficulty, which affects the likelihood of a miner to find a solution to the aforementioned problem. The Difficulty changes dynamically every 14 days according to the total computational power that miners expend, in order to keep the average of ten minutes per block(Nakamoto, 2008).

The result of mining, similarly as in the mining of gold, is obtaining new units of the mined asset, therefore new Bitcoins. However, unlike in the mining of gold, Bitcoin mining plays an important role in protecting the Bitcoin system, as an attacker would need to have at least the same computing power as all miners combined to do changes from which he could profit (so called 51% attack). Currently, due to the enormous combined computing power of the Bitcoin network, it is highly unlikely that an individual or a single organization would be capable of such an attack. However, Bastiaan (2015) argues that mining pools (*e.g.* F2Pool, AntPool, ViaBTC *etc.*) that group miners together might have a significant amount of the hashing power, and if organized, they could pose a threat.

Miners are rewarded for their effort by newly created Bitcoins and transaction fees. From each sent transaction, a small fee is taken and paid to a miner who found the correct answer to the system algorithm task and thus closed the block. The amount of newly created Bitcoin was initially set to 50 Bitcoins per block, but the reward halves approximately every four years (every 210,000 blocks mined). Thus, after the first halving in 2012, miners received 25 Bitcoins per mined block instead of 50 and after 2016, the reward further halved to 12.5 Bitcoins. The last halving happened in May 2020 and now, miners receive 6.25 Bitcoin per block. This also means that there is a limited amount of Bitcoin to be mined (21 million in total), as the reward will converge to 0 (as explained *e.g.* by Antonopoulos (2014)).

The Bitcoin network came into existence on January 3, 2009, as the first block (also called the Genesis block) was mined by Satoshi Nakamoto. At first, standard CPUs (Central Processing Units) from common computers were used for Bitcoin mining, but they were soon replaced by GPUs (Graphical Processing Unit), as in 2010 a programming language OpenCL was released and allowed for the hardware to be altered to better suit the miners' needs (Narayanan *et al.*, 2016). In 2011, another technological progress of Bitcoin mining was made, as FPGAs (Field Programmable Gate Arrays) started to be employed, because they provided better computing power. The largest step forward in the mining industry was the introduction of the first ASICs (Application-Specific Integrated Circuits) developed specifically for the needs of the mining of Bitcoin, at the end of 2013. It meant a large boost in the computing power, as well as in the efficiency of the electricity used for computing (Courtois *et al.*, 2014). The computing power of the Bitcoin network is called hashrate, *i.e.* how many hashes[2] per second is performed by all the mining machines combined. The computing power of a single mining unit is expressed by how many hashes per second it could perform when running on 100% of its capacity and its efficiency can be measured in J/Gh (joules per gigahash, how many joules of energy a machine needs to perform 1 billion of hashes).

As the popularity of Bitcoin grew, also its price was rising and in turn, the profitability of mining was increasing. This attracted a large number of investors, seeking profits from the selling of the mined Bitcoin. This and the evolution of the mining hardware led to the professionalization of the mining industry. Miners transformed from the Bitcoin enthusiasts mining on their home computers to large mining farms, running numerous mining units in an industrial-like manner and seeking locations with the cheapest electricity (the professionalization of the mining industry describes *e.g.* Blandin *et al.* (2020)).

This increase in the scale of mining inevitably led to an increase in the electricity consumed by the mining hardware, which became one of the main points in the criticism of Bitcoin. The problem of the electricity consumption of Bitcoin mining is extremely complex, from the normative perspective (is it justifiable to use so much energy to run the Bitcoin network?) as well as from the descriptive perspective (how much electricity is consumed by Bitcoin mining?). It is hard to quantify the amount of electricity consumed and its environmental effect because many variables are unknown. Important factors, such as which hardware is used exactly, location and energy sources of miners, or the exact mix of expenditures (electricity costs, capital expenditures, and other overhead costs) can be only speculated.

The next section provides an outline of works that tried to estimate these and similar factors. This topic is closely connected to the hashrate of the Bitcoin network and is important for a broad understanding of the Bitcoin

[2] One hash can be seen as one attempt to submit an answer to a question posed by the Bitcoin network.

topic. Although it is not connected to the price of Bitcoin directly, it is a necessity to keep it in mind when discussing the drivers of hashrate.

## 2.2 Research on the electricity consumption of Bitcoin mining

Bitcoin mining was in the academic literature explored only to a limited degree, with not many papers concerned with it. Most of the early research was analyzing technical aspects of Bitcoin and security questions of the network, later the focus shifted mainly towards financial aspects of Bitcoin, such as drivers of its price or its ability to serve as a currency, *e.g.* Androulaki *et al.* (2013); Kristoufek (2013; 2015). Questions about the electricity consumption of miners and its effect on the environment started to appear more significantly only after 5 years of the network working, as the amount of electricity consumed grew.

Papers tackling the issue of the electricity demand could be divided into three groups, the first being estimates of the electricity consumption at one particular point in time, the second group containing papers that estimate the electricity consumption and also consider the emissions that production of electricity used for Bitcoin mining causes, and the third group represents the estimates of the electricity consumption that are continuous through time (the third group being considerably smaller than the previous two). The next three subsections provide an overview of papers in the specified order.

### 2.2.1 Estimates of the electricity demand

One of the first, if not the first, papers trying to estimate the electricity consumption of cryptomining was published by Malone & O'Dwyer (2014). In their study authors concluded that Bitcoin mining demanded power from the range between 0.01 GW and 10 GW at the time of writing the paper and further presented their best plausible estimate of 3 GW, which was at that time a power consumption comparable to Ireland. For the power estimation, authors divided the total network hashrate by the power efficiency (in MH/J) of the best available mining hardware and by the power efficiency of the hardware that was still marginally profitable to use for the mining in order to compute the upper and lower bounds respectively (Malone & O'Dwyer, 2014). Due to the data unavailability, authors neglected additional electricity expenses, such as power required for cooling the mining hardware, and also took an assumption on the electricity price. Despite that, they marked a path in the previously unmapped territory for further research.

Different approach than Malone & O'Dwyer chose De Vries (2018). In this paper, De Vries decided to avoid uncertainty arising from the use of the data on the efficiency of mining hardware. This uncertainty might be present in a form of an unknown mix of various mining units, which together compose the total hashrate of the Bitcoin network. It is not possible to track devices on which the Bitcoin is being mined and a researcher trying to find the Bitcoin electricity consumption may either estimate bounds, between which the true consumption might lie, create a few scenarios that would represent various compositions of mining hardware, or can simply make an educated guess on the mix of the hardware used. Instead of this, De Vries estimated energy consumption of Bitcoin mining using the data on Bitcoin price and rewards received by miners, thus obtaining miners' total revenue in US Dollars. Then, he took 60% of this revenue, assuming that miners' expenditures are composed of electricity costs (including electricity for mining hardware, as well as cooling and other expenditures) of 60%. Finally, total electricity expenditures were divided by 0.05, assuming the price of electricity for all miners being 5 USD cents per kWh. Additionally, the author estimated the lower bound of the electricity consumption by dividing the total hashrate by the efficiency of the best mining hardware produced at the time (Antimenr S9 produced by Bitmain). This is though unlikely to be the true electricity consumption, due to a low supply of and a high demand for the hardware, shipping times, and other reasons. These estimates resulted in the lower bound of electricity consumption of 2.55 GW and possible consumption of 7.67 GW (De Vries, 2018).

In response to De Vries' estimates, Bevand (2017b) presented an article, in which he summarized flaws of the above-described approach. Later, Bevand (2017a) published his own estimate of the mining electricity consumption, coming with the values of 325 MW for the lower bound, 774 MW for the upper bound, and 470 to 540 MW for the best guess as of 26 February 2017. He made two updates on these estimates: 640 MW for the lower bound, 1.248 GW for the upper bound and 816 to 944 MW for the best guess as of 28 July 2017 and 1.62 GW for the lower bound, 3.136 GW for the upper bound, and 2.1 GW for the best guess as of 11 January 2018. It can be noted that the estimate of De Vries is approximately 1.5 times higher than the Bevand's estimate from the corresponding time. His approach used data on the total hashrate of the network and hashing efficiencies of ASICs, similar as Malone & O'Dwyer (2014), with additional information on the availability of hardware units to purchase and amounts of units sold, that were available thanks to the communication with one of the ASIC manufacturers (Bevand, 2017a).

Another estimate of power consumption of Bitcoin mining comes from Vranken (2017). In this paper, the overview of the Bitcoin mining system was presented and then the power consumption was estimated to lay between 100 MW and 500 MW (as of January 2017). This is a similar (although slightly smaller) estimate as the one made by Bevand (2017a), which came out two months later. Vranken uses the same method of estimation as Malone & O'Dwyer (2014), total hashrate divided by the efficiency of used miners, with updated data on the hardware in use. The main contribution of this paper is that it considers the economic aspect of running mining hardware from the point of view of a hardware manufacturer. Based on McCook (2014), Vranken makes a reasonable assumption that 80% of the mining hardware is run by manufacturers themselves, therefore reducing capital expenditures significantly, allowing them to mine for up to 30% cheaper than retail miners (Vranken, 2017).

Zade *et al.* (2019) made predictions on the future development of the power consumption of Bitcoin and Ethereum mining up until the year 2025, based on six scenarios. These scenarios presented different ways, in which the difficulty and efficiency of mining could develop. The authors used data on the mining efficiencies of various mining units from a dataset constructed one year earlier (Zade & Myklebost, 2018) and data on mining difficulties adjustments to create two scenarios for efficiency and three scenarios for difficulty, therefore, when combined, resulting in six scenarios for Bitcoin and six scenarios for Ethereum in total. Results showed that improvements in hardware efficiency are likely to have only a limited impact on the power demand of the network, as the rate at which the hardware is improving tends to slow down. Future difficulty (and therefore the total hashrate, as the difficulty is adjusted according to hashrate) might on the other hand be the main driver of the demand for electricity. Whether it will be stagnating, growing linearly, or growing exponentially, it will probably strongly influence the power consumption (Zade *et al.*, 2019). Authors also provided their estimates on the lower bound of electricity consumed by the mining of Bitcoin[3] so far: 168, 202, 364, 843 and

[3]Estimated lower bound of the power demand of the Ethereum network was provided as well: 3, 19, 367 and 991 MW in 2015, 2016, 2017, and 2018 respectively.

3852 MW in 2014, 2015, 2016, 2017, and 2018 respectively. Estimates are in similar ranges as those done by Bevand (2017a) or Krause & Tolaymat (2018), which is reasonable, as the same bottom-up method (Malone & O'Dwyer, 2014) was used. As for an estimate closer to reality, the authors concluded that if the mining process in October 2018 was done on the three most efficient available ASICs, the power demand of the Bitcoin[4] network would be approximately at 5 GW (Zade *et al.*, 2019).

De Vries (2020) presented an article, in which he broke the lifetime of the Bitcoin network into three categories according to periods when the total hashrate was growing, stagnating, or falling. De Vries argues that when the hashrate is falling, the network tends to converge to its optimum, as the inefficient miners exit the market and only the miners with the best prices and the most efficient hardware remain with non-negative profits, thus leading to the network's minimal power consumption. Further, De Vries shows that the hashrate declining period occurred only once and only briefly (November 2018 to December 2018, *i.e.* one out of total 33 months), and on the other hand the hashrate growth is dominant. The growing hashrate could mean that even sub-optimal miners remain relevant, as the profitability of mining cannot be rapidly exploited by buying and running the most efficient mining hardware due to low production volumes. Combining this observation with previous sales analysis (Stoll *et al.*, 2019), new IPO[5] filings from 2019 and independent market share estimates, De Vries (2020) states that previous estimates might have overestimated the efficiency of hardware used and thus underestimate the power consumed by mining. Additionally, the Bitcoin network was estimated to consume 87.1 TWh annually as of September 30, 2019 (corresponding to 9.9 GW), which is higher than estimates of CBECI (78.3 TWh), as well as Digiconomist's BECI (73.1 TWh) at that time (BECI and CBECI are described in section 2.2.3).

### 2.2.2 Estimates of the emission production caused by Bitcoin mining

In 2018, McCook published an updated version of his previous research (McCook, 2014) on the environmental costs and sustainability of Bitcoin mining.

[4]The power demand of the Ethereum network in October 2018 would be approximately at 0.9 GW if mined on the three best available GPUs.

[5]IPO stands for Initial Public Offering.

In the paper, the author extensively discusses the financial aspects of mining, estimates the energy demand of the Bitcoin network and its environmental impacts. Additionally, the author estimates the same metrics for the gold mining industry and compares it to Bitcoin mining. The Bitcoin network was estimated to consume 105 TWh annually (which, for comparison with the previous research, corresponds with consumption of nearly 12 GW[6] at any given time) and to be responsible for about 0.12% of global greenhouse gas emissions (37 Gt $CO_2$ + 16.5 Gt $CO_2$e) (McCook, 2018). The method used is the same one as in Malone & O'Dwyer (2014); Bevand (2017a); Vranken (2017) while assuming that all the mining is done by the best hardware available (Bitmain Antminer S9i or its equivalent). As for the comparison with gold mining, the mining of Bitcoin appeared to be less harmful to the environment than the mining of gold according to almost all considered indicators (all but the production of carcinogenics). The author also considered the argument that the energy consumption of the Bitcoin network is comparable to the consumption of gold mining (although lower), whereas the market capitalization of Bitcoin is just a fraction of the market capitalization of gold. He argued, that the long-term emission-per-unit trend of gold increases, while for Bitcoin it decreases, and proposed that increasing demand for cheap electricity by Bitcoin miners will in the future lead to faster development of green energy sources (McCook, 2018).

Mora *et al.* published an article that attracted the attention of other researchers, as well as regular media. In the title, the authors boldly state: "Bitcoin emissions alone could push global warming above 2°C" (Mora *et al.*, 2018). The paper attempts to calculate the carbon footprint of Bitcoin in 2017 and to show what would happen if it would continue in a similar manner. Results reveal that Bitcoin alone could be responsible for 69 $MtCO_2$e as of 2017 and if the same trend would continue unchanged, global warming might reach the limit of 2°C in 11 to 22 years (depending on how fast will be the blockchain technology adopted). The estimate on the energy consumption alone was not presented. Mora *et al.* could be criticized (*e.g.* in Dittmar & Praktiknjo (2019)) for several mistakes, such as extrapolating to the future based on the data from 2017, in which Bitcoin experienced sky-rocketing in price, which has

[6]Estimates of McCook include not only operational expenditures (*i.e.* energy required for running mining hardware and for cooling) but also capital expenditures, such as the energy required for construction of ASICs, their packaging and shipping. Therefore it is reasonable that the estimated use of energy is substantially higher than in the previously discussed papers (*e.g.* Bevand (2017a) or Vranken (2017)).

not been repeated in the next three years, or building the $CO_2$ estimate on the number of completed transactions (Mora *et al.*, 2018), although the number of transactions per second does not have a direct influence on the number of miners running and therefore on the electricity consumption.

