UNIVERSITY OF VAASA
FACULTY OF BUSINESS STUDIES
DEPARTMENT OF ACCOUNTING AND FINANCE

Henrik Wist

ASSESSING THE CONFORMITY OF CRYPTOCURRENCY MARKET DATA WITH BENFORD'S LAW

Master´s thesis in

Finance

Master's Degree Programme in Finance

VAASA 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

UNIVERSITY OF VAASA
Department of Accounting and Finance
Author: Henrik Wist
Title: Assessing the conformity of cryptocurrency market data with Benford's Law
Supervisor: Jussi Nikkinen
Degree: Master of Science in Economics
Programme: Master's Degree Programme in Finance
Programme Started: 2017
Year of Graduation: 2019 Pages: 78

ABSTRACT

The cryptocurrency markets have experienced a drastic increase in trading volume and prices since the global realization of blockchain technologies' potential in the mid-2017. The market capitalization of the cryptocurrency markets increased from only $18 billion on January 1st, 2017 to $565 billion on December 31st, 2017. Thus, it is important to understand the market's current operating state.

This research aims to test whether price and volume data from the cryptocurrency markets conform with Benford's Law and is the first research conducted in this manner. Similar research performed with data from the international financial markets show that price and volume data conform with the Benford's Law distribution, which is also the expected result of this study. The data is gathered from coinmarketcap.com and includes 30 of the top 100 largest cryptocurrencies determined by market capitalization.

Benford's Law offers distributions for first, second and first-two occurring digits of numbers in naturally formed data sets. If the distribution deviates from Benford's Law, it can be caused by an unnatural data formation process, low quality of data or insufficient data. This study also offers theoretical aspects affecting the cryptocurrency markets and its data formation. The authenticity of the use of Benford's Law as a data analyzation tool is emphasized by the fact that it is frequently used in research and as a fraud detection tool in auditing worldwide.

Three tests are conducted for both entire price- and volume data sets. As a further assessment, each subsample included is tested individually. The results show that when testing entire data sets, they closely conform with Benford's Law. Interestingly, when testing the subsamples, only two subsamples of the volume data set conform with Benford's Law at some level. Each price data subsample fails at least one of the three tests. Furthermore, the results show that the volume data conforms far more closely with Benford's Law than the price data. The results indicate that the majority of the subsamples experience something else than a natural data formation process. This gives direction for further research and assessments.

Key Words: Cryptocurrency, Benford's Law, data quality, data formation.

1. INTRODUCTION

1.1. Background of the study

The cryptocurrency markets have experienced a drastic increase in trading volume and prices since the global realization of blockchain technologies' potential in the mid-2017. Some expect cryptocurrencies to potentially have a disrupting effect on the financial system (Li & Wang, 2016). The market capitalization of cryptocurrencies was $565 billion on December 31st, 2017, while it was only $18 billion on January 1st, 2017 (CoinMarketCap, 2018a). The prices have come down since but are still significantly higher than in early 2017.

The growth of this scale also attracts market participants with financial motivations to try and benefit from enthusiastic new investors. According to a research conducted by Gandal, Hamrick, Moore and Oberman (2017), the first time Bitcoin rose to be worth $1,000 per coin, the rise was most likely driven by market manipulation. The price remained at that level only temporarily and it took more than three years for Bitcoin to reach the same level again. As cryptocurrencies have attracted more mainstream investors and countries are preparing to legalize payments by cryptocurrencies, it becomes more relevant to understand whether the cryptocurrency markets are performing on an expected level. Trading in inefficient cryptocurrency markets can be compared to trading in unregulated financial markets. (Gandal et al., 2017.)

This study tests whether price and volume data generated by the cryptocurrency markets conform with Benford's Law (BL). BL is a logarithmic distribution formed by the first, second and first-two digits of values in a data set, from 1 to 9, 0 to 9 and from 10 to 99 respectively. The frequency of the digits decreases as the values increase, resulting in the lowest frequency with the highest values. BL is applicable to naturally formed numerical data sets throughout many fields. (Benford, 1938.) Although, the phenomenon is called the Benford's Law, a mathematician and astronomer by the name of Simon Newcomb was the first one to notice that the first digits of naturally formed numbers do not appear with equal frequencies when examining large data sets (Newcomb, 1881).

BL is commonly used in assessing the quality and credibility of large numerical data sets. For example, in many countries it is used as a tool for identifying tax fraud (Nigrini & Wells, 2012). It can also be used in analyzing data from the financial markets (Ausloos, Castellano & Cerqueti, 2016, Riccioni & Cerqueti, 2018, Ley, 1996, Carrera, 2015). It is important to note that even though BL might indicate incredibility and low quality of the data, it can only be used as an indication that the subject should be more closely looked into. The understanding of data formation mechanisms is also valuable when analyzing price and volume data with BL, as psychologically influenced behaviors of investors can affect the process.

1.2. Purpose and hypothesis of the study

The Purpose of this study is to determine whether the distribution of the first, second and the combination of the two first digits of price and volume data from the cryptocurrency markets follow the distribution of naturally formed numbers in large data sets as analysed by Benford (1938). To the authors best knowledge, the cryptocurrency markets have previously not been studied in this manner. However, Riccioni and Cerqueti (2018) successfully applied BL when evaluating price and volume data obtained from the international financial markets and their study showed that the international financial markets conform with BL.

The cryptocurrency markets are similar to the international financial markets in many aspects. The main differences lie in the market's downtime and lack of regulatory framework. Unlike the international financial markets, the cryptocurrency markets experience no downtime, and thus, are open 24 hours a day for 7 days a week. This aspect also satisfies one of the requirements for BL as BL is applicable only for large data sets. As the markets for cryptocurrencies are relatively new, the fact that the markets are open for 7 days a week increases the amount of trading day data available. The timespan of collected data starts from April 28th, 2013 and ends in November 2nd, 2018, resulting in a 5-and-a-half-year research period for the oldest currencies. The data is collected from www.coinmarketcap.com, which gathers daily information of the cryptocurrency markets from all exchanges and combines them to represent the total daily price, volume, market capitalization and other indicators. The data for the first tradeable digital currency, Bitcoin, is available from April 28th, 2013, hence the given research period. The lack of regulatory framework also leaves the markets vulnerable to manipulative schemes. In this context, BL is a commonly used tool in assessing the credibility of the data.

The research problem can be written in the following way:

Is there a significant deviation from the Benford's Law distribution of the first and second digits of naturally formed numbers when examining a large data set of price and volume data from the cryptocurrency markets?

The hypothesis for the research problem is as follows:

H_0: *"The distribution of the first-, second- and first-two-digit values in the sample of price and volume data are approximately in accordance with Benford's Law."*

1.3. Motivation of the study

The increasing interest towards the emerging cryptocurrency markets generates a requirement for the understanding of the markets current state. The markets are currently unregulated and operate in a decentralized manner. The only regulation existing in the cryptocurrency markets is that of the trading platforms. This kind of regulation is, however, voluntarily provided and not in any way mandatory. Furthermore, some trading platforms choose not to regulate trading or their customer base, thus anyone can enter and trade in the market. Unregulated markets are also vulnerable against schemes which are categorized as manipulative. The research by Gandal et al. (2017) provides evidence of misconduct by a trading platform itself, which further indicates the riskiness of the market and undermines the credibility of the internal regulations issued by the trading platforms.

The motivation of this study generates from these standpoints. By gaining insight about whether the market data from the cryptocurrency markets conform with BL, one can better understand how the market functions. The results indicate how closely the distributions of BL and the cryptocurrency market data conform, which, on the other hand, can be used as an indication of the data formation process. The formation of the data can be affected by both psychological factors of traders and the presence of aspects which decrease market efficiency.

1.4. Structure of the study

This study aims to thoroughly examine the research data and determine whether it is in accordance with BL. Thus, it is necessary to explain some related concepts and theories. Section 2 provides information of some previous research using BL and cases where the markets have been targeted by manipulative schemes. Section 3 serves as an introduction to the cryptocurrency markets. Section 4 explains the theories affecting cryptocurrency markets and the price and volume formation process. The research methodology and data are presented in further detail in Section 5. In Section 6, the obtained results are shown and compared with expectations. Section 7 concludes the research and gives insight for further work.

2. PREVIOUS RESEARCH

This section introduces previous research related to Benford's Law and situations where the financial and cryptocurrency markets have been manipulated. It emphasizes the applicability of BL throughout many fields of research. Market manipulation can not be excluded as an explanation if the price and volume data does not conform with BL. Hence the relevance to the subject.

2.1. The cryptocurrency markets and manipulation

Even though market manipulation has become an important issue in the cryptocurrency markets, previous research of the subject is scarce. The research conducted by Gandal et al. (2017) is the first one published on the subject. Their research uses very detailed data to identify manipulative activity in late 2013 on the at-the-time largest cryptocurrency exchange Mt. Gox. The data used for the research was originally leaked by Mt. Gox itself, and contained trade information including transaction IDs, amounts, exact trade times, currencies and user's state and country codes. The user IDs enabled the linking of the transactions to the actors. The trading volumes and USD to BTC exchange rates were also compared to those reported by similar exchanges and the values obtained from web sites that gathered cumulative trading information from the markets.

The leaked data revealed two suspicious traders: the "Markus bot" and the "Willy bot". Markus' trading account was credited 335,898 Bitcoins ($76 million) through duplicated transactions during the period from February 14th, 2013 to September 27th, 2013 without any compensation paid to any other user. Willy, whose account was activated only several hours after Markus became permanently inactive, conducted its trades through 49 separate accounts, each of which bought $2,5 million worth of Bitcoin in sequential order and never sold the acquired coins. (Gandal et al, 2017.)

The difference between Willy and Markus was that Willy's trades happened with real users whose accounts were credited with fiat currency, even though Willy most likely dishonestly acquired the fiat currency. Willy acquired 268,132 Bitcoins ($112 million) within 65 days, thus being the more aggressive bot of the two. Afterwards, Mark Karpeles, the CEO of Mt. Gox, admitted that Willy was operated automatically by the platform itself. (Gandal et al, 2017.)