Another estimate of the energy consumption of Bitcoin network and emissions thus produced comes from Krause & Tolaymat (2018). The main purpose of the paper was to compare emissions produced by creating 1 US dollar worth of cryptocurrencies and 1 US dollar worth of precious metals. As for cryptocurrencies, Bitcoin, Ethereum, Litecoin, and Monero were selected, as all of them are based on a proof-of-work algorithm, so their mining requires a non-negligible amount of energy. As for metals, aluminium, copper, gold, platinum, and rare earth oxides were used for the comparison. The results showed that in general, excluding aluminium, mining of cryptocurrencies consumed more energy than the mining of minerals[7] to create 1 US$ of worth. Throughout the measured period (1 January 2016 to 30 June 2018), all four cryptocurrencies together were estimated to be cause for producing 3 to 15 million tonnes of $CO_2$ emissions, out of which Bitcoin alone seemed to be responsible for 3 to 13 million tonnes of $CO_2$[8]. Although it might be questionable, whether it is logical to compare dollar-value production of cryptocurrencies with precious metals[9], authors provided their estimates on the energy consumption of Bitcoin (as well as other cryptocurrencies). Bitcoin mining was estimated to consume 283 MW in 2016, 948 MW in 2017, and 3441 MW in the first half of 2018. The method of estimation is similar to the method used by Bevand (2017a), for each year, authors divided total network hashrate by assumed efficiency of miners, whereas cooling costs, as well as any other costs, were neglected. It might be argued (as did Koomey (2019)) that the cryptocurrency industry is such a dynamic environment, that averaging over a full year is a period too long for capturing all the potential changes, and for example, a month-based approach might be better suited.

One of the most evidence-based estimates of Bitcoin electricity consumption and emission production comes from Stoll *et al.* (2019). The authors took an advantage of the IPO files provided by the three major producers of mining hardware (Bitmain, Ebang, and Canaan). Information on the total market

[7]For the creation of 1 US$, Bitcoin, Ethereum, Litecoin, and Monero consumed on average 17, 7, 7 and 14 MJ respectively, while aluminium, copper, gold, platinum, and rare earth oxides consumed 122, 4, 5, 7 and 9 MJ respectively.

[8]Which is significantly lower estimate than 69 $MtCO_2e$ by Mora *et al.* (2018).

[9]As discussed for example by Carter (2018).

shares of producers was available, as well as somewhat accurate data on units sold. Therefore, the total efficiency of hardware in use could be estimated with a high level of precision. Based on the communication with large and medium-scale miners, Stoll *et al.* estimated a portion of the total electricity consumed, that has to be spent on cooling and other operational expenses (other than mining itself), depending on the size of the mining facility. Worldwide distribution among the categories of miners (small, medium, and large) was then assumed based on the publicly available data on miners' distribution in the Slushpool mining pool. This allowed the authors to average this additional electricity demand over the entire world, resulting in assumed 5% extra electricity needed, and to include it into the estimation as well. From July 2016 to January 2019, the lower and upper limits of electricity consumption were estimated as all miners using the most efficient hardware at the time for the lower limit and all miners earning zero marginal profit from mining as the upper limit. Three best-guesses[10] were then computed, using data on miming efficiency and the percentage of electricity required for cooling and overhead expenses, resulting in estimated consumption of 345 MW at the end of 2016, 1637 MW at the end of 2017, and 5232 MW in November 2018 (Stoll *et al.*, 2019). The authors also computed the carbon emissions resulting from Bitcoin electricity demand. They used three methods of estimation of the miners' location, two of them yielding reasonable results, in combination with average and marginal emission factors of power generation in the countries of interest. Two methods of estimating miners' location resulted in the range of Bitcoin emission production of 22.0 to 22.9 $MtCO_2$ (Stoll *et al.*, 2019). Additionally, in 2020, the same group of authors published a paper aiming at quantifying the electricity consumption of all major cryptocurrencies, not only of Bitcoin (Gallersdörfer *et al.*, 2020). 20 mineable cryptocurrencies with the highest market capitalization[11] were chosen, representing together more than 98% of the total market capitalization of all cryptocurrencies. Results indicated that at the end of March 2020, Bitcoin mining was responsible for approximately 4291 MW and also that Bitcoin is responsible nearly for 2/3 of total electricity consumption, whereas other major cryptocurrencies are responsible for the remaining 1/3. The authors chose a simple method of dividing total hashrate by efficiency of the best hardware available, therefore the estimate is considerably lower than in papers that used a more sophisticated method (especially if the

[10]It could be noted that all three best guesses are close to the lower bound.

[11]Acording to https://coinmarketcap.com

almost always rising total hashrate is considered). But the same method was used on all 20 cryptocurrencies alike, thus even if the total consumption might be higher, the ratio of consumption between Bitcoin and the rest might be correct (Gallersdörfer *et al.*, 2020).

Later in 2019, Köhler & Pizzol released a paper, in which they applied Life Cycle Assessment methodology on the Bitcoin mining network, thus allowing them to estimate the electricity consumption and emission production so far and develop three scenarios for the future. Authors utilized previous studies to get data on the geographical distribution of miners[12], hardware in use[13] and a percentage of electricity used for cooling[14]. The mining electricity consumption in 2018 has been estimated to be 31.29 TWh (which corresponds with a steady consumption of 3.57 GW). Additionally, the emission production was estimated to be 17.29 $MtCO_2e$ in 2018 (Köhler & Pizzol, 2019). The authors also argued that production and disposal of the mining hardware generate significantly fewer emissions than previously assumed (McCook, 2018), accounting together for less than 1% of the total carbon footprint of Bitcoin. The three scenarios presented situations where *a*) the geographical location of miners and their computing power and efficiency will stay the same as now, *b*) the geographical location will remain unchanged, but better mining hardware will be installed, or *c*) the geographical location, as well as the hardware, will change in order to provide the best competitive conditions to miners. Evaluation of the scenarios showed that the geographical location of miners is the most crucial factor, as it has the largest impact on the emission production caused by Bitcoin electricity demand. Authors also presume that in the long-term, the total hashrate (and thus the electricity consumption) will probably stagnate, as the main source of income for miners will shift from the block rewards to the transaction fees.

### 2.2.3 Continuous estimates of the electricity demand

The simplified approach of De Vries (2018) based on two major assumptions (the assumption of the electricity price of 0.05 USD/kWh and the assumption that 60% of miners revenue is spent on electricity costs) allowed the author to produce the estimates steadily over time, which the author publishes on

[12] The geographical distribution of mining activities was assumed according to Bendiksen *et al.* (2018); Rauchs *et al.* (2018) and authors' own research.

[13] Acording to Bendiksen *et al.* (2018), the mix of most up-to-date mining hardware was used.

[14] According to Stoll *et al.* (2019), the additional electricity demand of 5% was assumed.

his website Digiconomist[15] since early 2017 under the name Bitcoin Energy Consumption Index. However, Bevand (2017b) criticized De Vries for arbitrary changes in methodology, that were not explained.

In July 2019, Cambridge Center for Alternative Finance (CCAF) launched their Cambridge Bitcoin Electricity Consumption Index (CBECI[16]), that aims to track the electricity consumption of the Bitcoin network in real-time and, based on this, to provide its annual electricity consumption (Rauchs *et al.*, 2020b). Authors referenced to approach of Bevand (2017a) and employed a similar bottom-up method. The index provides three estimates. Firstly, the best-case scenario, the lowest possible bound of consumption, where it is assumed that mining is done by the most efficient mining hardware available and that miners will switch to the better hardware as soon as it is released (which is highly unlikely, if not impossible, due to a non-negligible shipping time or not-high-enough production volumes). Also, extremely good power usage efficiency (PUE[17]) of 1.01 is assumed. Secondly, the worst-case scenario, or the highest reasonable bound of electricity consumption, assumes that mining is done by the least efficient hardware that is still profitable. Under this assumption, miners will switch to more efficient hardware only when their current hardware becomes unprofitable (considering only operational costs, not capital expenses connected with the purchase of the mining units). Miners are also assumed to have the PUE of 1.20, which would still be considered efficient by common data center parameters, but not by mining facilities' requirements, where the electricity consumed plays a crucial role. Thirdly, the authors' best guess estimate, which assumes that miners are using an equally weighted mix of all mining units that are profitable at a time. Authors argue that such assumption is acceptable due to the unavailability of reliable comprehensive data on amounts of mining units sold and show that the estimate resulting from this assumption is comparable to the one done by Stoll *et al.* (2019), who used data from hardware manufacturers IPO filings. Also, in the best guess scenario, the PUE of 1.10 is assumed, as it has been confirmed by conversations with miners and mining experts, although it is more conservative than the best guess on PUE in the previous studies[18]. For all three estimates, the electricity

[15]https://digiconomist.net/bitcoin-energy-consumption

[16]Available at https://cbeci.org/.

[17]PUE refers to the amount of electricity required by a mining facility, but not directly used for mining. PUE of 1.01 means that 1% of additional electrical power is required for cooling, running other hardware components then ASICs *etc.*, on top of 100% electricity power needed for mining itself.

[18]PUE of 1.05 was assumed by Stoll *et al.* (2019) or Bevand (2018).

price of 5 USD cents per kWh is assumed[19], as it was supported by previous studies, as well as authors' communication with miners. Authors based the efficiencies and availability dates of mining hardware units on the findings of previous papers and their own research. CBECI provides the estimate on the immediate electricity consumption of the Bitcoin network in watts, as well as the annual consumption in watt-hours, which results from applying the 7-day moving average on the immediate consumption (Rauchs *et al.*, 2020b).

## 2.3 Drivers of the price of Bitcoin

For the proposed analysis, it will be crucial to properly identify factors, that drive the price of Bitcoin. Therefore, this section contains a summary of papers concerned with identifying such variables and specifying their effects.

Kristoufek (2013) was one of the first who examined the relationship between Bitcoin price and the frequency of online searches. He used data from Google Trends and Wikipedia, setting thus a standard frequently used in a similar type of analyzes. The vector autoregression (VAR) approach and vector error correction model (VECM) was used with an addition of differentiating between positive and negative sentiment in the bitcoin price, expressed by a variable that indicates whether the price of bitcoin is above or below its trend. The results revealed a strong, bidirectional relationship, *i.e.* not only the increase in the price can be explained by the increase in public attention, but also more public attention is caused by the rising price of Bitcoin. Moreover, when the sentiment is negative, the bidirectional relationship exists as well, but with a negative sign.

Garcia *et al.* (2014) followed with their analysis of Bitcoin's price bubble formation process. For this purpose, the vector autoregression (VAR) framework was used and three types of social signals were assessed: a volume of online searches, a number of new users, and a social media activity. Similarly to Kristoufek (2013), online searches were represented by data from Google Trends or by views of the Bitcoin Wikipedia page, new users by a number of downloads of the Bitcoin software client in a given day, or alternatively by a number of new addresses and social interactions by a number of Bitcoin-related tweets per million tweets or by a number of re-shares of Bitcoin-related posts on a selected Facebook group. Additionally, Bitcoin exchange rates were taken

[19] Nonetheless, the CBECI website allows users to change the assumed electricity prices, in order to see how the resulting estimates will change.

from various exchanges and for three currencies: USD, Euro, and CNY. Results revealed two feedback cycles that might drive the price of Bitcoin. The first of them starts with an increase in price positively affecting the search volume, word of mouth (social media activity) increases as the search volume increases, and price increases as the word of mouth increases. The second cycle starts the same, search volumes are increased by price growth, then in turn the amount of new users increases with high search volumes, and price grows with a spreading user base. Moreover, a negative effect of huge spikes in search volume on the price was observed. Garcia *et al.* found that three out of four largest price drops were preceded by significant increases in searches.

Kristoufek (2015) applied the wavelet coherence analysis on various possible sources of price movements. This allowed him to examine the correlation between series across the time span from the end of 2011 to the start of 2014 and for various frequencies, enabling him to distinguish the long-term and short-term effects and the leading variable. The first group evaluated were the Economic drivers. The Trade-Exchange ratio was defined as a ratio between the volume of the exchange transactions and the trade transactions and represented a measure of transactions used for purchases of products and services (*i.e.* the lower the ratio, the more frequent real-world application of Bitcoin). As expected, the analysis revealed a negative correlation between this ratio and the price with no evident leader. The price level was constructed according to standard economic theory as the average price of a trade transaction on a given day. A negative correlation was revealed with no significant leader in the long-run. The money supply of Bitcoin seemed to be positively correlated with price, against the expectation. This could be explained by the extreme predictability of the Bitcoin algorithm, which allows users to adjust their expectations. The second group of analyzed factors was the transaction drivers, represented by the trade volume and the trade transactions. Any stable relationship could not be seen for the trade volume, and the trade transactions were positively correlated in the long-term. However, the relationship between variables and the price grew weaker in time. The third group considered were the technical drivers. Total network hashrate and the mining difficulty were evaluated, revealing a positive correlation between both variables and the Bitcoin price in the long-run. The price seemed to be the leading variable in both cases, indicating that the increases in price attracted new miners to enter the market (or motivated increases in hashrate in general). The fourth group represented the measures of public interest in Bitcoin, expressed by the search volumes for the term “Bitcoin” on Google and Wikipedia search engines. The correlation of searches on both platforms was positively correlated with price in the long-run. In the short-run though, the relationship was more intricate, as it changed over time, depending on the price movements. It seemed, that during rapid price increases (forming bubbles), interest led the price, boosting it even higher, while during extreme price decreases (bubble bursting), interest still led the price, but this time with the negative correlation, pushing it even lower. Additionally, during the bubble burst, the interest had a more rapid effect on price compared to a period of bubble build-up. The fifth group analyzed were factors connected with a notion of Bitcoin being a safe haven. The Financial Stress Index (FSI) and the gold price in Swiss francs were chosen, as gold itself is a safe haven and FSI is an index of general financial uncertainty. The positive correlation with these factors would indicate Bitcoin itself is a safe haven. Nonetheless, Kristoufek (2015) found no long-term relationship between the Bitcoin price and either of the mentioned variables, thus concluding that Bitcoin cannot be considered a safe haven, based on the analyzed dataset. Finally, in the sixth group, the influence of China was examined, as a reaction to speculations of China having a strong influence over the entire Bitcoin market. For this analysis were used Bitcoin prices in USD and CNY and exchange volumes in USD and CNY, in CNY alone and in CNY while controlling for the exchange rate of USD. Results showed no proof of any causal relationship between the CNY and USD markets.