Overall, Gandal et al. (2017) found out that prices increased on 80 percent of the days on which suspicious activity occurred, though prices also increased on 50 percent of the days without suspicious activity. While the daily returns where slightly negative during days without suspicious activity, the returns ranged from 1.9 - 2.9 percent and 4.8 – 5.0 percent on all the largest exchanges combined when Markus and Willy where active, respectively. This suggests that the bot's activities had a significant effect on prices. Also, one key finding was that the trading volume increased significantly on days with suspicious activity. Even though Markus' daily trades accounted only for around 21 percent and Willy's for 18 percent of Mt. Gox's trading volumes and 12 and 6 percent of the global volumes, respectively, the increase in volume on suspicious trading days was significantly higher than what can be explained by the bots.

2.2. The financial markets and manipulation

Ausloos et al. (2016) used BL in assessing the credibility and quality of credit default swap (CDS) data from ten- and five-year periods. CDS data was chosen for analyzation for three reasons: first, before the financial crisis CDS were thought to enhance the risk transferring and signaling aspects of the financial markets, second, the CDS markets have grown significantly during the last decade, and third, some large banks have been accused of manipulating CDS prices. The countries included in the study were divided into three categories: core economies (France, Germany and United Kingdom), the most worrying economies (Ireland, Italy, Portugal and Spain), Eastern economies (Croatia, Czech Republic, Poland, Romania and Turkey) in addition with Greece as its' own group. The data set containing observations from ten years was found to be widely compliant with BL. However, BL was found to have been systematically violated in cases including five-year data sets. The most suspected manipulation was found in the five-year data sets of France, Czech Republic, Portugal, United Kingdom, Croatia, Spain and Ireland.

Aggarwal and Wu (2006) researched financial market manipulation in the U.S. by examining cases initiated by the Securities and Exchange Commission (SEC) against OTC market manipulators. They noted that the presence of information seeking traders enhanced the manipulators' opportunities to profit through the increase of competition for shares. Also, market manipulators were likely to be insiders, brokers, underwriters, large shareholders or market makers. The manipulators' goal was to create artificial demand for the targeted stock, luring others to buy based on the increase in volume and share price and enabling the manipulator to profit from selling the stock for the increased price. Also, one way of artificially affecting the demand of securities was spreading false information regarding the security in Internet chat rooms. As a result, Aggarwal and Wu (2006) concluded that during the manipulation period, stock prices, volatility and volume tended to be higher, but rapidly stabilize as the period ended. As for stock characteristics, the stocks targeted by manipulators were usually small and illiquid.

Azad, Azmat, Edirisuriya and Fang (2014) studied trade-induced manipulation in the South-Asian markets: Bombay Stock Exchange (BSE), Dhaka Stock Exchange (DSE) and Karachi Stock Exchange (KSE). They analyzed the price-volume relationship during time periods containing legally prosecuted manipulative schemes, where the manipulations had been of such extent that the whole price index had shifted. They found that the South-Asian markets were inefficient due to the presence of trading-induced manipulation. The manipulations were conducted by a "pump and dump" scheme or, in other words, excessive buying and selling, where the price was artificially inflated before crashing down. The scheme could also be started with excessive selling of the stock, which decreased the price. This enabled the manipulator to buy the stock at a discount, before the stock reaches its actual value again. Like Aggarwal and Wu (2006), also Azad et al. (2014) found that the trading volume tends to be significantly higher during the time of market manipulation than before and after the manipulation period.

2.3. Applications of Benford's Law

Riccioni and Cerqueti (2018) studied the validity and application of BL for daily series of price and volume data of various countries' stock indexes generated by the international financial markets. The timespan of the collected data ended on November 14th, 2014 but varied due to not all assets having been listed on a stock exchange on the same day. The individual indexes included in their analysis were restricted to those with more than 100 observations, to meet the requirements of BL. After the necessary restrictions, the data set included 4,166 stocks overall, resulting in 8,332 time series when investigating volume and price. They found that BL holds in most situations. There were some deviations, but, according to the authors, these deviations were probably to be explained by the technical characteristics and breakthrough events occurring in the financial markets.

Ley (1996) also studied the financial markets using BL. He tested whether the daily returns of Dow-Jones Industrial Average Index (DJIA) and Standard and Poor's Index (S&P) follow the BL distribution. He concluded that no significant deviations from the distribution were observed. When the results were tested for goodness of fit, the indexes would have ranked third and fourth among the original samples tested by Benford.

Examining exchange rates gives important information of the economic conditions and trading activities. They are also linked to short-term interest rates and reserve policies on an international level. Carrera (2015) studied the level of exchange rates management activities in Latin America by using BL. Exchange rates in Latin America suite the research well, as the United States Dollar is more frequently used in transactions than local currencies, increasing the reserve management activities. As a result of the study, the exchange rates of Latin American currencies and the USD deviated significantly from BL, indicating currency rate management of various degrees. On the other hand, exchange rates including the Euro did not deviate from BL. It was suggested that this was a result of the USD being the primary currency for currency rate management in Latin America.

Nigrini and Wells (2012) tested whether BL holds when examining accounting data. Their dataset included over 100,000 accounting numbers over a two-year timespan. After testing the frequencies of the digits, they concluded that accounting data closely followed BL. Thus, a deviation from BL should indicate further inspection for potential fraud. As an example, they mentioned Enron, which was a large American energy company that was found to have manipulated its accounting data for years before filing bankruptcy. The scandal was of such magnitude that it caused the global alteration of accounting standards.

When examining whether Enron's accounting data was in accordance with BL, they noticed that the distribution of *0*s as the second digit was as high as 7 out of 12 (Nigrini & Wells, 2012: 214). The true distribution of *0*s as the second digit in BL is approximately 1 out of 10.

The deviation of accounting data from the BL distribution can also be caused by earnings management, an unfraudulent accounting tactic. Earnings management is perceived as one of management's main objectives, thus being fairly common. Furthermore, it is expected that the management achieves a certain goal. Achieving this goal causes abnormal returns compared to just falling short. As a result, proceeds tend to become enhanced to an expected level to satisfy the user of the information. This phenomenon was first tested on companies listed on the New Zealand Stock Exchange, with data starting from January 1981 and ending in December 1985. Companies were required to have at least one profit announcement published to be included in the research. When examining the distribution of the numbers, the focus concentrates mainly on the second digit of appearing income values. The results showed that there are far more *0*s and far less *9*s as the second digit than what would have been required to match the expected distribution. In addition, the combination of the digit 1 followed by 0 was abnormally common. This supported the claim that management tended to round up the company's income to match expected levels. (Carslaw, 1988.) Later, Thomas (1989) found that the results obtained by Carslaw were similar to the results obtained when examining the phenomenon in U.S. based firms. Also, Keloharju and Niskanen (2010) studied the phenomenon in a Finnish environment. They noted that it was common for Finnish companies to strive for a lower income to avoid taxes, which also was the expected result of their research. However, they found that it was more common for companies to adjust income upward, thus presenting similar results as other studied.

Cho and Gaines applied BL with success while looking at fraud in funding of political campaigns. As data they used the Federal Election Commission's donation filings for six election cycles starting from 1994. The results showed that during the last three election cycles (2000, 2002 and 2004) the frequency of *1*s is decreasing while the frequency of *9*s is increasing. (Cho & Gaines, 2012.) The Iranian minister Boudewijn F. Roukema also utilized BL in a political context when the opposition leader Moussavi suspected electoral fraud in the presidential election of 2009 against President Ahmadinejad's election. The problem was approached through cities' voting volumes by testing the frequency of the first digit of volumes against the BL distribution. As a result, the number 7 appeared too frequently, thus supporting the claims by Moussavi. (Riccioni & Cerqueti, 2018.)

Brähler, Engel, Göttsche and Rauch (2011) were among the first to successfully use BL to assess the validity of macroeconomic data. To identify manipulations, their research compared the distributions formed by all relevant macroeconomic data from 27 EU member states against BL distribution. The data covered a time span of ten years, starting from 1999. The largest deviations for individual euro countries were from data representing Greece, Belgium and Austria, respectively. As for non-euro countries, Romania and Latvia had the largest deviations. The European Commission has confirmed data manipulation by Greece, which indicates the effectiveness of BL also in macroeconomic aspects.

Previous research show that BL has been widely used and is applicable when analyzing a variety of different data sets. The results are also usually interpreted as indications of the data's quality and validity. Thus, it can be expected that BL is applicable also when analyzing data from the cryptocurrency markets, and that it can be used to determine the quality and validity of such data.

3. INTRODUCTION TO CRYPTOCURRENCIES

Cryptocurrencies as a new phenomenon have gained increased attention during the last years. As recently as 2016, the term cryptocurrency was commonly associated with Bitcoin, an electronic currency with almost no economic use at that time. Nowadays, Bitcoin is merely one cryptocurrency in a sea of alternatives. However, academic research regarding cryptocurrencies is still relatively scarce. Thus, this section aims to introduce the cryptocurrency markets and some related concepts.

3.1. History and definition

Cryptocurrency as a term is complex, as it is widely associated with the terms virtual currency and digital currency. These terms can be classified into three different categories. The European Central Bank (ECB) in 2012 published guidelines on how to categorize various forms of electronic currencies. The three-stage categorization starts with plain virtual currencies, which refer to currencies within games or other virtual environments with no connections to the actual world. The second category contains currencies which are similar to the first category’s currencies, but can be purchased with actual currencies, thus having their own exchange rate. The final category adds the possibility to exchange the digital currencies back to actual currencies. The latter category is also known as cryptocurrencies.

Cryptocurrencies are based on a technology called cryptography. One of the founding fathers of cryptography is David Chaum, a doctor of cryptology from the University of California, Berkley (Dostov & Shust, 2014). Chaum developed and published papers on a method of so-called blind signatures in electronic payments, which would be as reliable as banks’ electronic stamps, but anonymous. Chaum’s earliest work is from 1982, which is also when the foundation for cryptocurrencies was developed (Chino & Subramanian, 2015).

David Chaum was also the first person to invent a working cryptocurrency in 1990, which was named E-Cash. Only in 1994, the first payment transaction was successfully completed using E-Cash. Some other cryptocurrencies emerged around the same time as E-Cash, one of which is called Peppercoin. Peppercoin was invented by Silvio Micali and Ronald Rivest as a solution for micropayments. However, cryptocurrencies failed to gain the public's acceptance in the early 90s leaving the crypto-world dormant for some time. (Chino & Subramanian, 2015.)

The crypto-economy, as we know it, started to evolve in 1998, when a paper visualizing about a new system was published by Wei Dai. Like blind signatures introduced by Chaum, also this method was grounded upon anonymity. The goal was to establish a currency system which enables its users to interact with each other efficiently through the internet while being completely anonymous and untraceable. This system would thus also be beyond the reach of governmental control. Inspired by Wei Dai's article, in 2009, a person under the pseudonym Satoshi Nakamoto invented the first decentralized cryptocurrency, Bitcoin. (Plassaras, 2013.)

Bitcoin is the first cryptocurrency which is totally disconnected from governmental control or authority, which can partially explain its rapid expansion. First, the electronic code and design was perfected by invested developers. Next followed the early adoption stage, where CEOs of various start-ups promoted products with Bitcoin as a payment option. Last came the public's financial and ideological attention along with critical thinking from both invested and uninvested authors. The fact which most fascinated or frightened the audience was that, compared to the global financial system where commercial - and central banks and non-profit associations must guarantee trust in the financial system for it to function, Bitcoin was built in a way which excludes all intermediaries. (Zimmer, 2017.)

The trust based on Bitcoin stems from its functionalities. One could expect an electronic currency to be copied and passed around as any other electronic document of file, leading to users trying to benefit from the same "coin" repeatedly. Bitcoin is built on a blockchain network, which publicly records all transactions to the same information "chain", eliminating the possibility of double transactions. The transactions, on the other hand, are added to the chain by users participating in the process of using their computer hardware to build and record transactions made by others. For their work, the participants are rewarded in Bitcoin. This process is also called "mining". The Bitcoin mining process becomes harder along with every new Bitcoin, thus eliminating the possibility of inflating the currency. (Nakamoto, 2009.) This “proof of concept” process is a built-in decentralized verification system, which obviates centralized authorities (Zimmer, 2017).

The widespread attention that Bitcoin rapidly gained after its launch, caused other cryptocurrencies to emerge mimicking the proof of concept process and ideology. In 2018, the total number of different tradeable cryptocurrencies reached over 2,000. Also, the internet platforms used for trading cryptocurrencies have risen to over 200 from just a few in early 2013. The cryptocurrency market reached its historical peak in early 2018, with the total market capitalization of over 800 billion. (CoinMarkerCap, 2018b.)

3.2. Cryptocurrencies compared to fiat-currencies

The term fiat comes from Latin and refers to a currency which is worthless by itself, but gains value through governmental laws and regulations. The value of fiat-currencies is also not tied to any commodity, like gold, but is determined by supply and demand. (Goldberg, 2012.) The most traded fiat-currencies in the world are the United States Dollar and the Euro. Unlike fiat-currencies, cryptocurrencies cannot be printed endlessly due to the deceleration of the mining process. This means that the mining process slows down in relation to the circulating supply. In addition, most cryptocurrencies have a so called “hard cap”, which determines the maximum supply. However, the value of cryptocurrencies is also based on the balance of supply and demand.

Fiat-currencies are produced and implemented by local governments and can only be used within a pre-determined geographic region. Thus, fiat-currencies need to be exchanged when moving between different regions. Cryptocurrencies, on the other hand, can be transferred and used between different regions without the need of exchanging. Due to their digital form, the utilization of cryptocurrencies is also more cost-efficient compared to fiat-currencies which are associated with relatively high transaction-, storing- and manufacturing costs. (Plassaras, 2013.)

The definition of money covers three main characteristics: can be accepted as a payment, can be used to store and measure value. Until this point, cryptocurrencies can only fulfill the first criteria: can be accepted as a payment. However, the acceptance of cryptocurrencies as means of payment is limited due to their low public acceptability. The value of cryptocurrencies has been relatively volatile through their existence, thus failing to store or measure value. Cryptocurrencies can rather be seen as speculative investment subjects. (Bjerg, 2016.) The daily volatility and feasibility of the Bitcoin as part of an investment portfolio was studied by Pandey and Wu (2014). The results were in line with the ones found by Bjerg (2016): when comparing to major currencies, stocks, real estate, indices and gold, Bitcoin had close to no correlation to daily returns.

The acceptance of cryptocurrencies as means of payment rely mainly on widespread acceptance and the public's interest towards a decentralized payment system. So far, mainly people associated with cryptocurrencies and with enough knowledge about them are interested in advancing cryptocurrencies towards becoming a new payment system. Most commonly they are seen as speculative investment alternatives. (Glaser, Haferkorn, Siering, Weber & Zimmermann, 2014.)

3.3. Cryptocurrency exchanges

Cryptocurrencies can be traded on online trading platforms similar to trading platforms for stocks and other financial instruments. Each trading platform has a specific list of cryptocurrencies that can be traded on the exchange. This mean that the exchange platform has access to the specific cryptocurrency's network, enabling it to create so-called wallets for storing and transferring cryptocurrencies between owners. Some exchanges verify user information in order to enhance security on the platform, but apart from that, anyone can create a trading account. When the account has been created, the user chooses the cryptocurrencies to be traded and the platform generates a wallet for them. Thereafter, the user either buys cryptocurrencies using fiat currencies, if the platform supports cryptocurrency-fiat currency trading, or transfers cryptocurrencies from another source to the platform. When the account has a balance, the cryptocurrencies can be traded in a similar manner as stocks on stock exchanges. According to CoinMarketCap, on December 2018, the largest cryptocurrency exchange platform ranked by trading volume occurring in the last 30 days was Binance (CoinMarketCap, 2018b).

The cryptocurrency exchanges are irregulated third party-owned service providers, which are vulnerable to security breaches and attacks. Trading platforms themselves are also provenly able to participate in schemes decreasing market efficiency. For example, Mt. Gox, one of the largest cryptocurrency exchanges in its time, admitted taking part in manipulating the price of Bitcoin in 2013. (Gandal et al., 2017.) The amount of trading platforms has increased significantly since the year 2013, reaching a total of 227 in late 2018 (CoinMarketCap, 2018b).

3.4. A legitimacy perspective

Due to the decentralized nature of cryptocurrencies, legislators are sceptic of accepting cryptocurrencies as means of payment or as part of an investment portfolio. As cryptocurrency transfers cannot be connected to any person, payments in cryptocurrencies are commonly preferred in criminal activities (Brown, 2016). For example, cyberhackers have been demanding payments in Bitcoin in exchange for not revealing sensitive information to the public. The problem is not deemed to be in cryptocurrencies themselves, but in the lack of regulation and anti-money laundering principles. (Brown, 2016.)

An important step towards creating a regulatory framework is defining the concepts and key definitions (Abend, 2008). Each country has its own definition and perspective towards cryptocurrencies and thus no universal regulatory framework can yet be formed. The European Central Bank’s (2015) publication: Virtual Currency Schemes – A Further Analysis mention that according to the laws of the European Monetary Union, only Euro is an acceptable currency. Thus, cryptocurrencies are not recognized as currencies in the European Monetary Union. In addition, according to the publication, the cryptocurrency markets were not regulated at that time.

In Finland, the tax authorities have presented instructions on the taxation of cryptocurrencies. The instruction follows the framework of the publication by the

European Central Bank (2015) stating that trading and using cryptocurrencies as means of payment, is not regulated, and the terms are solely determined by the contract between the participants as cryptocurrencies are not recognized as currencies. Also, according to the guidelines presented by the Finnish tax administration, the profits from cryptocurrencies, whether from mining or trading, are taxable once the cryptocurrency is exchanged to fiat currencies or other virtual currencies or physical goods. The losses generated by investments in cryptocurrencies are not deductible in one's personal taxation. However, if a company's main source of income involves investments in cryptocurrencies, the decrease in the investments value may be deductible, to some extent, in the company's income taxation. (Verohallinto, 2018.)

To address the problem of unregulated cryptocurrency markets, the Finnish government formulated in 2018 a proposition regarding the supervision of bank accounts. The proposition aims to enforce the law enforcement capabilities to prevent money laundering and funding of terrorism. One part of the proposition requires companies engaged in trading of cryptocurrencies to be registered under the Finnish regulatory authority for the financial markets, Finanssivalvonta. Hence, Finanssivalvonta would be able to monitor trading activities and cash flow movements of the registered parties and identify suspicious activities. In addition, the requirements of knowing your customer would be enforced. (HE 167/2018, 2018.)

The proposition by the Finnish government is an example of authorities working on building regulatory frameworks for the cryptocurrency markets. However, as one of the key attributes of the cryptocurrency markets is decentralization, it might be challenging to regulate the markets on a country-by-county level. As possessing a cryptocurrency wallet requires no disclosure of personal information, one might participate in trading cryptocurrencies through trading platforms that are registered in countries without regulation while staying completely anonymous. Also, transferring cryptocurrencies between wallets require no trading platform if the sender knows the wallet number of the receiver. Overall, country level regulation gives investors the opportunity to show their trading activities but does not exclude the possibility of total anonymity.

4. THEORETICAL APPROACH

This section aims to present a theoretical framework for the cryptocurrency market and its functionalities. The cryptocurrency market is young and without regulation, thus, reflecting similarities to emerging financial markets. The theory surrounding efficient markets explain the role of information in the markets, which is also relevant in a speculative environment such as the cryptocurrency markets. Market manipulation and different asset pricing phenomenon can affect the effectiveness of the market. This also connects the theories to this research.

4.1. Emerging financial markets

Emerging markets are commonly associated with higher risk due to low integration and regulation of the markets. Previously, for example, stock markets were widely institutional and legitimate and located in wealthy countries. This wealth has flowed to emerging markets in forms of foreign aid and national lending. In 1970's, banks in emerging markets initiated long term lending strategies supported by the foreign cash inflow. In Mexico, the lending strategy failed in the early 1980's causing a stop in private foreign investments. This reflects the instability of unregulated financial markets. Also, higher expected returns are straight in line with higher risk factors. As stated in the World Bank's World Development report, the most secure path to rapid economic growth is sufficient regulation in the financial markets. This ensures a state of stability which attracts more sophisticated investors. (Davis, Lounsbury & Weber, 2009.)

Financial markets that have a weak law-based regulation, provide individual investors with weak legal protection. The risks associated with investments increase significantly if the rights of the investors are not protected. For the markets to operate efficiently and contribute in developing the economy, a regulatory framework is necessary. (La Porta, Lopez-de-Silanes, Shleifer & Vishny, 1998.)