Based on previous studies, Ciaian *et al.* (2016) summarized drivers of Bitcoin’s price into three categories: forces of supply and demand, popularity among investors and users, and worldwide financial growth. Further, the authors argued that their predecessors neglected two important factors, the first one being that previously the drivers were analyzed separately and the interaction between them was not accounted for. This might have resulted in over-stressing the importance of some factors. The second flaw of previous research, according to Ciaian *et al.*, was the absence of analysis of potential structural breaks in the price of Bitcoin, which could have led to biased results. They attempted to preclude these shortcomings in their analysis, which was based on a model for a gold standard but enriched with the Bitcoin-specific factors. The daily data from 2009 to 2015 were analyzed with the help of the Vector Error Correction (VEC) model, the Multivariate Vector Autoregressive (VAR) model, and the Autoregressive Distributed Lag (ARDL) model. The first group of factors included in the analysis was total Bitcoins mined, the number of unique transactions per day, and the number of unique addresses per day, the days destroyed[20] of any transaction as a proxy for money velocity and the exchange rate between USD and Euro as a proxy for the global price level. These factors represented the supply and demand forces on the Bitcoin market. The second group, related to the popularity of Bitcoin, consisted of searches for the term "Bitcoin" on Wikipedia (as in Kristoufek (2013; 2015)), a number of new members, and a number of new posts on the online forum Bitcointalk.org[21]. The group covering the global macroeconomic situation was made of oil prices and the Dow Jones Industrial Average index. The results showed that, regarding factors of the first group, the number of transactions and the number of addresses had a significant positive impact on the price, while the total number of Bitcoins in circulation was significantly negative. The proxy variable for money velocity proved to be insignificant. From the second group, the only significant factor in the long-run was the new forum posts with a positive effect, while the new forum members and the Wikipedia searches appeared to be significant only in the short-run, without a clear direction. None of the factors from the third group were significant in the long-run, contrary to results from previous research, implying that global financial development has little to no impact on the price of Bitcoin. Ciaian *et al.* (2016) explained this by their innovative inclusion of factors from various areas.

Kjærland *et al.* (2018) examined various factors with the goal of determining which of them are the drivers of Bitcoin's price. For this purpose, the Autoregressive Distributed Lag (ARDL) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models were used. To account for the huge price changes of late 2017 and early 2018, the authors divided data into two subsections, the first from 2013 to 2016 and the second from 2017 to 2018. Also, to avoid potential autocorrelation, the daily data were transformed into weekly averages. As for the factors examined, in their analysis authors included the price of Bitcoin lagged by one week, hashrate, transaction volume, S&P500 index, Gold and Oil price indices, a measure of the expected market volatility VIX and the search statistics for the term "Bitcoin" from Google Trends. S&P500 represented general financial markets trends, the VIX index was supposed to provide a measure of risks connected with investing for the next 30 days. Gold and Oil prices were added as important commodities, that

[20]This variable was computed by multiplying the number of Bitcoins in the transaction by the number of days since these Bitcoins have moved to a different address (Ciaian *et al.*, 2016)

[21]https://bitcointalk.org/

might affect financial markets overall. Data from Google Trends were added as a measure of public attention directed to Bitcoin. Transaction volume was used as a traditional measure of supply and demand and the hashrate was included as an important technological factor. The most surprising result of the analysis was that the hashrate is insignificant in explaining price changes, which is a contradictory finding compared to previous studies. Authors argue that most likely the price explains the hashrate, not the other way around. On the other hand, in line with the previous studies was a finding that public attention has a significant positive relationship with Bitcoin price. S&P500 has also been found to have a positive impact on Bitcoin's price, which authors interpreted as optimism in financial markets is followed by optimism in the Bitcoin market. Another factor with a significant positive effect was the price of Bitcoin lagged by one week. Authors propose that this might be evidence against the Efficient Market Hypothesis and in turn support for the Momentum theory and the Greater Fool theory[22]. VIX index, a measure of fear in financial markets, was found insignificant in the first period, but in 2017 and 2018, it has been found to have a significant negative relationship with the price of Bitcoin. This might suggest that Bitcoin could to a certain degree have safe-haven properties. Gold and Oil prices have been found insignificant. Results from previous papers regarding these two commodities are mixed. Results on the transaction volume were not clear, each model yielding different results. Either the volume is not significant, or it is significant with a negative sign, as could be expected according to traditional supply and demand theory (Kjærland *et al.*, 2018).

In 2019, Kristoufek expanded upon his research of fundamental drivers of the Bitcoin price and examined it through the lens of classical economic equations. With the use of the Equation of Exchange[23] and the Law of one Price[24], the author was able to construct a price level index for the Bitcoin market and consequently to examine its relationship with the nominal exchange rate (with the USD). The price level index was computed as a monthly average of

[22]The Efficient Market Hypothesis says that investors behave rationally according to available fundamental information, Momentum theory is an empirical rule of thumb that says that rising and falling price rises and falls more than it rationally should have and according to the Greater Fool theory investors do not behave based on fundamental information at all.

[23]The equation states that a volume of money in circulation times the money velocity is identical as an average price of purchased goods and services (price level) times the number of realized transactions.

[24]The law says that the identical good has an identical value, independent of the location and other circumstances, and potential price differences are explained by the exchange rate. This could be written as a price level in one country is equal to a price level of the second country multiplied by the exchange rate.

the total transaction volume divided by the number of transactions, which was possible due to the unprecedented data availability of the Bitcoin network. As similar data for the USD are not accessible directly, the Consumer Price Index (CPI) was taken as the proxy for the price level of the USA. Because the CPI does not represent values of a price level directly, but refers to a point in the past and compares present prices with past prices, the relationship could be examined only proportionally. Kristoufek (2019) found a possible cointegration relationship between the time series, concluding that:

> ( . . . ) results suggest that there is a long-term equilibrium relationship between Bitcoin price and its price dynamics implied by the Equation of Exchange and Law of One Price.

Additionally, it has been found that the market price of Bitcoin copies the fundamental price implied by the model not perfectly, but very close. The author also suggests that the Bitcoin price development depends among other things on the increasing penetration and amount of users.

Wheatley *et al.* (2019) analyzed the bitcoin price bubbles with the aim of developing a methodology that would be able to identify an ongoing bubble and also to detect signals of increased risk of the bubble bursting. For this purpose, the authors used the generalized form of Metcalfe's law[25] and the Log-Periodic Power Law Singularity (LPPLS) model. Apart from detecting four Bitcoin price bubbles in the past and explaining a way how to detect a risk of a bubble bursting, the authors concluded that the number of active addresses is an important factor of the total market capitalization of the Bitcoin network (therefore also of the price itself).

Kristoufek (2020), similarly to Hayes (2017), examined the relationship of the Bitcoin price and the cost of mining (the cost of producing one Bitcoin). For that reason, Kristoufek employed VAR and VECM models and used the data from the start of 2014 up until the second half of 2018, as the introduction of ASICs in 2014 made prior data on hashrate hardly comparable to newer ones. For the analysis, the author used a variable representing the cost of mining, which he constructed from the total hashrate, an index of electricity prices[26] and the best-available miner efficiency. This approach accounts only for operational

[25] This law says that a value of a network (a telecommunications network in the original form) is proportional to the squared amount of active users of the network.

[26] The electricity price was averaged over the prices from countries in which the majority of mining facilities are assumed to be situated and also which disclose the electricity price publicly.

costs of mining and neglects the capital expenditures, as reliable data for all the fixed costs of mining are close to impossible to collect. Kristoufek (2020) argues that the capital expenditures can be assumed to be a fixed percentage markup to the operational costs that will be incorporated in the intercept of the final log-log model. Results showed that the influence of the mining cost on the price is weak and present only in the short-run. On the other hand, the effect of the price on the mining cost was significant even in the long-run, with a shock in price being followed by an adjustment in mining cost, ranging from three months to one year.

Paper from Fantazzini & Kolodin (2020) examined the discrepancy between the results of Kjærland *et al.* (2018) and Hayes (2017), who both analyzed the relationship of the hashrate and price, but came to different results. While Kjærland *et al.* found hashrate to be insignificant while explaining price movements and on the other hand concluded that the price might explain changes in hashrate, Hayes constructed the cost-of-production model (CPM) and argued that the "fair value" (production cost) is an important factor of Bitcoin's market price. Later, Hayes (2019) showed that the cost-of-production value of Bitcoin Granger-causes the price. To find what is the relationship between price and hashrate, Fantazzini & Kolodin (2020) used directly the hashrate, or alternatively the break-even cost of mining (computed by the CPM) as a proxy variable for the hashrate, in bivariate and multivariate models, employing the vector autoregression (VAR) and vector error correction model (VECM) methodologies. For the multivariate model, the authors used the same set of variables as Kjærland *et al.* (2018), but excluded the oil prices and VIX index and included the total transaction fees paid in the Bitcoin network (in USD). The cost of mining required additional information on the efficiency of miners and the electricity price. The efficiency authors estimated by using data from websites listing the mining hardware and applying Holt-Winters double exponential smoothing. The electricity cost was either assumed to be constant at 0.13 USD per kWh or was taken as an equilibrium price from Nord Pool, which is a power exchange operator in Northern and Western Europe. To avoid the potential effect of Bitcoin reward halving on price, the examined period was from August 2016 to February 2020, *i.e* between the two halving events. The analysis showed that it is probably better to consider the hashrate directly, rather than its proxy in the form of the cost of production while modeling its relationship with price. Also, similarly to Kjærland *et al.* (2018), Fantazzini & Kolodin (2020) found that the hashrate seemed to be driven by price, not the other way around.

Next, a section where a model to be used is proposed and the construction of a dataset is described follows. Also, a deeper motivation for the analytical part of the work and a methodology to be utilized are discussed.

# Chapter 3

# Methodology and data

This section presents a motivation for the research part of the thesis and specifies the model that will be estimated. Reasons for including specific variables are provided and the methodology that will be put to use is discussed. Furthermore, the process of data collection is described, together with a way in which some of the variables were constructed.

## 3.1 Motivation

As Kristoufek (2020) found, there is a relationship between the bitcoin price and the operational cost of mining (which consists of the electricity price, the mining efficiency, and the hashrate). However, changes in the hashrate were much larger compared to the changes in the electricity price and the efficiency in the past several years[1]. According to Ciaian *et al.* (2016), the inclusion of multiple factors from various areas and their simultaneous analysis improved their results greatly. Therefore, examining the relationship of the price and the hashrate with the inclusion of the electricity price, the mining efficiency, and other factors (closely discussed in section 3.4) that proved to be significant in the previous research as explanatory variables, might provide insightful results.

The importance of understanding the dynamics of the relationship between the Bitcoin price and the hashrate is at hand. It could help and improve the efficiency of investors in cryptocurrency markets, as changes in the Bitcoin price are often closely followed by changes in the pricing of other cryptocurrencies.

[1]Since the start of 2014 for example, the mining efficiency improved approximately 50 times, the electricity price oscillated in the range of approximately 50% to 150% of the mean value and the hashrate increased more than 12000 times. Note that these are only back-of-the-envelope calculations.

Additionally, governments and decision-makers may benefit from the deciphering of the price-hashrate relationship as well, as a good understanding of the properties of cryptocurrencies is a necessary condition for their just and fair regulation.

It should also be kept in mind that the total hashrate is one of the two driving factors of the total amount of electricity, that is consumed by the Bitcoin network, the other factor being the mining efficiency. Therefore, if a strong influence of price, or some other factor, on the hashrate should be revealed, implications to the electricity consumption could be drawn from the price changes in the future.

In the research field of the Bitcoin price dynamics, the hashrate was assumed to be an important technological factor that might help to explain the movements of the price. Nonetheless, several authors (such as Fantazzini & Kolodin (2020); Kjærland *et al.* (2018); Kristoufek (2015)) suggested that instead of the Bitcoin price being affected by the hashrate, the situation could be quite the opposite, *i.e.* changes in the price might be causing the network hashrate to change. This indicates that a system of equations should be constructed, with the Bitcoin price and the hashrate both being dependent variables and also included as explanatory variables transversely. As suggested by the previous research, both the price and the hashrate will probably be endogenous explanatory variables in such a system, as there is likely a mutual relationship among them, changes in one might cause a change of the other variable and vice versa. The system of equations used to model the relationship of the price and the hashrate and also their relation to other factors shall be discussed in the next section.

## 3.2 Model proposal

When the aim is to explain changes in the Bitcoin price and in the network hashrate, the selection of the explanatory variables is affected firstly by the logic and rules that drive the Bitcoin network, and secondly by the data availability, as some of the variables are not accessible, because it would be hard or even impossible to measure them. In some cases, an explanatory variable can be assumed to be exogenous, as its values are determined solely by the forces that come outside of the Bitcoin system. However, there are variables that could be able to explain some of the variance in the price or the hashrate, and simultaneously might be at least partially affected by the forces already contained in the system, thus being endogenous. The variables with such characteristics are most likely the Total transaction fees paid in the network (measured in BTC), denoted as $\{transaction\_fees_t\}$, and the search volume, representing the public interest in the topic of Bitcoin, denoted by $\{search\_volume_t\}$.

Therefore, together with the Bitcoin price and the hashrate, these four variables were assumed to be endogenous in the system, and thus for each of them, an equation was constructed. The four equations are presented in this section, together with reasons for including the explanatory variables and the expected relationships that ought to be revealed. Also, the reasons for assuming the four mentioned variables to be endogenous are further explored.

The four equations were constructed as follows:

$$\begin{aligned}\log(price_t) = \alpha_1 &+ \beta_1 \log(hashrate_t) + \beta_2 \log(addresses_t) \\ &+ \beta_3 \log(price\_level_t) + \beta_4 \log(USD\_M2_t) \\ &+ \beta_5 \log(search\_volume_t) + \beta_6 \log(transaction\_fees_t) + \epsilon_{1t}\end{aligned} \tag{3.1}$$

$$\begin{aligned}\log(hashrate_t) = \alpha_2 &+ \beta_7 \log(price_t) + \beta_8 \log(addresses_t) \\ &+ \beta_9 \log(efficiency_t) + \beta_{10} \log(transaction\_fees_t) \\ &+ \beta_{11} reward\_phase_t + \epsilon_{2t}\end{aligned} \tag{3.2}$$

$$\begin{aligned}\log(transaction\_fees_t) = \alpha_3 &+ \beta_{12} \log(price_t) + \beta_{13} \log(addresses_t) \\ &+ \beta_{14} \log(search\_volume_t) + \epsilon_{3t}\end{aligned} \tag{3.3}$$

$$\log(search\_volume_t) = \alpha_4 + \beta_{15} \log(price_t) + \epsilon_{4t} \tag{3.4}$$

where $t = 1 \ldots T$ is a time index, $\alpha_1$, $\alpha_2$, $\alpha_3$ and $\alpha_4$ are intercepts, $\beta_j$ for $j \in \{1 \ldots 15\}$ are the coefficients of the explanatory variables and $\{\epsilon_{1t}\}$, $\{\epsilon_{2t}\}$, $\{\epsilon_{3t}\}$ and $\{\epsilon_{4t}\}$ are the error terms.