Glaeser, Johnson and Shleifer (2001) point out that for a law-based regulatory framework to function properly, significant investments are required. In many countries, courts lack financing, motivation, are unclear of the interpretation of the law, are unfamiliar with the economic issues or are corrupted. In such environments, the verification of contractual facts cannot be conducted in a trusted manner. The verification of facts needs to be driven by powerful incentives. Thus, it can be argued that the regulatory framework can be more efficiently developed by regulators than judges. Regulators can be more easily motivated as they are more involved in developing the markets, hence they can potentially be more aggressive in terms of regulation enforcement. In an optimal case, both regulators and judges work together towards a best solution.

Morck, Yeung and Yu (2000) showed that stock prices in emerging markets tended to co-move in a synchronized manner, while the prices in mature markets were relatively unsynchronized. Three explanations for this phenomenon can be considered. First, firms in low GDP per capita economies might be more correlated with fundamentals, resulting in more synchronized price movements. Secondly, the regulatory framework in emerging markets tend to be poor, enabling political events and rumors to cause market-wide swings in stock prices. Lastly, low regulatory framework can deflate the value of firm-specific information, resulting in stock prices that do not reflect all information. This reduces the variation between stock prices and increases return synchronicity.

The lack of regulation in the markets also enable the possibility of manipulative schemes. For example, price manipulation induced by artificial trading between two co-operating brokers is likely to exist throughout the emerging financial stock markets. For young markets such manipulations are a problem that needs to be overcome. Mature markets are more protected from manipulation through law-based regulation. (Khwaja & Mian, 2005.)

The emerging cryptocurrency markets relate to emerging financial markets in many ways. The regulatory framework discussed is necessary for the markets to function properly. However, as the cryptocurrency markets are decentralized, also a decentralized regulatory system would be required for the regulation to be effective. When observing the price movements in the cryptocurrency markets, one might notice that different cryptocurrencies' prices tend to co-move as if there were some common factors affecting the price. As shown by Morck et al. (2000), this is common for emerging markets.

4.2. Market efficiency

The purpose of the capital market is to efficiently allocate capital to where it is most needed. To be able to do this, the signals given by the market should be reliable. The ideal situation is when investors can make investment decisions and can choose from a variety of securities based on prices which fully reflect the information available to the market. Thus, each market participant should always have access to the same information. Fama (1970) studied the theory of efficient markets through three models: the weak form, the semi-strong form and the strong form. The weak form refers to a market situation where only historical price data is available. The semi-strong form took into account all reports and news published, such as stock splits and annual reports. The strong form considered situations where some investors have unique information on securities which information is not publicly available. As a result, Fama (1970) found that the financial markets are efficient in case of the weak and the semi-strong forms. The strong form has best use as a benchmarking tool for deviations.

On the other hand, searching for market information is a time-consuming and costly process. The costs related to searching information regarding specific securities might not be in line with the potential gains. Hence, it is arguable whether the markets are, in fact, efficient. Gârleanu and Pedersen (2018) argue that for the markets to be efficient, the securities' prices should also reflect the price of asset management. If asset managers were easier to find, more investors would actively use asset management, their fees would be lower and, as a result, the overall level of knowledge would increase among the investors. In practice, asset management is used by investors with relatively large amounts of capital, while smaller investors lack the ability to utilize the information sources of asset management companies. Investors investing in asset management tend to outperform the market.

In the case of efficient markets, informed investors would actively use every piece of market information available, and, when doing so, eliminating the profiting possibilities as the prices would instantly reflect all available information. As many investors try to benefit from the same information at the same time, the price movements are random and unpredictable. As new investment strategies are invented, old strategies may become less profitable and, in the end, obsolete. This idea has led to a theory of the financial markets being like a biological subject that evolves. While evolving, some old subjects, such as securities and investment strategies, get extinct, and more profitable alternatives take their place. In this sense, investing in asset management and active information seeking might be profitable until the information becomes outdated. (Farmer and Lo, 1999.)

From a behavioral perspective, investors also tend to overreact on positive or negative information. The effect was studied by De Bondt and Thaler (1987) who showed that a fall in securities' prices tend to be followed by a sequential rise in prices, or the other way around. This happens to such extent that the securities' prices do not reflect their fundamental values after the initial movement. The difference indicates the level of overreaction. The efficient market hypothesis introduced by Fama (1970) is, however, in contradiction with these findings. Zarowin (1990), on the other hand, argue that the returns experienced after losses are not due to overreaction, but are due to size factors when examining periods over three months. In this size phenomenon, if the losers are smaller than the winners, the losers tend to outperform the winners, and if the winners are smaller than the losers, the winners tend to outperform the losers.

The utilization of market information is a crucial part of the price formation process in the cryptocurrency markets. The difference between the financial markets and the cryptocurrency markets is that the cryptocurrencies' prices are mainly based on expectations. These expectations are mainly based on rumors as the teams behind cryptocurrencies have usually only presented an idea on how the currency would be utilized. Thus, identifying relevant and reliable information and determining its effect on prices is a time demanding process. Also, manipulative schemes reflect false information to the markets. As the cryptocurrency markets are unregulated and relatively young, the presence of market manipulation is possible. Hence, it is important to understand the theories surrounding market efficiency to understand the functionalities of the cryptocurrency markets.

4.3. Market manipulation

Market manipulation is a negative aspect related to poorly regulated markets. Manipulated markets are inefficient in terms of information transparency and asset allocation, resulting in the exploiting of uninformed investors. One way of manipulating markets is the so-called “pump and dump” scheme, where co-operating brokers trade amongst themselves to artificially inflate the price while signalling a growth trend to the market. This is followed by uninformed investors entering the market in hope of a growing stock price, which increases the stock price even further. Once the price is high enough, the brokers exit, leaving the uninformed investors with an overvalued stock. This scheme is mainly initiated on days when a certain stock price is low. Vice versa, the scheme can be conducted also as “dump and pump”, where co-operating brokers sell a specific stock to each other, signalling a decreasing price trend to the market. This causes uninformed investors to sell the stock, decreasing the price even further. As the price is low enough, the brokers buy the stock from the market with a discounted price. Dump and pump schemes are likely to take place on days with relatively high stock prices and require an initial ownership of the stock. Brokers also have an advantage in engaging in manipulative schemes, as they tend to have lower transaction costs, better real-time information and are able to spread false information regarding stocks. (Khwaja & Mian, 2005.)

One form of market manipulation is the distribution of false market rumours. As stated by Buckner (1965), a feature of a rumour is that information is without confirmation when passed forward between persons. It is also argued that when hearing a rumour, the recipient tends to be unsure whether the information is valid or not and whether to use time in evaluating the authenticity or to react instantly (Banerjee, 1993). For example, Bertin, Torabzadeh and Zivney (1996) provided evidence of positive abnormal returns surrounding takeover rumours. When a company is taken over using a stock buyout, a premium is normally paid upon the initial stock price. When investors hear rumours regarding a takeover, they tend to buy the stock until the price matches the initial price plus the premium added. If the rumour is proven to be false, the stock price plummets back to its initial level.

Spreading false rumours deliberately and profiting from the artificial price movement is an act of market manipulation. Brokers have an information and network advantage in rumour spreading (Khwaja & Mian, 2005). In addition, the increasing use of the internet and other communication networks make rumour spreading easier and more effective.

In some cases, brokers have deliberately traded their customer's portfolio back and forth to increase their own commissions. If caught, the brokers are obligated to pay the commissions and possible losses back to the customer. This kind of activity, which is also known as churning, increases the volume of the traded stocks and makes them seem more active to uninformed investors. (McCann, 1999.) Based on the increased trading activity, other investors are given false signals about new information in the market and can lead to unprofitable trading decisions. Simply trading stocks back and forth to increase trading activity and prices, but in one's own account, is called "wash trading". Wash trading and churning are two of the most common ways of volume-based market manipulation (Cumming, Johan & Li, 2011). Stock returns are also shown to be higher on days when the trading volume is higher than usual (Garvais, Kaniel & Mingelgrin, 2001).

Having the control over a large amount of capital enables the possibility to corner the market. This means that one can purchase the majority of the available stock supply and significantly affect the price of the stock. As shown by Allen, Litov and Mei (2006), cornering can also be a result of rational investor behaviour. However, deliberately cornering the market significantly increases the volatility of prices, negatively effects market efficiency and causes severe distortions in prices.

As market manipulation is more common in emerging and unregulated markets, it is important to take into account its possibility to exist also in the cryptocurrency markets. Market manipulation also affects the price and volume formation process, rendering it unnatural. When using Benford's Law as a method of data analyzation, the unnatural data formation process may be captured as the data no longer conforms with Benford's Law.

4.4. Pricing theories

As market manipulation, also different pricing phenomenon can affect the data formation process and the conformity with Benford's Law. As the cryptocurrencies are traded in a similar manner as securities in the financial markets, similar phenomenon may be present.

In effective markets, the market price is determined by the available information and expectation probabilities perceived by market participants. The rational expectations theory is thus an important theory surrounding the pricing process. The theory is based on the wisdom of the crowd ideology, which states that the average expectation of the crowd is closer to the right answer than the individual expectations. Thus, it is assumed that the average expectations made by market participants are correct while using all information available to the market (Muth, 1961).

Various pricing theories can be used to explain unnatural price and volume formation processes. One example of an irrational pricing process is clustering. Clustering involves investors to prefer round numbers rather than halves, halves rather than quarters, quarters rather than other fractions. For example, when looking at real estate prices, far more real estates are offered at round numbers than odd numbers. Harris (1991) argue that clustering exists, as it reduces negotiation costs. Negotiations require more interaction between traders if the prices are in odd numbers, as it offers a larger range of possible bids and increases the amount of required explanatory information. Another line of reasoning around clustering takes a behavioural perspective. Ikenberry and Weston (2008) propose that investors willingly choose rounded numbers when trading, as their study finds 50 percent of trades being completed in only 20 percent of the available trade price intervals. As an explanation, they cite studies including one of Niederhoffer and Osborne (1966), which provides a psychological framework for clustering by claiming that people tend to think in round numbers, thus are driven to trade in a similar manner.

Clustering is not only a phenomenon occurring in price formation, but also in trade sizes. This kind of clustering takes place particularly when investors try to disguise their trades by rounding medium-sized trades on days when trading is unusually active. Rounded trades also occur more consistently in large trades. However, rounded trades are less likely to occur in high priced stocks and shortly before the end of quarters. The latter is argued to be due to balancing of portfolios before the end of reporting periods. (Alexander & Peterson, 2007.)

Another irrational phenomenon affecting investor’s investment decisions is overconfidence. Overconfidence exists in situations where more weight is set on one’s own beliefs and expectations than on those of the market (Daniel, Hirshleifer & Subrahmanyam, 1998). This can lead to mispricing and irrational trading activity. If the market values a security below the overconfident investor’s valuation, the investor tends to hold on to the position and even purchase more of the underlying security, and vice versa. In markets with a relatively large number of overconfident participants, the overconfidence effect may affect the market price, thus signalling a new level of the market’s expected price. This kind of a price effect decreases the information value of the market price, which, then again, decreases market efficiency.

5. DATA AND METHODOLOGY

In this section, the methodology and data used to conduct this study are described in more detail. As mentioned earlier, the method used is called Benford's Law. Through describing the various aspects and restrictions for using BL in section 5.1., the data set is derived. The data set and the cryptocurrencies included are described in section 5.2. The last section concentrates on the tests used to verify the significance of the results.

5.1.Benford's Law

Benford's Law is a logarithmic distribution which indicates the occurrence of a specific first and second digit in a large numerical data set. The first digit may vary from one to nine and the second digit may vary from zero to nine. Although the phenomenon is called the Benford's Law, a mathematician and astronomer by the name of Simon Newcomb was the first to notice that the first digits of naturally formed numbers do not appear with equal frequencies when examining large data sets (Newcomb, 1881). Newcomb noted that one using a lot of logarithmic tables might have noticed that the first pages are more worn out than the last pages. A common way of perceiving the distribution of first and second digits occurring in a random numerical data set would be that each digit has an equal chance of occurrence. Newcomb's discovery led him to question whether the digits actually occur with the same frequency or follow a different distribution instead. After comparing anti-logarithmic tables with logarithmic tables, he concluded that the pages of anti-logarithmic tables are used with equal frequencies unlike the logarithmic tables. The frequencies were calculated for the occurrence of first and second digits in a random numerical data set. The calculation for determining frequencies for third occurring digits was deemed unnecessary as the frequencies would closely follow each other.

Over a half century later, a physicist called Frank Benford picked up on the same phenomenon. Benford (1938) gathered a random sample data set which contained 20,229 observations from 20 different categories and tested whether the first and second digits follow a specific logarithmic distribution. The categories varied from random numbers picked from newspapers to the number of rivers in regional areas and volts used in x-rays. The fact that the numbers were picked randomly is emphasized. The numbers chosen also contained three or more digits and were not too conditioned. If the data set contained numbers which started with the digit 0, the next digit larger in value than 0 would be counted as the first digit. Benford calculated the frequency distribution of first occurring digits for each category. The distribution obtained by Benford (1938) was close to identical with the one obtained by Newcomb (1881). The distribution is shown in Table 1.

Digit	First Place	Second Place
0	0.000	0.120
1	0.301	0.114
2	0.176	0.109
3	0.125	0.104
4	0.097	0.100
5	0.079	0.097
6	0.067	0.093
7	0.058	0.090
8	0.051	0.088
9	0.046	0.085

Table 1. Frequencies for first and second place digits (Benford, 1938).

As shown in Table 1, the average chance of occurrence is decreasing as the value of the digit increases. Numbers that have the digit 1 as a first digit get a 30,6 percent chance to occur as the digit 2 has an 18.5 percent chance. For the digit 9, Benford's (1938) calculations gave an occurrence of only 4.7 percent. Benford's original values are shown in Table 2. The value for logarithm 2 is 0.301, which is close to Benford's chance of getting the digit 1 as the first digit. When subtracting logarithm 3 from logarithm 2, the resulting logarithmic interval is close to the chance of getting 2 as the first digit. The same effect applies when calculating log 10 minus log 9, which gives the value of 0.046. Thus, the results follow closely the logarithmic relation:

(1) $$F_a = \log\left(1 + \frac{1}{a}\right)$$

where F_a stands for frequency of the specific digit a as a first digit in a number (Benford, 1938).

	First Digit									
Category	1	2	3	4	5	6	7	8	9	Count
Rivers, Area	31,0	16,4	10,7	11,3	7,2	8,6	5,5	4,2	5,1	335
Population	33,9	20,4	14,2	8,1	7,2	6,2	4,1	3,7	2,2	3259
Constants	41,3	14,4	4,8	8,6	10,6	5,8	1,0	2,9	10,6	104
Newspapers	30,0	18,0	12,0	10,0	8,0	6,0	6,0	5,0	5,0	100
Spec. Heat	24,0	18,4	16,2	14,6	10,6	4,1	3,2	4,8	4,1	1389
Pressure	29,6	18,3	12,8	9,8	8,3	6,4	5,7	4,4	4,7	703
H.P. Lost	30,0	18,4	11,9	10,8	8,1	7,0	5,1	5,1	3,6	690
Mol. Wgt.	26,7	25,2	15,4	10,8	6,7	5,1	4,1	2,8	3,2	1800
Drainage	27,1	23,9	13,8	12,6	8,2	5,0	5,0	2,5	1,9	159
Atomic Wgt.	47,2	18,7	5,5	4,4	6,6	4,4	3,3	4,4	5,5	91
$n^{-1}, \sqrt{n}$...	25,7	20,3	9,7	6,8	6,6	6,8	7,2	8,0	8,9	5000
Design	26,8	14,8	14,3	7,5	8,3	8,4	7,0	7,3	5,6	560
Digest	33,4	18,5	12,4	7,5	7,1	6,5	5,5	4,9	4,2	308
Cost Data	32,4	18,8	10,1	10,1	9,8	5,5	4,7	5,5	3,1	741
X-Ray Volts	27,9	17,5	14,4	9,0	8,1	7,4	5,1	5,8	4,8	707
Am. League	32,7	17,6	12,6	9,8	7,4	6,4	4,9	5,6	3,0	1458
Black Body	31,0	17,3	14,1	8,7	6,6	7,0	5,2	4,7	5,4	1165
Addresses	28,9	19,2	12,6	8,8	8,5	6,4	5,6	5,0	5,0	342
$n^1, n^2 \ldots n!$	25,3	16,0	12,0	10,0	8,5	8,8	6,8	7,1	5,5	900
Death Rate	27,0	18,6	15,7	9,4	6,7	6,5	7,2	4,8	4,1	418
Average	30,6	18,5	12,4	9,4	8,0	6,4	5,1	4,9	4,7	1011
Probable Error	±0,8	±0,4	±0,4	±0,3	±0,2	±0,2	±0,2	±0,2	±0,3	-

Table 2. First digits of Benford's subsamples (Benford, 1938).

When examining the second digit in a number, we must account for ten different values for digits as the digit 0 is included, thus dividing logarithmic intervals in to ten parts. If the first digit is a and the following digit is b, then the two-number combination is ab where the second digit is in relation to the first. Logarithmic intervals between ab and ab+1 can be written as log (ab – 1) – log ab, which leads to the frequency equation of F_b:

(2) $$F_b = \sum_{k=1}^{9} \log\left(1 + \frac{1}{10k+b}\right).$$

The frequencies for first and second appearing digits are also shown in Table 1. (Benford, 1938.)

Following the same method as for the first and second digits, a distribution can also be calculated for the combination of first and second digits ranging from 10 to 99:

(3) $$F_{ab}=\log(1+\frac{1}{ab})$$

where *a* determines the first number in order and *b* the second (Benford, 1938). The distribution of first and second digits combined is shown in Table 3.

	0	1	2	3	4	5	6	7	8	9	**Sum**
1	4.14	3.78	3.48	3.22	3.00	2.80	2.63	2.48	2.35	2.23	**30.10**
2	2.12	2.02	1.93	1.85	1.77	1.70	1.64	1.58	1.52	1.47	**17.61**
3	1.42	1.38	1.34	1.30	1.26	1.22	1.19	1.16	1.13	1.10	**12.49**
4	1.07	1.05	1.02	1.00	0.98	0.95	0.93	0.91	0.90	0.88	**9.69**
5	0.86	0.84	0.83	0.81	0.80	0.78	0.77	0.76	0.74	0.73	**7.92**
6	0.72	0.71	0.69	0.68	0.67	0.66	0.65	0.64	0.63	0.62	**6.69**
7	0.62	0.61	0.60	0.59	0.58	0.58	0.57	0.56	0.55	0.55	**5.80**
8	0.54	0.53	0.53	0.52	0.51	0.51	0.50	0.50	0.49	0.49	**5.12**
9	0.48	0.47	0.47	0.46	0.46	0.45	0.45	0.45	0.44	0.44	**4.58**
Sum	**11.97**	**11.39**	**10.88**	**10.43**	**10.03**	**9.67**	**9.34**	**9.04**	**8.76**	**8.50**	

Table 3. Frequencies for combined first and second digits (Benford, 1938).

As indicated by previous research, the frequency distribution appears to be widely applicable when tested on different sets of numerical data. Then again, some mathematical tables, such as square-root tables, and, for example, telephone numbers of a specific region are not a good fit with BL. Telephone numbers in a specific region tend to start with the same combination of numbers. Even though a set of numbers does not agree with the BL distribution, the distribution will, however, converge into the BL distribution when taking random samples from this non-agreeing distribution and testing the samples for frequencies. (Hill, 1995.) Despite this, some requirements for the use of BL has been introduced. The requirements are mostly in line with those introduced by Benford (1938):

1. The numbers are randomly formed,

2. The numbers are not too conditioned,

3. The numbers contain three or more digits, *4. The number may contain 0 as the first digit, but the next digit larger in value than 0 is counted as the first digit,*

5. The data set is large.

Durtschi, Hillison and Pacini (2004) listed examples for when to expect the data set to be in accordance with BL:

1. Numbers resulting from mathematical combinations,

2. Transaction level data,

3. Large data sets,

4. Mean is greater than the median and skew is positive.

The first example refers to situations where the number is formed by a mathematical combination such as sales divided with price. Transaction level data include numbers as they appear, such as sales data. A large data set tends to be undefined in previous research. However, in large data sets, the observations range from roughly 1,000 (Carslaw, 1988, Niskanen and Keloharju, 2000) to 70,000 (Thomas, 1989, Cho and Gaines, 2007). Wallace (2002) noted that a data set tends to follow BL more closely when it has a positive skew and the value of the mean divided by the median is large. The claim gains proof from the fact that Benford's (1938) distribution is dominated by small numbers.