**Price equation** The equation 3.1 explores relation of some variables to the Bitcoin price, represented by $\{price_t\}$. The relationship with the total network hashrate, depicted by $\{hashrate_t\}$, is one of the main questions of this work. The expectation is that the increasing hashrate will positively affect the price, as increased security of the network might make it more attractive for users and investors. Additionally, a positive relationship was already shown in previous studies. The variable $\{addresses_t\}$ represents the number of active users of the Bitcoin network, as Wheatley *et al.* (2019) showed that the number of active users plays a role in the valuation of the technology. In line with Metcalfe's Law, it can be expected that the price will be positively affected by the growing user base. The price level of the Bitcoin market is represented by the variable $\{price\_level_t\}$, which was constructed according to Kristoufek (2019) as the volume of Bitcoin transactions divided by their number. This relationship directly follows fundamental economic theory, which also says that an increasing price level of Bitcoin should have a negative effect on the exchange rate (an increase in the price level will lead to 1 Bitcoin being equivalent to fewer US dollars). In order to capture the ability of Bitcoin to serve as a safe haven for investors in times of strong inflation, the M2 money supply of the USD was added as $\{USD\_M2_t\}$. In the previous papers, the price of gold or the S&P 500 index were used, as they represent overall indicators of the global market situation, but the results did not indicate a clear relationship. The Covid-19 pandemic crisis led to an unprecedented quantitative easing in the United States, which in turn fueled the speculations about the Bitcoin being a hedge against inflation. The inclusion of the USD money supply might bring an insight into the situation. It could be expected that the increasing money supply will lead to growth in the price of Bitcoin. The public attention towards Bitcoin was captured by the variable $\{search\_volume_t\}$, which represents the Google search data. The increasing number of people interested in the Bitcoin technology might signify an increase in demand for Bitcoins, which could result in price increases, and on the other hand in situations when the price is decreasing, the increased interest in (potentially negative) news about Bitcoin might push the price even lower. Therefore, the expectations are mixed, but in the previous research, the overall positive effect outweighed the negative one. The variable $\{transaction\_fees_t\}$ depicts the total transaction fees paid by users of the Bitcoin network (those who send transactions) and received by miners in one day (measured in Bitcoin). The height of the fee depends on the senders' willingness to pay for completing the transaction, as the fee is an incentive for the miner who finds a block to include a transaction into the block. Therefore, a transaction with a fee set high has a higher priority and gets completed earlier than transactions with a low fee. It could be expected that high fees will motivate higher prices because they represent the interest of users to complete their transactions, *i.e.* to use the network. And, similarly as in the case of the active addresses, an increase in the interest for the use of the network might lead to an appreciation of Bitcoin. However, the opposite relationship would be also possible, as extremely high fees might discourage users from sending transactions, thus effectively lowering the price.

Hashrate equation The equation 3.2 describes the relation of the hashrate to its explanatory factors. The Bitcoin price is clearly linked to the total hashrate, as a major part of miners' revenue is dictated by the algorithm in the form of the block reward (the number of Bitcoins awarded to a miner who mines a block). The price of Bitcoin, therefore, affects whether miners will be profitable and able to continue mining, or whether they will operate in loss and will be forced to end their mining activities. The expectation is that the increasing price will drive the hashrate higher, as new miners will be motivated to enter the business. Another variable that might be able to capture the changes in the hashrate is the number of active addresses. Similarly, as in the case of the Price equation, it could be expected that a growing user base might demand higher security of the network, which in turn is reflected by the growing hashrate. The efficiency of the mining hardware was in the model captured by the variable $\{efficiency_t\}$. It represents the combined efficiency of the best mining units available at time $t$ in J/GH (joules per gigahash). The efficiency of mining is one of the crucial segments of the total hashrate, as more efficient hardware enables miners to output more computing power for the same amount of electricity. It could be expected the more efficient hardware will be (the less J per GH will be required), the higher the hashrate will be, as miners will be able to mine with lower electricity expenses. The variable $\{transaction\_fees_t\}$ is a two-sided coin, the first side being the costs that users incur while sending transactions, as described before. The other side is the revenue of miners, who receive the fees. The share of miners' revenue formed by the transaction fees depends on how many users are willing to spend their Bitcoins on the fees, the higher the demand for the completed transactions, the higher the transaction fees. It can be expected that high fees will result in a higher hashrate, as miners' incentive to mine will increase. A variable representing the Bitcoin reward halving events was added as $\{reward\_phase_t\}$. It is equal to a number of the halving events that occurred prior to a time $t$. There is no clear consensus in the literature for how to account for the effects of the Bitcoin halving, even though the empirical evidence suggests that some relationship with the hashrate exists. Adding such a variable might be a way of disentangling the relationship. The expectation is that a reward halving will have a negative influence on the hashrate, as it decreases the miners' reward.

As mentioned before, the Bitcoin price and the hashrate are very likely to be endogenous variables, as an increase in the price is expected to have a positive impact on the hashrate, and growth of hashrate is expected to increase a price (although the price to hashrate relationship is likely to be stronger). However, these are not the only endogenous variables in the system. The variables $\{transaction\_fees_t\}$ and $\{search\_volume_t\}$ were expected to be able to explain some of the variance in the Bitcoin price and hashrate and thus they were used as explanatory variables in 3.2 and 3.2. However, it is reasonable to assume that their values might be at least partially driven by the forces already contained in the system, and thus to be endogenous as well.

Transaction fees equation The transaction fees are driven by the demand for completing transactions as fast as possible. This need consists of several factors explored by equation 3.3, the first of them being the price. A sudden change in Bitcoin price might motivate users to send transactions more than usual[2], increasing thus transaction fees. This, however, does not indicate whether the relationship will be positive or negative, but because the price of Bitcoin is overall increasing, a positive influence of price could be expected. The number of active addresses can influence the transaction fees in a very simple way. The more active addresses there is on a given day, the more users are trying to send their transaction. However, the block-space is limited, there is a maximum number of transactions that can be sent per block. Therefore, the transaction fees will be naturally higher, as with more transactions in line there is more competition for space in a block. The effect of $\{addresses_t\}$ is thus expected to be positive. The search volume could have a similar effect

[2]Typically, the transaction fees spike when there are extreme growths or falls in the price, as many users want to buy or sell Bitcoins as fast as possible.

as the number of active addresses. The more people are interested in Bitcoin, the more potential users exist, and thus the higher demand for the transactions completed there might be[3]. Following this logic, the influence of the search volume on the transaction fees is expected to be positive.

Search volume equation The last equation of the proposed system, 3.4, captures how the search volume is affected by the Bitcoin price, as there is likely no other variable that could significantly influence levels of the online search queries. The extreme changes in the price have usually been assumed to be the reason why people started to be interested in Bitcoin. Similarly to transaction fees in equation 3.1, the influence of price on the search volume is ambivalent. Both the spikes and the falls of price are likely to cause the increased attention of people toward Bitcoin. The expectations are thus mixed, but, as the price is growing in general, the positive effect is more likely.

It should be noted that with the exception of the variable signifying the phase of the reward halving, on all of the variables were used the (natural) logarithmic transformation. This was done in order to capture the effect of percentage changes of variables, as in most of the cases, an impact of a single unit change might not be very meaningful.

The presence of the four endogenous variables in the model means that at least four additional exogenous variables will likely be needed for the estimation as instrumental variables. These instruments should not be contained in any of the equations[4] and should be able to explain at least some of the variance of the endogenous variable on the right-hand side of the equation, while not affecting the explained variable on the left-hand side of the equation. For this purpose were used the total Bitcoin supply, the price of electricity used for mining, the Bitcoin price lagged by one time period, and the hashrate lagged by one period. The Bitcoin supply represents the number of Bitcoins that were issued (mined) and the electricity price was constructed as the average industrial electricity price in the USA and Northwestern Europe (a detailed description of constructing this and other variables is provided in section 3.4.2).

[3] Note that this relationship is present regardless of whether the context for the online searches is positive or negative. In the case of positive news, new investors are likely to buy Bitcoin, in the case of negative news, investors owning Bitcoin might be willing to sell it.

[4] At least the instrumental variable for an endogenous variable should not be contained in the equation, in which the endogenous variable is.

It is difficult to find proper instruments in a system as convoluted as the system of Bitcoin price and hashrate. Thus, the lagged values of these two variables were used, as they may carry useful information about the whole system.

An improvement of the analysis could be the use of the true mining efficiency of the hardware used and the true electricity price that miners have to pay. However, such information is unavailable, hence simplified versions of the variables were used. Another improvement would perhaps be the inclusion of the variable representing the sentiment of the public opinion on Bitcoin, telling *e.g.* whether the news articles are critical or appreciative towards Bitcoin. Such a measure could be though quite subjective, therefore constructing a variable of this kind objectively is beyond the scope of this work.

The next section describes the methodology, that will be utilized to perform the regression of the model. After that, the construction of the dataset is described.

## 3.3 Methodology

The four equations indicated in the previous section could be estimated separately by the method of Ordinary least squares (OLS). However, in case that the assumption of the four variables being endogenous would be at least partially correct, the resulting estimates of coefficients would probably be biased and inconsistent (in each equation, in which an endogenous explanatory variable would be present), as there would be an explanatory variable correlated with the error term, which would violate the exogeneity assumption of the Gauss-Markov theorem. According to Wooldridge (2015), this assumption states that the error term $\epsilon$ has a zero expected value given values of any independent variable $x$, more formally stated

$$E(\epsilon|x_1, x_2, \ldots, x_k) = 0.$$

And as shown by Kristoufek (2020), there is a significant relationship between the price and the cost of mining (which consists of the hashrate, the electricity price, and the mining efficiency). Therefore, it can be assumed that a better method for the estimation method than OLS exists.

Another option would be to use the Two-stage least squares (2SLS), which is an appropriate solution in the case of multiple instrumental variables. The 2SLS accounts for the endogeneity of variables by regressing the endogenous variables on the instrumental variables (all exogenous variables in the system of equations) in the first stage and using the fitted values from the first stage instead of the endogenous variable for the OLS estimation in the second stage (Wooldridge, 2015). As there are fewer endogenous explanatory variables than exogenous variables in the system, all four equations could be considered identified.

A potentially better option than the 2SLS estimator could be the 3SLS (Three-stage least squares) estimator, originally specified by Zellner & Theil (1962). The 3SLS is a system estimator, while the 2SLS is only a single equation estimator. It is a combination of the 2SLS estimator and the Seemingly unrelated regression (SUR) estimator, the first stage being the obtaining of the residuals of each structural equation through the 2SLS procedure. Thus estimated residuals are used in the second stage for estimating the error variance-covariance matrix $\hat{\Sigma}$. In the third stage, $\hat{\Sigma}$ is used in the GLS (Generalized least squares) regression instead of the true variance-covariance matrix $\Sigma$ to estimate the entire system (Baltagi, 2011). Just as in the case of 2SLS, all equations could be considered identified.

The advantage of the 3SLS estimator compared to the 2SLS estimator is that it allows for the correlation of unobserved errors across equations (while the errors of equations are assumed to be themselves uncorrelated). This might be useful in the case of the four equations specified above, as it is not impossible that errors are correlated because *e.g.* both price and hashrate might react to exogenous shocks simultaneously. It could be shown on a recent real-world example, where the Chinese regulators started banning the cryptominig in China in the second quarter of 2021 (as reported *e.g.* by Khatri (2021)). The reaction of miners followed as the hashrate dropped by almost 50% in approximately one month. At the same time though, the price dropped by roughly 40%. It could be speculated that these two events are connected.

The 3SLS estimate was shown to be at least as asymptotically efficient as the 2SLS estimator, however in case that at least one equation in the system is specified incorrectly, the 3SLS estimator is inconsistent (Baltagi, 2011). Additionally, there exist several special conditions under which the 2SLS and the 3SLS estimators are identical, but very likely this will not be the case of the two equations specified above. The Hausman test can be used for choosing either the 2SLS or the 3SLS estimator.

For the analysis, the statistical software R (R Core Team, 2019) and a package *sytemfit* (Henningsen & Hamann, 2007) shall be used. The code that performs the analysis is attached to the thesis.

## 3.4 The construction of the dataset

This section describes the collection of data, their sources, and also the way some of the variables were constructed. The description was split into three subsections, the first of which being formed by variables easily available thanks to the blockchain data, the second group containing variables that had to be constructed from multiple sources, and the third group for variables from various other sources.

### 3.4.1 Blockchain variables

As mentioned before, blockchain technology is unique for its unprecedented data availability, as information on every transaction is being saved into the blockchain. This allows for the analysis of many variables that would be hard to measure otherwise. Variables of such type included in the analyzed dataset are the Bitcoin price, the total network hashrate, the number of active addresses, the total transaction fees paid to miners (in BTC), the total Bitcoin circulating supply, and the dates of the block reward halving events. All of these variables, with the exception of the price and the halving events, were downloaded from Blockchain.com[5], where they are available on daily basis. The data on the Bitcoin price was downloaded from Coindesk.com[6], as Coindesk.com provides a thorough explanation of the criteria for choosing the exchanges with verifiable trade data (Acheson *et al.*, 2020), from which the Bitcoin price is taken and averaged. As the majority of exchanges between Bitcoin and fiat currencies are conducted in the United States dollars, the exchange rate between the Bitcoin and the USD was assumed to be the market price of Bitcoin. The total network hashrate is measured in terahashes per second (an average number of terahashes per second the network was performing in the last 24 hours) and is estimated from the number of mined blocks and the difficulty. The number of active addresses shows the number of unique addresses, that were active on the blockchain on a given day (*i.e.* they received or sent Bitcoin). The total transaction fees, measured in Bitcoin, express the total number of Bitcoins that were paid to miners as transaction fees in one day. The total circulating supply

[5]https://www.blockchain.com/
[6]https://www.coindesk.com/

of Bitcoin indicates how many Bitcoins were issued so far. The dates of the halving events were retrieved from Bitcoinblockhalf.com[7]. Figures depicting all hereby mentioned variables are provided in Appendix A.

### 3.4.2 Variables constructed from multiple sources

Some of the variables were not so easily accessible and had to be derived in a more or less complicated way. These are the efficiency of miners, the electricity price, and the price level of Bitcoin.