Durtschi et al. (2004) also introduce situations where the distribution is not in accordance with BL:

1. Numbers are assigned,

2. Numbers are influenced by human thought,

3. Accounts with a large amount of firm-specific numbers,

4. Number sets with a built-in minimum or maximum,

5. Where no transaction is recorded.

Numbers being assigned refer to specific logic of forming numbers, such as numbering invoices or checks. The numbers follow a specific logic and thus are not formed randomly. The same principle applies to numbers that are formed to satisfy psychological thresholds. Numbers from store sales, such as $0,99, are assigned to this category. Accounts with firm specific numbers are formed to follow a specific transaction type and thus the ledger shows the same numbers repeatedly. Also, if the number set has a built in minimum or maximum, it is not fitting with BL. Lastly, where no transaction is recorded means that the distribution for a set of zeros is not in accordance with the restrictions of BL.

Unrelated to BL, but relevant for defining restrictions for the data, are the elements of an optimal target of manipulation. As explained by Aggerwal and Wu (2006), stocks that are:

1. small in market capitalization,

2. illiquid and

3. targeted by information seeking traders,

are usually targeted by manipulators. Smaller market capitalization gives the manipulator the opportunity to significantly affect the price, volume and volatility with a relatively small amount of capital. Illiquidity refers to relatively inefficient markets, such as OTC markets. In such markets, trading and the information disclosure requirements are much less regulated. Information seekers tend to compete for shares, which leads to better profitability opportunities for the manipulator.

The data set used in this research is selected respecting the guidelines described. More specific reasons for selection will be described in the next section, where the data is also introduced.

The significance testing is conducted by using Pearson's chi-squared test, Z-statistic and mean absolute deviation (MAD). The equation for Pearson's chi-squared test can be written as follows:

$$\mathbf{X^2(n\text{-}1)=\sum_{i=1}^{n}\frac{(AC_i\text{-}EC_i)^2}{EC_i}} \tag{4}$$

where AC_i is the actual detected count for digit i and EC_i is the expected count. Degrees of freedom for the test are calculated as n-1 and are 8 for the first digit frequencies, 9 for second digit frequencies and 89 for first-two digit frequencies. Critical values for the Pearson's chi-squared test are 20.09, 21.67 and 122.94 for 8, 9 and 89 degrees of freedom respectively on a 1 percent level and 15.51, 16.92 and 112.02 for the same degrees of freedom on a 5 percent level. If the value obtained by the chi-squared test exceeds the critical value, the null hypothesis is rejected. The chi-squared test is more reliable when testing smaller data samples. (Nigrini & Wells, 2012: 153-154.)

The equation for the Z-statistic is

(5)
$$\mathbf{Z=\frac{|AP-EP|-(\frac{1}{2N})}{\sqrt{\frac{EP(1-EP)}{N}}}}$$

where AP stands for actual proportion, EP for expected proportion and N for the number of records. Critical values for the Z-statistic are 1,96 and 2.33 for 5- and 1-percent significance levels, respectively. If the value of the Z-statistic exceeds the critical value, the null hypothesis is rejected. The general level of significance used in Benford's Law-testing is 5 percent. The Z-statistic is also more reliable when testing smaller data sets, as even small deviations are flagged significant in large data sets. (Nigrini & Wells, 2012: 150-153.)

MAD is calculated as the average distance of each digit frequencies from the expected frequency:

(6)
$$\mathbf{MAD=\frac{\sum_{i=1}^{K}|AP-EP|}{K}}$$

where AP and EP stand for actual and expected proportions respectively and K stands for the number of digits under observation (9 for first-, 10 for second- and 90 for first-two-digit tests). There are no pre-determined critical values for the MAD. (Nigrini & Wells, 2012: 158-160.) Nigrini and Wells (2012) presents a table calculated by using everyday data examples and personal experience to offer critical values for the MAD-measure. The results are presented in Table 4.

Digits	Range	Conclusion
First Digits	0.000 to 0.006	Close conformity
	0.006 to 0.012	Acceptable conformity
	0.012 to 0.015	Marginally acceptable conformity
	Above 0.015	Nonconformity
Second Digits	0.000 to 0.008	Close conformity
	0.008 to 0.010	Acceptable conformity
	0.010 to 0.012	Marginally acceptable conformity
	Above 0.012	Nonconformity
First-Two Digits	0.0000 to 0.0012	Close conformity
	0.0012 to 0.0018	Acceptable conformity
	0.0018 to 0.0022	Marginally acceptable conformity
	Above 0.0022	Nonconformity

Table 4. Critical values and conclusion for MAD-testing (Nigrini & Wells, 2012: 160).

To conclude, the most suitable measure of significance is the MAD measure as it has no limitations regarding the size of the data set. Thus, the hypothesis testing will be conducted by using the MAD value. The values for Z-statistic and Pearson's chi-squared test are presented as reference.

5.2. Defining the data

The previous section offers good guidelines for selecting a data set which gives accurate results of the state of the data formation process in the cryptocurrency markets based on whether the results fit to BL. This section describes the relevance of each set of guidelines and introduces the assumptions used for data selection. As there are 2,063 cryptocurrencies listed on CoinMarketCap (2018c), the cryptocurrencies included are selected randomly from the top 100 cryptocurrencies based on market capitalization. Benford (1938) selected 20 different categories from which he collected 20,229 observations. This study includes 30 different cryptocurrencies which result in 26,709 observations regarding price and 25,948 observations regarding volume.

The randomness of data mentioned by Benford (1938) increases as more “categories” (cryptocurrencies) are added to the data set. If the data set would consist only of one cryptocurrency, each number would be formed based on the previous number, as the previous number gives a specific starting point for the following value. However, this interpretation applies only when considering the value of the cryptocurrency. By adding more cryptocurrencies, the data set also becomes less conditioned. However, over time, the observations become less related. The price and trading volume data get a significantly larger range of values when adding cryptocurrencies with different prices and number of investors participating in trading.

Cryptocurrency data is available in most major currencies, including valuation in Bitcoin, which is the main trading currency in the cryptocurrency markets. The selected data concentrates on valuations in USD and Bitcoin (BTC). Even though one single coin or token might be valued under USD 0.01, CoinMarketCap reports the values with a precision of six decimals. Bitcoin as a currency can be reported with a precision of 8 decimals, as the smallest recordable amount of Bitcoin on the block chain is one satoshi. A satoshi is a one hundred millionth of a single Bitcoin. Thus, the numbers contain three or more relevant digits even though the first digits are zeros.

As mentioned earlier, cryptocurrency markets are different from stock markets in the way that they are open seven days a week and throughout the whole day. This enables the data set to contain two more days of observed data per each week compared to the international financial markets. A year of data from one cryptocurrency leads to 365 observations. The earliest records of Bitcoin data are available from April 28th, 2013, thus the first observations used in this study are from that date. The last observations used in this study are from November 2nd, 2018, which leads to a five-and-a-half-year research period.

Most of the cryptocurrencies used in this study have been listed on exchanges later than Bitcoin. Thus, the research period is shorter for the other cryptocurrencies. However, the total amount of price data observations is 26,709 and 25,948 for volume observations covering the 30 cryptocurrencies included in this research. The amount of observations included in this research can thus be categorised as a large data set. The volume data set contains fewer observations than the price data set due to discrepancies in accumulated data in the beginning of the observation period and days with no reported volume.

The cryptocurrency markets are a new phenomenon for many investors. Thus, the market capitalization of cryptocurrencies is still relatively small. The market capitalization of all the tradeable cryptocurrencies combined was slightly over 205 billion on November 1st, 2018. Bitcoin alone accounted for 54 percent of the market capitalization at that time. In comparison, the 100th largest cryptocurrency based on market capitalization had a market capitalization of only 64 million. (CoinMarketCap, 2018a.)

The descriptive statistics regarding price data are shown in Table 5. It includes the statistics for the amount of observations (N), average of values, median of values, skew of values and minimum and maximum values. The reason for the mean minus median statistic is that one indicator for BL's validity in a data set is that the mean minus median statistic is positive. Only three cryptocurrencies stand out from the data set in this regard: Decentraland, Binance Coin and MetaverseETP. The positiveness of the skew can also be taken as an indicator for the data set fitting BL. Only Binance Coin's skew gets a negative value. The statistics for the whole price data set combined is shown on the last row. The skew of the whole data set is 9.010 and the mean minus median value is 187.250. According to these statistics, the price data is expected to follow BL. The values are shown with the precision of five decimals.

Descriptive statistics for volume data are shown in Table 6. The volume data set does not contain as many observations as the price data set because the volume data tracking contains discrepancies in the beginning of the data-period. The volume data is entirely within the guidelines of a data set that would result in a BL distribution. The skew of the volume data set is 11.632 and mean minus median results in a value of 147,434,574.

Table 5. Descriptive statistics – price data.