As mentioned before, it is not known exactly which hardware units are being used for mining and the total mining efficiency (combined energy efficiency of all miners deployed in the network) thus can not be computed with high confidence. In the research papers (see section 2.2), a theoretical optimum of the mining efficiency was computed and used (all miners using the best available hardware at a time), as well as the worst feasible efficiency (all miners mining at break-even costs). Some of the researchers used the data from IPO filings of mining hardware producers in order to approximate the number of sold units, thus estimate the real mining efficiency and Bitcoin electricity consumption (*e.g.* Stoll *et al.* 2019). As the usefulness of the information from IPO files declines with time (because new and more efficient mining units are being produced, which were not accounted in the files), this analysis makes use of a simplified method of choosing the mix of the most efficient available ASICs at each point in time. The technical specification of ASICs, as well as the dates of their release to the market, were taken from a list constructed by Zade & Myklebost (2018) and from Asicminervalue.com[8], where all the necessary information is available. Additionally, data on release dates and efficiencies were further validated against other websites[9] listing the mining hardware. The first publicly sold ASIC was released in May 2013, which sets the starting point for the analysis. The efficiencies of previously used FPGAs are hard to verify and also the introduction of ASICs meant a few-orders of magnitude improvement in the hashing power, as well as in the mining efficiency, which makes it a natural staring point[10].

[7] https://www.bitcoinblockhalf.com/

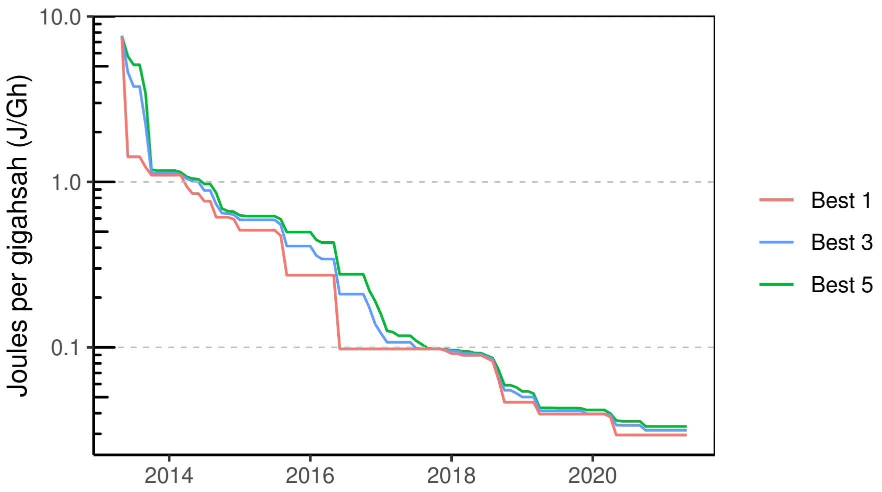
[8] https://www.asicminervalue.com/efficiency/sha-256

[9] https://cryptomining.tools/compare,
https://en.bitcoin.it/wiki/Mining-hardware-comparison,
https://www.bitcoinmining.com/bitcoin-mining-hardware/

[10] Note that the list of ASICs, that was used for further analysis, as well as the dataset, are appended to the thesis as a standalone file.

Three options for the energy efficiency were created: the efficiency of the best available ASIC, simple arithmetic mean of efficiencies of the three best available ASICs and of the five best available ASICs in time (see Figure 3.1). From these three alternatives, the most appropriate one was selected by us-

Figure 3.1: Efficiency of the mining hardware



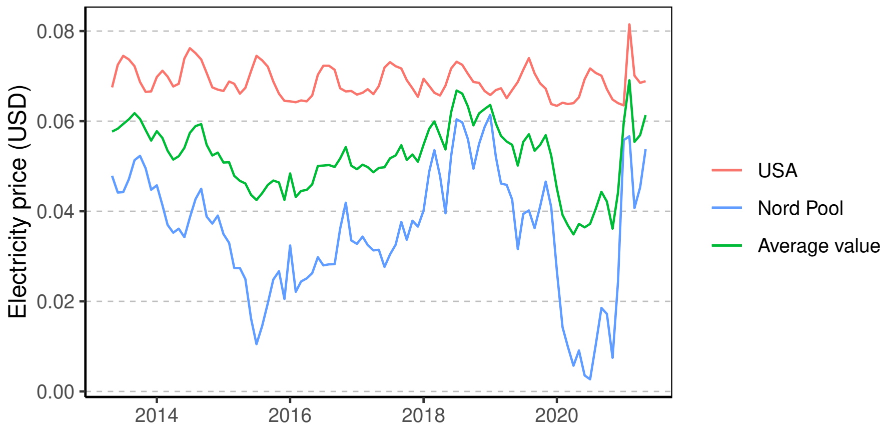

ing a comparison with the estimated electricity consumption computed by the CBECI[11]. The CBECI electricity consumption was used as a benchmark because the methodology used in this estimate is well described and robust (see 2.2.3 for detail), but it uses data that is not publicly available. Additionally, the estimate is continuous through time, which makes the comparison possible. As for the comparison itself, three versions of the electricity consumption were computed based on the three scenarios for the mining efficiency. Then, using the Root Mean Square Error (RMSE), they were compared to the CBECI electricity consumption estimate and a series with the smallest RMSE value was selected. Based on this approach, the arithmetic mean of the five most efficient ASICs was the most preferable to be used in further analysis.

Similarly to the mining efficiency, the electricity prices that miners have to pay are not known, as most miners keep it secret. Therefore, it can be only speculated about based on unverifiable claims, or a proxy such as industrial electricity price index can be used. In this work, the latter alternative was opted for, because using unverifiable data might bring a strong element of uncertainty to the analysis. Fantazzini & Kolodin (2020) used only the equilibrium price of Nord Pool, a power exchange operator situated in Northwestern Europe.

[11] See Rauchs *et al.* (2020b) or https://cbeci.org/.

Kristoufek (2020), on the other hand, used industrial electricity prices in several countries that were considered the main players in Bitcoin mining at the time of writing the paper (among others including the Nord Pool prices for Estonia and Sweden and the USA industrial electricity price index). However, as Bitcoin

Figure 3.2: Mining electricity price



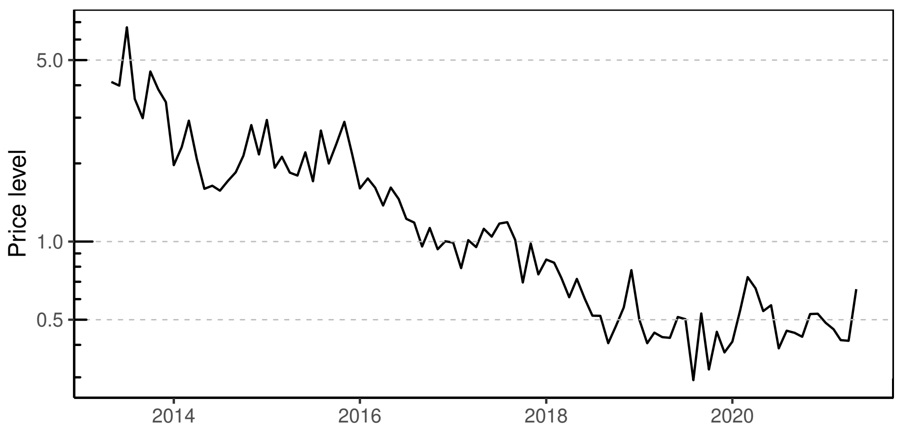

mining is a very dynamic business, the location of miners might have changed (implied *e.g.* by Rauchs *et al.* (2020a)). Therefore, for the purpose of this analysis, the System clearing price from the Nord Pool website[12] was used, as it was supported both by Kristoufek and Fantazzini & Kolodin. Additionally, the USA Average industrial electricity price[13] was used, as it was supported by Kristoufek and Rauchs *et al.* consider the United States to be one of the major mining locations. The final electricity price is the average of the electricity price from these two sources (see Figure 3.2). The electricity price from the Nord Pool exchange is provided in Euros, thus the Euro-USD exchange rate was needed to convert both prices to a common unit. The exchange rate was taken from the website[14] of the Federal Reserve System of the USA (FED).

The price level of Bitcoin was constructed in line with the economic theory described by Kristoufek (2019), as a ratio of the total transaction volume and the number of transactions. Similarly, as in 3.4.1, both of these variables were easily accessible on Blockchain.com thanks to the properties of the Bitcoin network. The transaction volume (measured in BTC) expresses how many Bitcoins are sent through the network in one day. The number of transactions is accessible on Blockchain.com in two forms: the number of confirmed transac-

[12]https://www.nordpoolgroup.com/
[13]Provided by the U.S. Energy Information Administration, https://www.eia.gov/
[14]https://www.federalreserve.gov/

tions in one day and the number of confirmed transactions in one day excluding those transactions, that were sent from or to the 100 most popular addresses. The reason for the exclusion of the 100 most used addresses is to block the effect of people sending Bitcoins to or withdrawing Bitcoins from the crypto-exchanges, keeping only the information about real peer-to-peer transactions. However, similarly, as in Kristoufek (2019), the difference between those two

Figure 3.3: Price level of Bitcoin




options in the analysis was negligible, therefore they were averaged out. The final price level was then obtained by dividing the transaction volume by the number of transactions on any given day (see Figure 3.3).

### 3.4.3 Search volume and USD money supply

The remaining variables to be described are the USD money supply and the Google Trends data. The USD money supply is published by the FED weekly as the trend-adjusted or the not-adjusted M1 and M2. Recently, the composition of the M1 was changed, which made it not directly comparable to its older values, thus the not-adjusted M2 was used in the analysis.

The Google Trends[15] statistics was in the previous studies[16] often used as a sign of public interest in Bitcoin. Therefore, the search data for the term "bitcoin"[17] downloaded from Google Trends were used as one of the variables. Data have monthly frequency and are normalized, with the point in time when the attention was highest being equal to 100, and the point with the lowest

[15] https://trends.google.com/

[16] See for example Garcia *et al.* (2014); Kjærland *et al.* (2018).

[17] Google Trends search engine does not differentiate between lowercase and capital letters, therefore the search statistics for the terms "bitcoin" and "Bitcoin" are the same.

attention being equal to 1. The Google Trends data, together with the data on electricity price, are the only two variables in the analyzed dataset that are reported with monthly frequency, which set the overall frequency for the analysis. Other variables that are reported more often, were averaged over the period of one month. Figures depicting the USD money supply and the search volume are provided in Appendix A.

# Chapter 4

# Reporting the analysis results

In this chapter, the process of choosing the methodological approach, that will be used for the analysis, is described, the analysis results are reported and commented on and various statistical tests, together with their results, are described.

## 4.1 Choosing the regression method

In the methodology section (3.3), two viable options were proposed for regressing the dependent variables on the independent ones: the 2SLS or the 3SLS estimators. For choosing the one that better fits the needs of the analysis, the Hausman test can be utilized, as specified by Hausman (1978). The $H_0$ of this test states that all exogenous variables are uncorrelated with all disturbance terms. Under this hypothesis both the 2SLS and the 3SLS estimators are consistent but only the 3SLS estimator is asymptotically efficient. Under the alternative hypothesis the 3SLS estimator is inconsistent while the 2SLS is consistent (Henningsen & Hamann, 2007). The test was performed with 20 degrees of freedom, the test statistic being equal to 32.81, which corresponds to a p-value of 0.0354. Therefore, the null hypothesis can be rejected with a high level of confidence and the analysis can proceed using the 2SLS estimator.

The 3SLS estimator seemed like a promising option as a problem that it is supposed to solve (correlated error terms across equations) did not appear to be an unrealistic scenario. However, according to the results described above, the 2SLS is a preferable option

It should be noted that the same Hausman test can be used for comparing the 2SLS estimator with the OLS estimator. When the test was performed, the results indicated the OLS estimate to be the more efficient option, signifying thus that the OLS should be used. However, the logic of working of the Bitcoin system and the previous research clearly points to the endogeneity being present (at least within the price-hashrate relationship), and, as it turns out, the Durbin-Wu-Hausman test confirms this (see the next section for details). The contradicting results of the Hausman tests might be caused by the relatively small number of observations (97), or by the subtle changes of daily data values that were smoothed by the monthly averaging. In the next section, the results of various statistical tests of the 2SLS estimator are presented.

## 4.2 Statistical tests

For the chosen 2SLS estimator, a battery of statistical tests was performed.

To confirm the assumption of the endogeneity being present in the case of the four variables specified above, the Durbin-Wu-Hausman (DWH) test was performed, equation by equation. The DWH test consists of two steps. In the first step, the variable that is suspected to be endogenous is regressed on a set of exogenous variables, and fitted values of the (presumably) endogenous variable are used to obtain the residuals. In the second step, the originally proposed structural equation is estimated, with the tested endogenous variable included on the right-hand side, together with exogenous variables and residuals from step one. Then, if the estimator of the residuals[1] is significantly different from 0, the $H_0$ that a tested variable is exogenous can be rejected (Hausman, 1978).

This procedure was carried out for the four endogenous variables from the model. The Bitcoin price, the hashrate, the transaction fees, and the search volume were regressed on exogenous variables from equations 3.1, 3.2, 3.3, 3.4 respectively and on the four instrumental variables (electricity price, total Bitcoin supply and lagged values of the price and the hashrate). Next, residuals thus obtained were added[2] as explanatory variables to the structural equations, which were then estimated by the OLS one by one. In all cases, the estimators of the residuals were statistically significant (with $p = 0.025$ or lower), therefore

[1]Residuals from the first step that are included in the structural equation can be thought of as "a part of the tested variable", that can not be explained by exogenous variables and thus comes from the system, *i.e.* is endogenous. If this part is statistically significant in explaining the variance of a left-hand side variable, the variable in question can be assumed endogenous.

[2]Residuals of an endogenous variable were added to a structural equation, only if the endogenous variable was present in the structural equation.

it was concluded that the assumption of the four variables being endogenous holds.

Next, it was tested for the stationarity of residuals, as non-stationary residuals (or residuals containing a unit root) in the context of a time series might, according to Wooldridge (2015), lead to a problem of spurious regression. Three tests were employed for this, the first of them being the Augmented Dickey-Fuller (ADF) test. The ADF tests the null hypothesis that a unit root of a time series is present (Fuller, 1996). Next, the Phillips-Perron (PP) test was employed, testing the same null hypothesis (Phillips & Perron, 1988). But, unlike the ADF test, the PP test is robust to the unspecified autocorrelation and heteroscedasticity in the errors. Finally, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test with a null hypothesis of tested time series being stationary was used (Kwiatkowski *et al.*, 1992).

Table 4.1: Tests for stationarity

| Equation | ADF | PP | KPSS |
|---|---|---|---|
| Price | 0.01 | 0.01 | 0.1 |
| Hashrate | 0.01 | 0.01 | 0.1 |
| Fees | 0.0599 | 0.01 | 0.1 |
| Search | 0.01 | 0.0123 | 0.1 |

0.01 for $p \leq 0.01$; 0.1 for $p \geq 0.1$

The results of the ADF, PP, and KPSS tests are summarized in table 4.1 (the p-values are shown). Note that the PP test and the KPSS test have both options to test for a series with no drift and no deterministic trend, series with a drift and without a trend, or a series with both a drift and a trend. In Both the PP and the KPSS test, the test result for the second option is displayed, however, the other two options did not differ significantly. In almost all cases the null hypothesis of the unit root was rejected with a high level of confidence, in the case of the ADF test of the Transaction fees equation, the confidence is still high ($p = 0.0599$). The null hypothesis of stationarity in the KPSS test could not be rejected on any major level of significance. Therefore, it could be assumed that the residuals are stationary and the analysis may proceed as intended.