Cryptocurrency	N	Mean	Median	Mean-Median	Skewness	Minimum	Maximum
Bytom	452	0.327	0.293	0.033	1.193	0.039	1.170
Lisk	941	4.932	2.710	2.222	2.006	0.105	34.110
Mithril	224	0.516	0.413	0.103	1.220	0.088	1.340
Decentraland	412	0.084	0.087	**-0.002**	0.144	0.009	0.249
XRP	1,917	0.159	0.008	0.151	4.043	0.003	3.380
Aion	381	2.119	1.700	0.419	1.716	0.364	10.520
Bitcoin	2015	2,245.800	592.690	1,653.110	2.064	68.430	19,497.400
Ardor	765	0.232	0.134	0.098	3.139	0.009	2.100
GXChain	496	3.017	2.730	0.287	1.408	0.931	10.020
Binance Coin	466	8.617	9.850	**-1.233**	**-0.207**	0.100	22.760
Nano	591	3.713	1.890	1.823	2.465	0.007	33.700
MaidSafeCoin	1,650	0.175	0.079	0.096	1.654	0.011	1.170
Siacoin	1,164	0.006	0.001	0.006	3.365	0.000	0.094
NEM	1,312	0.130	0.007	0.123	3.463	0.000	1.840

Bitcoin Gold	376	101.302	53.200	48.102	1.402	17.780	500.130
EOS	490	6.351	5.815	0.536	0.620	0.493	21.540
DigiByte	1,731	0.009	0.000	0.009	2.657	0.000	0.127
Monero	1,626	53.526	3.135	50.391	2.063	0.224	469.200
Metaverse ETP	516	2.407	2.500	**-0.093**	0.396	0.341	5.980
BitShares	1,566	0.071	0.009	0.062	2.900	0.003	0.892
Dash	1,723	123.590	8.820	114.770	2.675	0.315	1,550.850
Stratis	813	3.633	2.730	0.903	1.545	0.011	21.750
Funfair	494	0.035	0.026	0.009	2.769	0.010	0.193
Tron	416	0.036	0.035	0.001	2.018	0.001	0.221
Electroneum	366	0.040	0.023	0.017	1.583	0.005	0.195
0x	444	0.767	0.716	0.051	0.756	0.171	2.370
Decred	997	30.434	25.820	4.614	0.902	0.422	122.740
DigixDAO	929	81.245	52.870	28.375	1.933	6.420	555.440
Waltonchain	433	9.535	7.130	2.405	1.513	0.612	41.730

Augur	1,003	19.957	14.420	5.537	1.828	1.470	108.470
Data set	**26,709**	**187.785**	**0.535**	**187.250**	**9.010**	**0.000**	**19,497.400**

Table 6. Descriptive statistics – volume data.

Cryptocurrency	N	Mean	Median	Mean-Median	Skewness	Min	Max
Bytom	452	35,446,958	21,798,100	13,648,858	1.86339	5,085	272,244,992
Lisk	941	15,131,787	4,967,850	10,163,937	4.13281	6,142	309,340,000
Mithril	224	58,091,637	32,878,000	25,213,637	2.41461	4,557,390	358,718,016
Decentraland	412	11,341,398	5,828,910	5,512,488	3.11055	181,704	107,251,000
XRP	1,772	206,017,154	1,031,565	204,985,589	7.08287	8,316	9,110,439,936
Aion	381	5,481,914	3,214,820	2,267,094	3.23829	10,755	60,766,500
Bitcoin	1,772	1,589,674,925	71,501,300	1,518,173,625	2.96747	2,857,830	23,840,899,072
Ardor	755	4,991,612	1,542,490	3,449,122	7.80333	7	199,863,008
GXChain	493	8,852,482	7,223,360	1,629,122	1.38642	1,455	54,450,500
Binance Coin	466	56,120,360	40,400,650	15,719,710	3.56842	9,284	637,020,992
Nano	590	13,385,476	3,917,685	9,467,791	6.30171	1,917	396,790,016
MaidSafeCoin	1,639	1,084,692	275,660	809,032	4.00884	1,011	22,130,000
Siacoin	1,121	1,031,9342	1,073,960	9,245,382	9.65422	1,027	612,913,024
NEM	1,237	12,960,097	661,495	12,298,602	4.96187	1,009	332,371,008

Bitcoin Gold	376	62,466,110	16,722,200	45,743,910	6.52568	1,792,550	1,688,999,936
EOS	490	584,294,151	477,827,500	106,466,651	2.30789	4,556,540	4,870,720,000
DigiByte	1,611	4,476,889	29,656	4,447,233	7.96242	1,007	232,192,992
Monero	1,626	22,266,278	824,358	21,441,920	4.42267	7,900	543,884,032
Metaverse ETP	516	5,258,050	4,106,275	1,151,775	2.75712	14,164	50,356,000
BitShares	1,566	10,759,256	267,570	10,491,686	10.46687	11,052	726,073,024
Dash	1,723	39,420,963	560,892	38,860,071	2.85873	9,604	816,872,000
Stratis	813	10,167,028	4,326,670	5,840,358	4.17676	6,461	167,018,000
Funfair	494	4,861,399	1,537,810	3,323,589	6.35440	8,783	131,221,000
Tron	416	268,463,306	172,711,996	95,751,310	4.94424	26,475	4,089,410,048
Electroneum	366	2,576,424	1,089,785	1,486,639	5.28826	146,541	48,002,900
0x	444	15,650,175	11,516,300	4,133,875	4.46054	704,645	208,910,000
Decred	997	1,961,739	788,266	1,173,473	6.21947	2,725	61,502,000
DigixDAO	925	4,402,412	284,247	4,118,165	9.40405	1,026	295,958,016
Waltonchain	433	13,973,756	8,067,540	5,906,216	8.38303	1,102,610	325,576,000

Augur	897	4,606,886	1,912,630	2,694,256	11.29242	1,048	23,639,008
Data set	**25,948**	**149,942,384**	**2,507,810**	**147,434,574**	**11.63221**	**7**	**23,840,899,072**

6. OBTAINED RESULTS

The applicability of Benford's Law on price and volume data obtained from the cryptocurrency markets and represented by 30 different currencies is demonstrated below. The data consists of observations gathered by utilizing CoinMarketCap which combines data from every market available, thus representing the close-to-entire cryptocurrency market. This section aims to describe the state of data formation occurring in the cryptocurrency markets by analysing daily observed price and volume data. The state of market data formation is determined by the level of deviation from the Benford's Law distribution and is tested for significance by using MAD as the main measure and chi-squared and Z-statistic as reference points.

The research is conducted on two different levels. First, the results for the entire data set of price and volume data is presented in order to gain view of the whole market and to determine whether a larger data set conforms more closely than the subsamples individually. Second, the testing is conducted for each cryptocurrency included in this study to determine which are the ones resulting in the highest deviation from Benford's Law.

6.1. Entire price data test results

The below table (Table 7) shows the results of testing whether the daily price data for all the included cryptocurrencies fit the Benford's Law distribution in terms of first digit frequencies. In the table (and the following tables), AP stands for the actual observed proportion, EP for expected proportion, Diff for the difference between the actual and expected proportions, AbsDiff for the absolute value of the difference, Z-stat for the Z-statistic and X^2-test for the Pearson's chi-squared statistic. The largest deviations are produced by digits 1 and 2 but are relatively small (1.4% and 1.1% respectively). The table indicates that digits 1, 2, 5, 6, 7 and 9 appear more frequently or unfrequently than expected according to the Z-statistic. However, this does not cause a significant enough deviation to flag the data set as manipulated according to the MAD-measure. The significance testing is determined by using the mean absolute deviation (MAD) score. In this case, the first digit frequencies result in a MAD-score of 0.00535 which, by using the guidelines of table 4, indicates close conformity. Thus, the null hypothesis is accepted in terms of first digit frequencies of the entire set of price data. Both Z-statistic and X^2-statistic indicate that there are significant deviations from the expected proportions, but, as explained in Section 5.1., both tests become unreliable when examining large data sets.

Table 7. First digit distribution of price data.

Digit	Count	AP	EP	Diff	AbsDiff	Z-stat	X^2-test
1	7,673	0.287	0.301	-0.014	0.014	4.881**	446,032
2	4,996	0.187	0.176	0.011	0.011	4.735**	495,185
3	3,417	0.128	0.125	0.003	0.003	1.441	49,141
4	2,619	0.098	0.097	0.001	0.001	0.573	8,214
5	1,978	0.074	0.079	-0.005	0.005	2.983**	220,594
6	1,926	0.072	0.067	0.005	0.005	3.328**	278,081
7	1,404	0.053	0.058	-0.005	0.005	3.786**	363,110
8	1,376	0.052	0.051	0.001	0.001	0.371	3,756
9	1,320	0.049	0.046	0.003	0.003	2.655**	181,552
N	26,709						
X^2-test	2,045,665**						
MAD	0.00535				Close conformity		

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

The next step is to determine whether the proportions of second digits of price data are in line with the Benford's Law distribution. The results shown in Table 8 indicate that the second digit distribution follows Benford's Law more closely than the first digit distribution, as the largest deviations are produced by the digit 0 and 2 with only 0.9% and 0.6% deviations, respectively. Also, when tested for conformity by using the MAD, the data shows close conformity. Furthermore, even though the Z-statistic and the X^2-statistic do not agree with large data sets, the results show that fewer digits score a high significance on the Z-statistic (digits 0 and 2) and the X^2-statistic value is lower also relative to the critical value (largest contributors to the value are digits 0 and 2). Thus, the null hypothesis is accepted also for the second digit distribution of price data.

Table 8. Second digit distribution of price data.

Digit	Count	AP	EP	Diff	AbsDiff	Z-stat	X^2-test
0	3,432	0.128	0.120	0.009	0.009	4.430**	463,354
1	2,968	0.111	0.114	-0.003	0.003	1,414	47,940
2	2,733	0.102	0.109	-0.006	0.006	3.399**	276,663
3	2,845	0.107	0.104	0.002	0.002	1.160	32,759
4	2,618	0.098	0.100	-0.002	0.002	1.235	37,256
5	2,529	0.095	0.097	-0.002	0.002	1.090	29,223
6	2,419	0.091	0.093	-0.003	0.003	1.566	60,154
7	2,432	0.091	0.090	0.001	0.001	0.390	3,907
8	2,381	0.089	0.088	0.002	0.002	0.900	20,232
9	2,352	0.088	0.085	0.003	0.003	1.784	78,734
N	26,709						
X^2-test	1,050,222**						
MAD	0.00265				Close conformity		

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

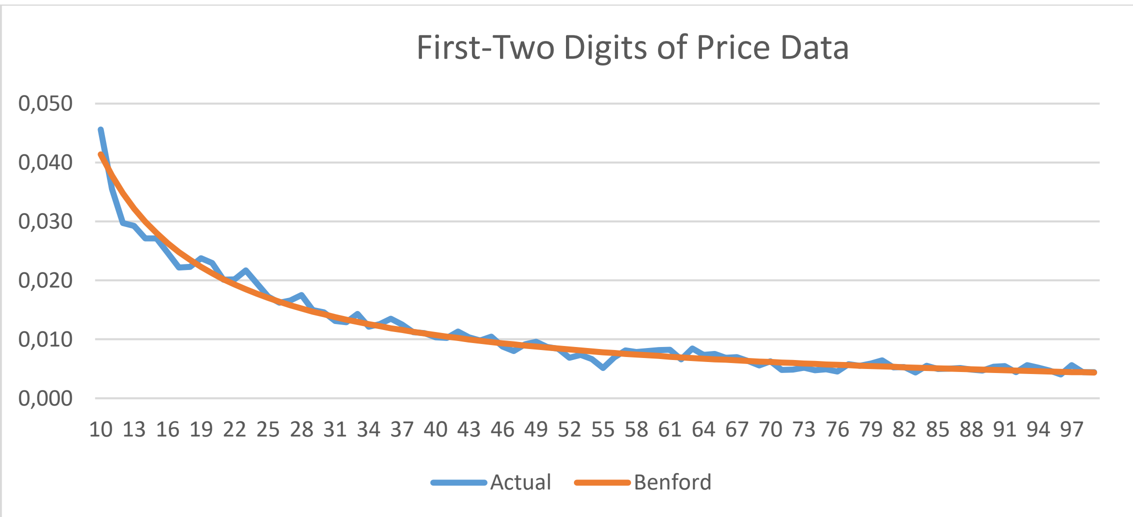

Figure 1. First-two-digit distribution of price data.