The following step was to check for the fundamental assumption of the time-series analysis, *i.e.* for the normality, heteroskedasticity, and for the remaining serial correlation (autocorrelation) of error terms (residuals). The normality was tested with the Doornik-Hansen test (Doornik & Hansen, 2008), which has a null hypothesis of normal distribution. The test has an option to test for univariate or for multivariate normality. The multivariate normality was tested, resulting in $\chi^2(8) = 38.744$, which corresponds to $p < 0.001$. Therefore, the multivariate normality was rejected with a high level of significance. In the case of univariate normality, the null hypothesis of normality could not be rejected in the case of the price and the hashrate equation, however, it was rejected for the remaining two equations. Thus, overall, residuals of the model cannot be considered to have a normal distribution. However, as there are close to 100 observations (97 in total), the residuals still can be assumed to be at least asymptotically normal, which is sufficient.

In order to test the homoskedasticity of residuals, the Breusch-Pagan Test was used (Breusch & Pagan, 1979). The null hypothesis of this test says that there is no heteroskedasticity present. According to the results summarized in

Table 4.2: The Breusch-Pagan test

| Equation | $\chi^2$ | DF | *p-value* |
|---|---|---|---|
| Price | 9.91 | 6 | 0.126 |
| Hashrate | 24.63 | 6 | $< 0.001$ |
| Fees | 21.064 | 3 | $< 0.001$ |
| Search | 2.922 | 1 | 0.087 |

the table 4.2, the homoskedasticity was rejected in all but the first equation.

For the purpose of detecting autocorrelation of residuals, the Durbin-Watson test was used (Durbin & Watson, 1971) with the null hypothesis that the autocorrelation of disturbances is 0. The test returns a test statistic between 0 and 4, values close to 0 signify positive autocorrelation, values close to 4 negative autocorrelation and values close to 2 no autocorrelation[3]. The test in-

Table 4.3: The Durbin-Watson test

| Equation | DW value | *p-value* |
|---|---|---|
| Price | 1.201 | $< 0.001$ |
| Hashrate | 0.806 | $< 0.001$ |
| Fees | 0.481 | $< 0.001$ |
| Search | 0.432 | $< 0.001$ |

dicates that residuals of all four equations are autocorrelated with a high level

[3]A rule of thumb says that values between 1.5 and 2.5 mean autocorrelation weak enough that it may not need to be adjusted for.

of probability (the autocorrelation of residuals in the case of the price equation is weaker than in other equations). It should be noted that in the case of the Durbin-Watson test, as well as the Breusch-Pagan test, the model was tested equation-by-equation.

The results of the last two tests suggest that the model might be miss-specified in some way, as only the first equation passed both of them (to a certain degree, to say the least). To account for this potential issue at least partially, the heteroskedasticity and autocorrelation consistent (HAC) standard errors shall be used in the estimation instead of the unadjusted standard errors. Since the residuals were assumed to be asymptotically normally distributed, this solution is feasible. A superior solution would be an inclusion of explanatory variables that might capture the dynamics in the model better than currently used variables. However, as discussed in section 3.2, this is not an option due to unavailable data. The next section briefly summarizes the findings of the model.

## 4.3 Results

In the table 4.4, the results of the 2SLS estimation are reported, together with the HAC standard errors and respective t-statistics and p-values.

The results regarding the price-hashrate relationship are surprising. The effect of the hashrate on the price is positive and significant ($p = 0.02$), which was expected, however, the effect of price, which was assumed to be an important driver of the hashrate, is statistically insignificant ($p = 0.331$). The theory[4], as well as previous research, says that the price of Bitcoin is a motivation for miners to mine and thus it should be significant in explaining the hashrate changes. Despite that, the research conducted hereby could not reject the hypothesis that the effect is equal to zero.

The effect of the addresses on the price is negative and significant ($p = 0.015$), which is a contradicting result compared to expectations. The effect of the price level, however, being also negative and significant ($p < 0.001$), was expected. Growth of the search volume has a positive significant ($p < 0.001$) effect on price. The expectation was not clear, as both directions in

[4]In a hypothetical situation when the Bitcoin price would drop to zero, miners would receive effectively no reward for their activity and would be thus forced to exit the mining industry. The total hashrate would thus also converge to zero, as all the mining would be funded by miners and would yield absolutely no profit.

the relationship are logical, however, the positive one was confirmed also by previous research. The effect of the increasing money supply of the USA is significantly positive ($p < 0.001$), which is reasonable, as it leads to more dollars that can be spent on Bitcoin. The effect of an increase in transaction fees is negative and significant ($p = 0.007$), which is interesting information, as the expectations were mixed, both effects being possible.

The effect of the price on the hashrate is statistically insignificant, as already discussed. The effect of an increase of the number of active addresses on the hashrate is positive, which was expected, and also quite strong ($\beta_8 = 2.6$, $p < 0.001$). The improving efficiency of mining increases hashrate, which is logical[5] ($p < 0.001$). The effect of the mining phase did not reach any major level of significance ($p = 0.314$ and $p = 0.305$), therefore it can not be considered different from 0. However, a negative effect would be logical, as it means lower rewards for miners. Note that the effect of phase one is hidden in the intercept. The effect of transaction fees on the hashrate is significantly negative ($p < 0.001$), which is surprising, as an increased reward for miners should lead to an increase in hashrate.

The expectations regarding the effect of price changes on the transaction fees were not clear, as both variants would be reasonable. The model indicates a negative significant relationship ($p < 0.001$), which could be explained as a fear of loss is stronger than a fear of lost opportunity (see the next chapter for a detailed discussion). The number of active addresses affects the transaction fees positively ($p = 0.006$), which is logical, as increased demand for completing transactions unavoidably leads to high transaction fees. The effect of high search queries on transaction fees is also significantly positive ($p < 0.001$), which could be expected as well, because more people interested in the technology probably leads to more people sending transactions.

In the last equation, the effect of price increases on the search volume was estimated to be positive and significant ($p < 0.001$). The expectations were not clear, as both directions of the relationship would be sensible, however, the positive effect was indeed more probable.

All the explained (right-hand side) variables are log-transformed, therefore interpreting the estimated equation intercepts is not very informative. In the next section, the model results are analyzed and their implications are developed.

[5] Note that the efficiency is measured in J/GH, therefore decreasing the absolute value of the variable means improvement in efficiency.

Table 4.4: The 2SLS analysis results

| | $\widehat{\beta_j}$ | SE | *t-statistic* | *p-value* |
|---|---|---|---|---|
| **Price equation** | | | | |
| Intercept | −8.93 | 3.017 | −2.959 | 0.003 |
| log(hashrate) | 0.09 | 0.039 | 2.338 | 0.02 |
| log(addresses) | −0.38 | 0.153 | −2.453 | 0.015 |
| log(price_level_average) | −0.68 | 0.079 | −8.59 | <0.001 |
| log(search_volume) | 0.97 | 0.019 | 50.421 | <0.001 |
| log(USD_M2) | 1.91 | 0.227 | 8.427 | <0.001 |
| log(transaction_fees) | −0.08 | 0.03 | −2.687 | 0.007 |
| **Hashrate equation** | | | | |
| Intercept | −18.87 | 2.316 | −8.149 | <0.001 |
| log(BTC_price) | 0.14 | 0.149 | 0.972 | 0.331 |
| log(addresses) | 2.60 | 0.179 | 14.563 | <0.001 |
| log(efficiency) | −1.38 | 0.189 | −7.328 | <0.001 |
| reward_phase2 | −0.21 | 0.212 | −1.009 | 0.314 |
| reward_phase3 | −0.29 | 0.284 | −1.027 | 0.305 |
| log(transaction_fees) | −0.71 | 0.062 | −11.341 | <0.001 |
| **Fees equation** | | | | |
| Intercept | −8.05 | 4.543 | −1.772 | 0.077 |
| log(BTC_price) | −0.94 | 0.269 | −3.485 | <0.001 |
| log(addresses) | 1.25 | 0.447 | 2.806 | 0.006 |
| log(search_volume) | 1.43 | 0.28 | 5.102 | <0.001 |
| **Searches equation** | | | | |
| Intercept | −2.04 | 0.358 | −5.693 | <0.001 |
| log(BTC_price) | 0.55 | 0.049 | 11.083 | <0.001 |

| | $R^2$ | Adj. $R^2$ |
|---|---|---|
| Price eq.: | 0.987 | 0.986 |
| Hashrate eq.: | 0.989 | 0.988 |
| Fees eq.: | 0.583 | 0.569 |
| Search eq.: | 0.857 | 0.855 |
| Num. obs. (total) | 97 | |

# Chapter 5

# Discussion

This chapter contains a discussion of the estimated model, an analysis of revealed relationships between variables. Additionally, the further development of the hashrate and the electricity consumption is suggested, tying the model results together with the discussion of environmental impacts of Bitcoin mining from the second chapter (section 2.2). Lastly, the limitations of the model are described and potential improvements, as well as possible avenues for further research, are proposed.

## 5.1 Implications of the model

The estimated effect of the hashrate on the Bitcoin price is in line with previous research, however, an overall stronger relationship than revealed ($\widehat{\beta_1} = 0.09$) was expected. As mentioned in 3.3, a drop of hashrate from the first half of 2021 of nearly 50% was followed by a price dropping approximately by 40%. According to the estimate, a 50% decrease of hashrate results in a price decrease of only 6%. Therefore, it can be concluded that in this particular example, the price decrease was probably a consequence of several phenomena, the hashrate drop being only one of them. The same conclusion holds in general, the hashrate is likely to be only one of several factors that cause the dynamic changes of the Bitcoin price.

The other side of this coin is the effect of the price changes on the hashrate. However, no comments can be made on this relationship based solely on the results of the proposed research due to a low significance level. It can be noted that the introduction of the heteroskedasticity and autocorrelation consistent standard errors rendered the estimate insignificant ($p = 0.331$), while when computing the t-statistic and subsequent p-value from unaltered standard errors, the estimator $\widehat{\beta_7}$ seemed to be significant (in such case $\widehat{\beta_7} = 0.14$ with $p = 0.035$). This might indicate that the heteroskedasticity and autocorrelation truly are present in the data and that they need to be accounted for. Based on the insignificant estimator, it can be argued that the assumed endogeneity of price and hashrate is not very strong or the price is on the edge of being endogenous (it can be noted that the Durbin-Wu-Hausman test rejected the exogeneity of price in hashrate equation, however out of all endogenous variables with the lowest probability). The relationship from at least one side might be weak or non-existent. This could cause the estimate to be insignificant. Another possible interpretation could be as follows. The Bitcoin system is unique in its interconnectedness. An exogenous shock on one variable (example could be the already discussed ban on Bitcoin mining in China from 2021) should, according to pure logic, have no effect on another variable (the ban should not affect the price, as for an end-user, nothing has changed, the security of the network remained extremely high and the time to process a transaction increased only slightly and only for a short period of time). However, as it turns out, it is probably the fear of investors that decreases the price. This results in a seeming correlation of variables, which however is not a result of a causal relationship, but is caused by external events that cannot be (or only with a high level of imprecision) captured by the data, leading thus to unexpected result such as the one described above. When the previous research findings and workings of the Bitcoin network are taken into consideration, the expectation of price being a strong motivation for hashrate remains unchanged.

One of the unexpected conclusions is that the number of active addresses on the network decreases the price, which is in contradiction with the findings of Wheatley *et al.* (2019). It would mean that active use of the network for sending transactions, not just holding Bitcoins on addresses passively for speculative purposes, lowers the price of a single Bitcoin. It is hard to explain why such a thing should be a case since a higher price of one Bitcoin should theoretically pose no limitations on its use. One possible reason might be that those users, who value Bitcoin less than is its current market price feel no restrictions on selling it or using it as a currency (*e.g.* paying with it for goods and services), and those who value Bitcoin more than is its current market price hold on in, expecting the price to increase in the future, thus not participating in the set of active addresses. The difference between the estimate of this study and Wheatley *et al.* (2019) might be caused either by newer data, that were used for the analysis (as in 2021 the Bitcoin price experienced the largest increase since the renowned peak at the end of 2017) or by the inclusion of a different set of variables. Alternatively, it could be caused by an inaccurate model specification, as the relationships in the bitcoin network are tangled and the causal relation of some variables might have not been identified correctly.

The search volume was shown to have a positive effect on the price, which corresponds with the previous research results. Both directions in this relationship would be reasonable, however, the positive effect seems to be stronger. It can be interpreted as people are more prone to buy Bitcoin after searching for information about it than to sell it after searching it. Increased attention towards the technology is likely to attract new investors, who then increase the price by increasing demand for Bitcoins.

A noteworthy finding is a positive (and rather strong: $\widehat{\beta_4} = 1.91$) effect of the US money supply on the Bitcoin price. This time series has not been used in the previous research[1] as an explanatory variable of the price, and, as it turns out, it is highly significant. This could be interpreted as the Bitcoin truly serves as a hedge against inflation. The notion of Bitcoin as a form of "digital gold" was strengthened during the Covid-19 pandemic, since high price increases were often thought to be connected to extraordinary quantitative easing in the USA, and the results of the analysis seem to agree.

The effect of transaction fees on the price is negative, which might be explained by users' unwillingness to pay extraordinarily high fees. High transaction fees seem to discourage investors from buying Bitcoin, lowering thus demand and also the price. An opposite relationship would be also reasonable, high fees might drive price high, and they might be interpreted as users are eager to use the network despite the high fees. However, the unwillingness to pay for transactions seems to dominate.

The effect of active addresses on the hashrate is positive and quite strong ($\widehat{\beta_8} = 2.6$). The strength of this relationship was unexpected, as the hashrate was not assumed to be connected to the addresses very tightly. Miners were expected to mine regardless of whether in some time $t$ the number of active users is the same, smaller, or larger than in $t-1$ (or generally $t-k$). Nonetheless, it turns out that it is an important factor. One possible explanation might be that the expected positive effect of price on hashrate is hidden in the effect of

[1]It has not been used at least in terms of the papers that were assessed for the purpose of this work. A study exploring this relationship was not found, however, it does not mean it has never been conducted.

addresses. Further implications of this relationship are explored in the next section.

Another estimate, that goes against the expectations, is the effect of the transaction fees on the hashrate. It would be reasonable to expect a positive relationship since the fees are clearly a motivation for miners to spend money on electricity (although it is a small part of the motivation, compared to the mined Bitcoins). However, the estimate is negative. A feasible reason for this effect to be negative is that the transaction fees are measured in Bitcoin, while the miners are expected to be motivated primarily by revenues in USD (they need to pay for electricity and other expenses regularly, for which they need common currencies). This situation results in that in times of strong price increases, the transaction fees might decrease in Bitcoin-terms, however, increase when expressed in USD, and thus also the revenue of miners decreases when measured in Bitcoin and increases when measured in USD. When the price is falling, the situation might be reversed, fees are lower in USD-terms, but higher in Bitcoin-terms, same as the revenue of miners, thus resulting in a lower hashrate. However, it should also be noted that if such interpretation is not true, it might be the case that an unobserved effect plays a role, some causality that was not identified by the model might have disrupted the results.

The effect of the Bitcoin price on the total transaction fees was estimated to be negative. The expectations were uncertain, the estimate thus brings a useful insight. The negative relationship implies that transaction fees grow when the price drops. This might be explained as in case of a sudden price drop, users are afraid of loss in form of their Bitcoins losing value even more and thus send their coins on exchanges to exchange them as fast as possible, despite high transaction fees. If the situation would be reversed and the estimate would be positive, the explanation might be that people are afraid of lost opportunity and buy Bitcoin as fast as possible, disregarding the high transaction fees. Empirically, both notions, fear of loss as well as fear of lost opportunity can be observed. Despite that, since the estimate is negative, the fear of loss seems to be stronger.

Similarly, there was no clear expectation regarding the effect of the price changes on the search volume, as the search volume can be expected to increase when the price suddenly moves both up and down. The estimate indicates a positive relationship, which can be explained as people being interested more in positive news (price increases) than the negative ones (price drops).

The effects that were not commented on in this chapter were expected and the results only confirmed the previous research (this is the case of the price level in the first equation, the mining efficiency in the second equation and the number of active addresses, and the search volume in the third equation). The interpretation of such effects is already proposed in section 3.2. The estimated effect of a halving of the block reward (represented by the variable reward phase) on hashrate was shown to be insignificant, therefore no conclusions can be made on this relationship. The next section further discusses the effect of explanatory variables on hashrate through the lens of its environmental impact.

## 5.2 Environmental outlook of the Bitcoin network in future

The results of the analysis described above revealed the effects of various explanatory variables on hashrate. These effects can be put into perspective of the electricity consumption of Bitcoin, as the hashrate is one of the main factors that drive the total amount of electricity that is consumed by the Bitcoin network. The total electricity demand of the network is equal to the total hashrate divided[2] by the overall efficiency of all mining units.

The mining efficiency was also shown to have a negative effect on hashrate (*i.e.* improvements in efficiency lead to higher hashrate). Therefore, it can be assumed that these two consequences of the mining efficiency will at least partially nullify themselves, improvements in efficiency will lead to less consumed energy, but simultaneously it will lead to more hashrate being deployed, therefore increasing the energy demand. Also, it was shown that the rate at which the efficiency is improving has slowed down dramatically in the last few years. For these reasons, the hashrate can be expected to remain the most important and dynamic factor of the total electricity demand.

The intuition says that the Bitcoin price should be the main driver for hashrate in the long-run, as it determines how much total revenue will be distributed among miners. However, as the estimated effect of the price is statistically insignificant, it can not be relied upon.

The two remaining significant estimates presented by this thesis are the effects of active addresses and of transaction fees. However, both should be taken with caution, as they both go slightly against the expectations, as explained previously. Based on these two factors, the future development of hashrate can

[2]Or multiplied, depending on how the efficiency is measured.

be estimated[3]. The number of active addresses can be extrapolated into the future more easily since there is a natural limitation posed to this number by the Bitcoin algorithm, therefore the outlook for the hashrate in the future will be based on this variable.

The maximum theoretical number of transactions in one day is approximately 600000, as due to the limited block space of 1 MB and the time between two blocks of 10 minutes there can be at most 7 transactions per second (Croman *et al.*, 2016). From the start of 2017 until the half of 2021, there were, on average, less than 550000 active addresses per day, while at the highest point, the number reached almost 1 million. The theoretical maximum of active addresses per day is 1200000 (maximum number of transactions multiplied by two, as each transaction needs a sender and a receiver). However this is in fact extremely unlikely, as many transactions are made by people who send their funds to exchange platforms, therefore not using a unique address on the receiver's side (even if many people decide to exchange cryptocurrencies, the receiving address will be the one of the exchange platform, therefore not unique). Additionally, if Bitcoin should be used as a currency more widely, merchants receiving payments for goods and services on their addresses would also reduce the number of unique active addresses used every day (however, this is in contradiction with the notion of Bitcoin being digital gold, which is used for storing value, not for day to day purchases). Therefore, it could be assumed, that in the case of extreme Bitcoin popularity in the future the average number of unique active addresses per day could double. This would result, according to the estimators, in hashrate increasing approximately 20 times (taking into account the effect of addresses on transaction fees and the consequent effect of transaction fees on hashrate). This is a lot, however, it is far from the catastrophic scenario indicated by Mora *et al.* (2018). And it should be kept in mind that those calculations are for the case of a theoretical maximum of the number of active unique addresses (which is in reality highly unlikely).

In such a case, the electricity demand of the Bitcoin network would increase. However, this does not have to directly imply that a carbon footprint of Bitcoin would be bigger, as a second and just as important part of the environmental question is the source of the energy that is used for mining. As it was noted by many researchers, the mining can be expected to shift from areas with a

[3]Keep in mind that such estimation is just an educated guess and should be treated as such.

large portion of electricity coming from fossil fuels to areas with a high share of renewable sources, since in the long-run, miners have to optimize for the cheapest electricity possible (which is expected to come from renewable energy). However, a detailed discussion of this topic is beyond the scope of this work.

In cases of Bitcoin popularity stagnating or declining, the hashrate could be expected to follow and either stagnate in a range of contemporary values or fall. No signs of this though cannot be seen consistently, thus the popularity of Bitcoin is expected to rise in the future. A discussion of the limitations of the analysis and of the areas worth attention of researches follows.

## 5.3 Limitations and further research

In this section, firstly the limitations of this work are described, and secondly, options for further research are proposed.

In the analysis, four variables were considered endogenous (the price of Bitcoin, the hashrate, the total transaction fees, and the search volume) and other variables were considered exogenous. In reality, however, the distinction is not entirely clear. Some variables are given by the Bitcoin algorithm or independent institutions, so there is no doubt about their exogeneity (*e.g.* the reward phase is dictated by the algorithm, and the M2 money supply is given by the FED). But some other variables, especially the number of active addresses or the price level, might in fact be at least partially driven by price (or some other variable) and be thus endogenous. In such a case, the resulting estimators might be biased. The relationships in the Bitcoin system are extremely tangled as causalities and correlations are crossing each other, and variables might be affecting each other in unexpected ways. Identifying all the relationships correctly is thus difficult.

Another inaccuracy might be caused by the way, in which the variables representing the efficiency of mining and the electricity price incurred by miners are constructed. Both of these variables are simplifications of reality, as the true data are close to impossible to collect. The true efficiency of the mining hardware is likely to be worse than assumed, since the efficiencies of 5 best available units were taken, and the true prices of electricity are likely to be lower than assumed, as it is believed that the large-scale mining companies are optimizing the electricity price as much as possible, *e.g.* by situating the mining facilities in locations with extremely cheap electricity (cheaper than the industrial average). This imprecision might have also caused a slight bias in the analysis.

The variable representing the number of unique active addresses might be questionable. In the analysis, it was used as a proxy for the number of users of the network. As was already mentioned, there is a limitation for the number of transactions per second, that can be completed in the network, due to the limited block space. To mitigate this problem, the community of Bitcoin developers works on the second layer network for Bitcoin transactions, the so-called “Lightning network”. The advantage of it is that transactions completed in the Lighting network are much faster than transactions from the first layer (blockchain) and are not limited in numbers. However the network is less secure (Lee & Kim, 2020), so the Lighting network is expected to be used for small transactions (*e.g.* for payments for goods) and the blockchain directly will be used for large transactions. The transactions from the Lightning network were not included in the analysis, as it is still considered to be in the early stage of adoption by Bitcoin users, however, it can be expected that it will form a non-negligible portion of active users in the future. Thus, a better proxy variable for active users should be considered in further research. Additionally, the price level, which is computed as the volume of Bitcoin transactions divided by their number, will be also affected, as transactions from the Lightning network will contribute to these measures as well.

Apart from quantifying transactions on the Lightning network, further research could concentrate on estimating the efficiency of the used mining hardware and the price of electricity used for mining with the highest precision possible. It would make the analysis of the drivers of the price and the hashrate more accurate, as these two variables were the strongest assumptions. It would also be helpful for the discussion of the environmental impacts of Bitcoin mining. For the sake of this topic, also estimating the locations and thus also energy sources of miners would also be immensely beneficial, as existing estimates are themselves based on strong assumptions, and also are likely to get outdated.

Also, performing a similar study like the one proposed using data with a higher frequency (*e.g.* weekly or even daily) might reveal the relationships more clearly, since it could better reflect the subtle changes in data. The problem of questionable endogeneity might thus be resolved at least to a certain degree. The main bottleneck for the data frequency was the monthly reports of the Google Trends search volume and the electricity prices from the USA. Perhaps a different set of variables with similar relations to the Bitcoin system might be feasible to use.

# Chapter 6

# Conclusion

The main motivation of the thesis was to identify the factors, that drive the price of Bitcoin and the total hashrate of the Bitcoin network and to explore the mutual relationship of these two variables. For this purpose, a system of equations was built, with the aim of including explanatory variables that would be able to capture the dynamics of the examined relationship. Since there were strong reasons to assume that endogeneity might be present in the case of some explanatory variables, the Two-stage least squares, and the Three-stage least squares estimators were proposed as a solution, and based on an appropriate statistical test, the method of Two-stage least squares was chosen as a more reliable option.

In particular, four equations were created, one explaining the price of Bitcoin, one for the total network hashrate, one for the transaction fees paid in the Bitcoin system, and one for the search volume for term "bitcoin" (representing the public interest in Bitcoin), as each of those four variables was assumed to be endogenous, which was later confirmed by a statistical test.

Data on some of the variables were not directly accessible and were thus constructed for the purpose of the thesis. This was the case of the efficiency of the mining hardware, the electricity price incurred by miners, and the price level of Bitcoin. Time series with a total of 97 observations were used, with a monthly frequency.

To a large degree, the results of the analysis confirmed expectations based on the previous research results, however, deviations were also found. A unique set of variables was constructed to be analyzed simultaneously, which helped to uncover some relations. The hashrate was found to drive the Bitcoin price positively, however, the effect of the price on the hashrate was statistically insignificant, therefore no conclusion on this relationship could be drawn based on the analysis. It was argued that such an inconclusive result might have been caused by the price being on the edge of endogeneity and also possibly by unobserved exogenous shocks to the whole Bitcoin system. The money supply of the United States dollar, a variable previously not evaluated in a similar type of analysis, was found to positively affect the Bitcoin price. The number of unique active addresses on the Bitcoin network was found to negatively affect the price, which is a result contradicting the previous research, and a negative effect of transaction fees on the hashrate was revealed. Both these results were not expected, however, in the context of the Bitcoin system they are reasonable and can be well explained.

In addition, the environmental effects of Bitcoin mining were briefly discussed, considering the results of the proposed analysis. It was concluded that opinions envisioning Bitcoin as an instrument of a huge environmental catastrophe might be exaggerated.

To conclude, the analysis found some of the drivers of the Bitcoin price and the hashrate, however, no definitive conclusions could be made on the price-hashrate relationship. Two other endogenous factors were identified and the USD money supply was shown to be a significant driver of the price, which might be useful in the future for explaining the price changes. However, it is up to further research to uncover with higher precision, whether the price is truly endogenous or not, and what is its exact relation to hashrate.

# Bibliography


ACHESON, N., L. LEWITINN, & G. MOORE (2020): "Coindesk Asset Data Coverage: Asset Selection Methodology." *Coindesk Research* .

ANDROULAKI, E., G. O. KARAME, M. ROESCHLIN, T. SCHERER, & S. CAPKUN (2013): "Evaluating user privacy in Bitcoin." In "International Conference on Financial Cryptography and Data Security," pp. 34–51. Springer.

ANTONOPOULOS, A. M. (2014): *Mastering Bitcoin: unlocking digital cryptocurrencies*. O'Reilly Media, Inc.

BALTAGI, B. h. (2011): *Econometrics*. Springer, Berlin, Heidelberg.

BASTIAAN, M. (2015): "Preventing the 51%-attack: a stochastic analysis of two phase proof of work in Bitcoin."

BENDIKSEN, C., S. GIBBONS, & E. LIM (2018): "The Bitcoin mining network-trends, marginal creation cost, electricity consumption & sources." *CoinShares Research* **21**: pp. 3–19.

BEVAND, M. (2017a): "Electricity consumption of Bitcoin: a market-based and technical analysis." *http://blog.zorinaq.com/bitcoin-electricity-consumption* .

BEVAND, M. (2017b): "Serious faults in Digiconomist's Bitcoin Energy Consumption Index." *http://blog.zorinaq.com/serious-faults-in-beci* .

BEVAND, M. (2018): "Reviewing Morgan Stanley's Bitcoin research reports." *http://blog.zorinaq.com/morgan-stanley-bitcoin-research-reports/* .

BLANDIN, A., G. C. PIETERS, Y. WU, A. DEK, T. EISERMANN, D. NJOKI, & S. TAYLOR (2020): "3rd global cryptoasset benchmarking study." *Cambridge Centre for Alternative Finance* .

BREUSCH, T. S. & A. R. PAGAN (1979): "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* **47(5)**: pp. 1287–1294.

Carter, N. (2018): "Digesting "Quantification of energy and carbon costs for mining cryptocurrencies"." *Medium* .

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

Courtois, N. T., M. Grajek, & R. Naik (2014): "Optimizing SHA256 in Bitcoin mining." In "Cryptography and Security Systems," pp. 131–144. Springer, "Springer Berlin Heidelberg".

Croman, K., C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. Gün Sirer, D. Song, & R. Wattenhofer (2016): "On Scaling Decentralized Blockchains." In J. Clark, S. Meiklejohn, P. Y. Ryan, D. Wallach, M. Brenner, & K. Rohloff (editors), "Financial Cryptography and Data Security," pp. 06–125. Berlin, Heidelberg: Springer Berlin Heidelberg.

De Vries, A. (2018): "Bitcoin's growing energy problem." *Joule* **2(5)**: pp. 801–805.

De Vries, A. (2020): "Bitcoins energy consumption is underestimated: A market dynamics approach." *Energy Research & Social Science* **70**: p. 101721.

Dittmar, L. & A. Praktiknjo (2019): "Could Bitcoin emissions push global warming above 2°C?" *Nature Climate Change* **9(9)**: pp. 656–657.

Doornik, J. A. & H. Hansen (2008): "An Omnibus Test for Univariate and Multivariate Normality." *Oxford Bulletin of Economics and Statistics* **70(s1)**: pp. 927–939.

Durbin, J. & G. S. Watson (1971): "Testing for serial correlation in least squares regression. III." *Biometrika* **58(1)**: pp. 1–19.

Fantazzini, D. & N. Kolodin (2020): "Does the Hashrate Affect the Bitcoin Price?" *Journal of Risk and Financial Management* **13(11)**.

Fuller, W. A. (1996): *Introduction to statistical time series*, volume 428. New York: John Wiley & Sons.

Gallersdörfer, U., L. Klaassen, & C. Stoll (2020): "Energy consumption of cryptocurrencies beyond Bitcoin." *Joule* **4(9)**: pp. 1843–1846.

GARCIA, D., C. J. TESSONE, P. MAVRODIEV, & N. PERONY (2014): "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy." *Journal of The Royal Society Interface* **11(99)**: p. 20140623.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics." *Econometrica* **46(6)**: pp. 1251–1271.

HAYES, A. S. (2017): "Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing Bitcoin." *Telematics and Informatics* **34(7)**: pp. 1308–1321.

HAYES, A. S. (2019): "Bitcoin price and its marginal cost of production: support for a fundamental value." *Applied Economics Letters* **26(7)**: pp. 554–560.

HENNINGSEN, A. & J. D. HAMANN (2007): "systemfit: A Package for Estimating Systems of Simultaneous Equations in R." *Journal of Statistical Software* **23(4)**: pp. 1–40.

KHATRI, Y. (2021): "Bitcoin hashrate drops by nearly 50% following China's mining crackdown." *https://www.theblockcrypto.com/post/109315* .

KJÆRLAND, F., A. KHAZAL, E. A. KROGSTAD, F. B. G. NORDSTRØM, & A. OUST (2018): "An Analysis of Bitcoin's Price Dynamics." *Journal of Risk and Financial Management* **11(4)**.

KÖHLER, S. & M. PIZZOL (2019): "Life cycle assessment of Bitcoin mining." *Environmental science & technology* **53(23)**: pp. 13598–13606.

KOOMEY, J. (2019): "Estimating Bitcoin Electricity Use: A Beginner's Guide 1.0." *https://www.coincenter.org/estimating-bitcoin-electricity-use-a-beginners-guide/* .

KRAUSE, M. J. & T. TOLAYMAT (2018): "Quantification of energy and carbon costs for mining cryptocurrencies." *Nature Sustainability* **1(11)**: pp. 711–718.

KRISTOUFEK, L. (2013): "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era." *Scientific Reports* **3**: pp. 1–7.

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

Kristoufek, L. (2019): "Is the Bitcoin price dynamics economically reasonable? Evidence from fundamental laws." *Physica A: Statistical Mechanics and its Applications* **536**: p. 120873.

Kristoufek, L. (2020): "Bitcoin and its mining on the equilibrium path." *Energy Economics* **85**: p. 104588.

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

Lee, S. & H. Kim (2020): "On the robustness of Lightning Network in Bitcoin." *Pervasive and Mobile Computing* **61**: p. 101108.

Malone, D. & K. O'Dwyer (2014): "Bitcoin mining and its energy footprint." pp. 280–285.

McCook, H. (2014): "An order-of-magnitude estimate of the relative sustainability of the Bitcoin network." *A critical assessment of the Bitcoin mining industry, gold production industry, the legacy banking system, and the production of physical currency* **2**: p. 25.

McCook, H. (2018): "The cost & sustainability of Bitcoin." *https://www.academia.edu/37178295* .

Mora, C., R. L. Rollins, K. Taladay, M. B. Kantar, M. K. Chock, M. Shimada, & E. C. Franklin (2018): "Bitcoin emissions alone could push global warming above 2°C." *Nature Climate Change* **8(11)**: pp. 931–933.

Nakamoto, S. (2008): "Bitcoin: A peer-to-peer electronic cash system." *https://bitcoin.org/bitcoin.pdf* .

Narayanan, A., J. Bonneau, E. Felten, A. Miller, & S. Goldfeder (2016): *Bitcoin and cryptocurrency technologies: a comprehensive introduction.* Princeton University Press.

PHILLIPS, P. C. B. & P. PERRON (1988): "Testing for a unit root in time series regression." *Biometrika* **75(2)**: pp. 335–346.

R CORE TEAM (2019): *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

RAUCHS, M., A. BLANDIN, A. DEK, & Y. WU (2020a): "Bitcoin Mining Map." *https://cbeci.org/mining-map* .

RAUCHS, M., A. BLANDIN, A. DEK, & Y. WU (2020b): "Cambridge Bitcoin Electricity Consumption Index (CBECI)." *https://cbeci.org/* .

RAUCHS, M., A. BLANDIN, K. KLEIN, G. C. PIETERS, M. RECANATINI, & B. Z. ZHANG (2018): "2nd global cryptoasset benchmarking study." *Cambridge Centre for Alternative Finance* .

STOLL, C., L. KLAASSEN, & U. GALLERSDÖRFER (2019): "The carbon footprint of Bitcoin." *Joule* **3(7)**: pp. 1647–1661.

VRANKEN, H. (2017): "Sustainability of Bitcoin and blockchains." *Current opinion in environmental sustainability* **28**: pp. 1–9.

WHEATLEY, S., D. SORNETTE, T. HUBER, M. REPPEN, & R. N. GANTNER (2019): "Are Bitcoin bubbles predictable? Combining a generalized Metcalfe's Law and the Log-Periodic Power Law Singularity model." *Royal Society Open Science* **6(6)**: p. 180538.

WOOLDRIDGE, J. M. (2015): *Introductory econometrics: A modern approach.* Cengage learning.

ZADE, M. & MYKLEBOST (2018): "Bitcoin and Ethereum Mining Hardware." *Mendeley Data* **V1**.

ZADE, M., J. MYKLEBOST, P. TZSCHEUTSCHLER, & U. WAGNER (2019): "Is Bitcoin the Only Problem? A Scenario Model for the Power Demand of Blockchains." *Frontiers in Energy Research* **7**: p. 21.

ZELLNER, A. & H. THEIL (1962): "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations." *Econometrica* **30(1)**: pp. 54–78.

# Appendix A

# Figures

Figure A.1: Price of Bitcoin

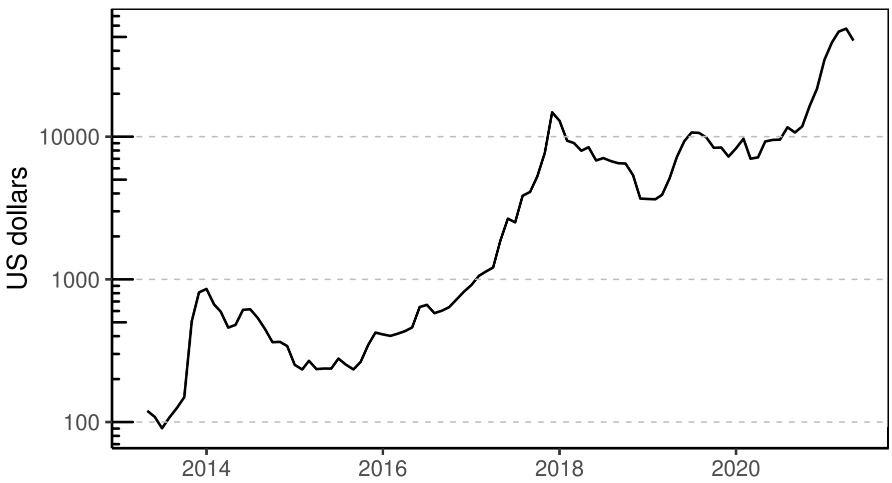


Figure A.2: Hashrate of the Bitcoin network

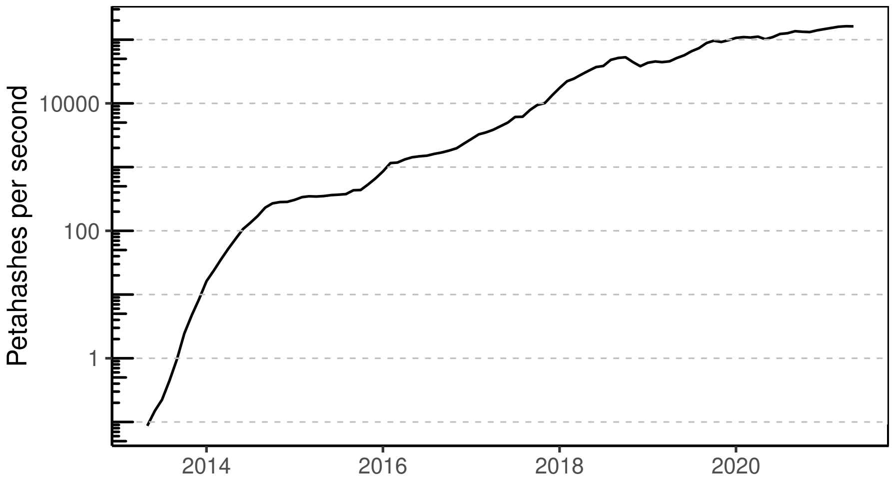



Figure A.3: Number of unique active addresses per 24 hours

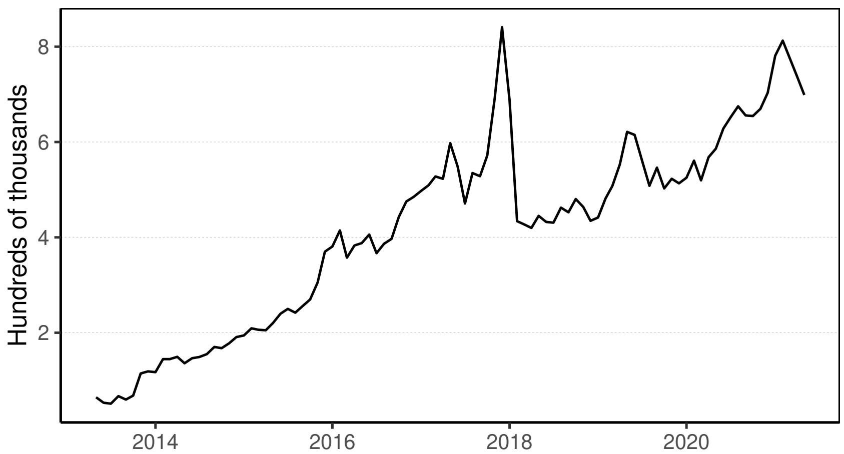


Figure A.4: Search volume from Google Trends

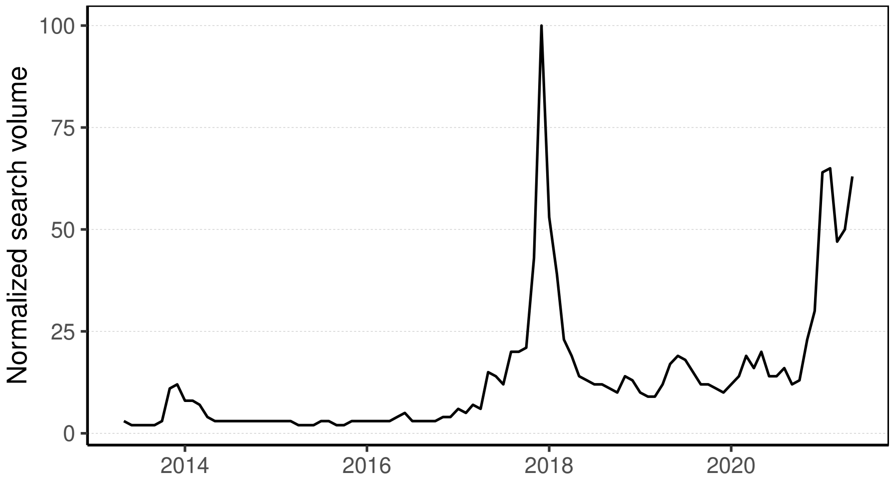



Figure A.5: Transaction fees paid in the Bitcoin network

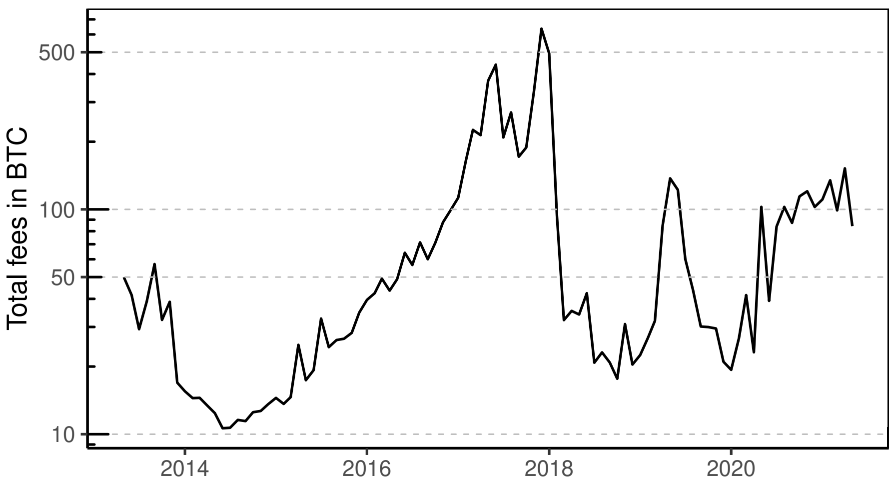


Figure A.6: Total Bitcoin supply

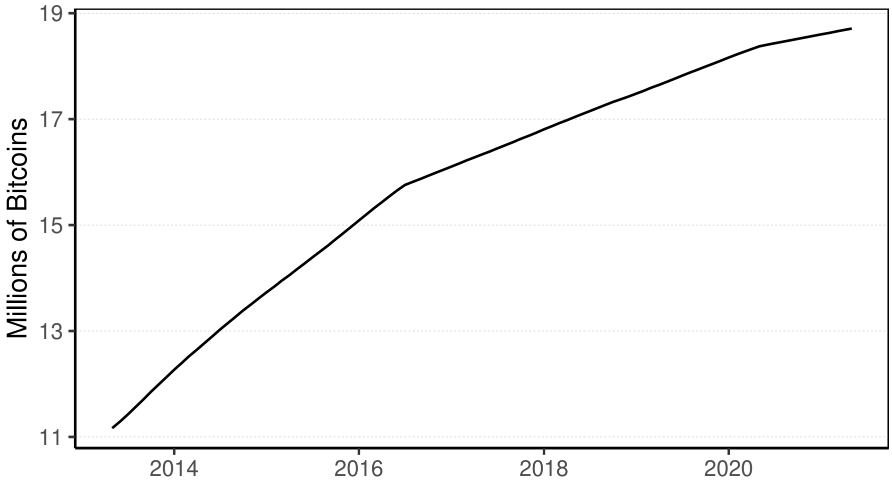



Figure A.7: US dollar M2 money stock