The first-two-digit distribution of price data is also demonstrated as a figure (Figure 1). As seen in the figure, there is some deviation from Benford's Law. The largest deviations result from digits 10 and 12, with differences of 0.4% and 0.5% from the expected proportions, respectively. Digits 13, 14, 17, 23 and 55 deviate by approximately 0.3%. However, the overall deviation is relatively small. The MAD-score obtained from the distribution is 0.00088, which also indicates close conformity. Thus, the null hypothesis that the data does not deviate significantly from Benford's Law is accepted also in the case of the distribution formed by the first-two digits.

6.2. Entire volume data test results

Table 9. First digit distribution of volume data.

Digit	Count	AP	EP	Diff	AbsDiff	Z-stat	X^2-test
1	7,968	0.307	0.301	0.006	0.006	2.116*	81,750
2	4,566	0.176	0.176	0.000	0.000	0.044	59
3	3,068	0.118	0.125	-0.007	0.007	3.256**	242,077
4	2,439	0.094	0.097	-0.003	0.003	1.576	59,009
5	2,054	0.079	0.079	0.000	0.000	0.002	4
6	1,809	0.070	0.067	0.003	0.003	1.773	77,144
7	1,507	0.058	0.058	0.000	0.000	0.046	85
8	1,369	0.053	0.051	0.002	0.002	1.161	33,985
9	1,168	0.045	0.046	-0.001	0.001	0.559	8,153
N	25,948						
X^2-test	502,267**						
MAD	0.00234				Close conformity		

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

After determining the validity of price data, the same tests are conducted for the daily volume data from the same cryptocurrencies and same time period. The results for the first digit distribution of volume data is presented in Table 9. The results show that three digits appear in almost exact accordance with Benford's Law: 2, 5 and 7. The largest deviation is caused by the digit 3 with a 0.7% deviation. The next most deviating digit is 1 with a deviation of 0.6%. Digit 3 is, however, flagged as the most deviating one because the deviation is largest relative to the expected proportion. The MAD-score obtained indicates close conformity with Benford's Law. Thus, the null hypothesis cannot be rejected for the first digit distribution of volume data.

Table 10. Second digit distribution of volume data.

Digit	Count	AP	EP	Diff	AbsDiff	Z-stat	X^2-test
0	3,163	0.122	0.120	0.002	0.002	1.091	27,686
1	2,902	0.112	0.114	-0.002	0.002	1.030	24,870
2	2,729	0.105	0.109	-0.004	0.004	1.878	82,411
3	2,761	0.106	0.104	0.002	0.002	1.084	27,802
4	2,661	0.103	0.100	0.002	0.002	1.192	33,772
5	2,443	0.094	0.097	-0.003	0.003	1.367	44,487
6	2,393	0.092	0.093	-0.001	0.001	0.627	9,567
7	2,365	0.091	0.090	0.001	0.001	0.434	4,672
8	2,295	0.088	0.088	0.001	0.001	0.488	5,901
9	2,236	0.086	0.085	0.001	0.001	0.668	10,936
N	25,948						
X^2-test	244,417**						
MAD	0.00184				Close conformity		

** Represents a significance level of 1% ($p<0.01$).

* Represents a significance level of 5% ($p<0.05$).

Table 10 shows the results for the second digit distribution of the volume data. The distribution is completely within the statistical guidelines with the largest deviation resulting from digit 2. Even though the first digit distribution is partially a better fit to Benford's Law, the second digit distribution conforms better overall. This is indicated by the lower value of the X^2-test and by no significant Z-statistics. The MAD-score indicates that the null hypothesis is accepted for the second digit distribution of the volume data.

The results for the first-two-digit distributions are shown in Figure 2. The only relatively large deviation results from the digit 13 with the difference of 0.3% from the expected. The MAD-score obtained for the distribution is 0,00058, which indicates close conformity. The score is better than for price data from which can be concluded that the volume data follows Benford's Law more closely.

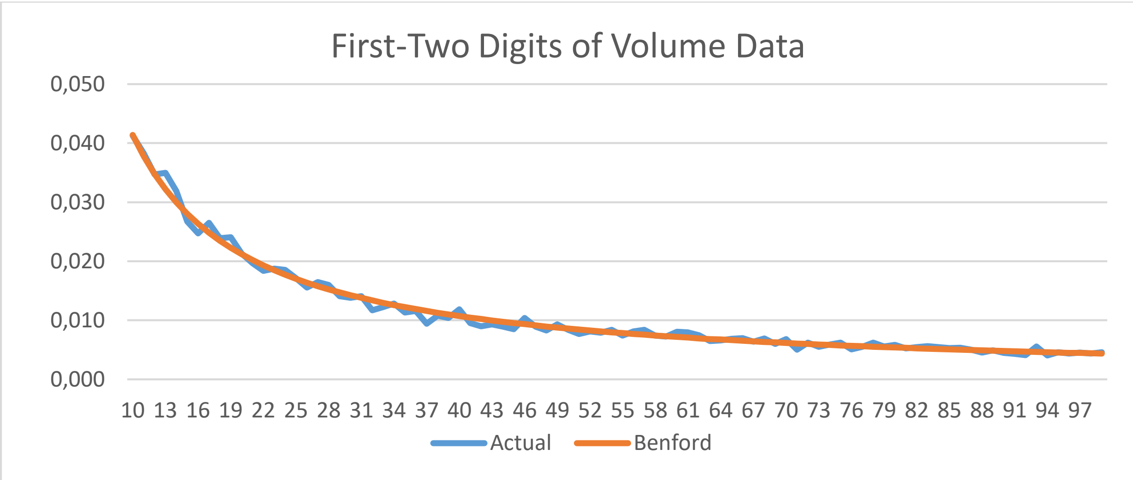

Figure 2. First-two-digit distribution of volume data.

6.3. Summary of total data tests

The tests concerning price and volume data show that as the data set becomes large, it tends to follow the Benford's Law distribution. This being true, we then need to investigate the different sub-samples included in the data set, as a large data set can hide deviations of smaller sub-samples within itself. For example, consider two data sets of ten observations each, one including four values that start with the digit 1 and the second one including two values that start with the digit 1 and the expected amount is three for each data set. Combining these data sets will give a total amount of six observed digits valued 1, which again result in the expected proportion. Thus, by only examining the combined larger data set, deviations of smaller sub-samples might go unnoticed. Next, the data regarding each individual cryptocurrency will be tested to find out whether the sub-samples contain deviations. It is expected, that also the sub-samples follow the distribution formed by the total data set, which again provenly follows the Benford's Law distribution.

6.4. Price data results for individual cryptocurrencies

Table 11 contains the distributions of first digit values for each individual cryptocurrency included in this study. First, when looking at the MAD-score it can be noticed that none of the currencies conform closely with Benford's Law. The only currency which scores an acceptable conformity is Decred with the MAD-score of 0.01172. All other scores indicate nonconformity. Decred is also the only currency with no significant Z-statistics indicating significance on a 1 percent level. Based on the results, a conclusion can be made that the null hypothesis can only be accepted when considering Decred.

The largest deviations based on the MAD-score result from Binance Coin, Decentraland, Tron, EOS, Ardor and XRP, respectively. The score from Binance Coin is by far the highest (0.10410) and the other ones mentioned range from 0.06 to 0.07. For Binance Coin, the only digit that does not deviate significantly is 8. All other digits deviate at a magnitude that is statistically significant at a 1 percent level. The most deviating digit is 1, with a probability of 65,9 percent. For Decentraland, the most deviating digit is also 1, with a probability of 50 percent, and the only digit that is not statistically significant is 6. XRP forms the largest value of the X^2-test (1,611,094 and 78,7 percent of the X^2-value of the total sample), which indicates that proportions of larger digits deviate more than the smaller digits while the amount of observations is relatively large. In the case of XRP, there is fewer digit ones than there is sixes and the larger digits seem to dominate the data. From the table it can be noticed that proportions are evenly under and over the expected values, which leads to close conformity regarding the entire data set.

Table 12 contains the second digit distributions of individual cryptocurrencies' price data. Overall, the second digit distributions deviate less from the expected than the first digit distributions. The observed MAD-scores lead to the same conclusion. There are three currencies regarding which second digit distribution confirms closely with the expected distribution: Siacoin, BitShares and Augur. However, all three scored nonconformities when examining first digits. In the acceptable conformity range there is XRP, Nano, Bitcoin Gold, EOS, DigiByte and MetaverseETP. The rest got a score of marginally acceptable conformity or nonconformity. Interestingly, according to the Z-statistic, Waltonchain should conform closely with Benford's Law but it gets a score of nonconformity when measured with MAD. This means that each digit deviates from the expected enough to result in a high MAD, but not enough for the digits to be significant according to the Z-statistic. Then again, if one digit deviates significantly according to the Z-statistic and the rest conform closely, the MAD-score "forgives" the significant deviation by taking the average deviation.

None of the digits come close to the 1 percent significance level critical value of the chi-squared test for 9 degrees of freedom (21,67). This could mean that all currencies deviate significantly from the expected distribution on a 1 percent level. As the chi-squared-test is only used as a reference, the take-away is that the larger the test value, the larger the deviations relative to the expected values. The largest deviations relative to expected values are found in Bitcoin's data (chi- squared value of 123,589). This is mostly the result of digit 1's deviation, which is approximately 20 percent. Decred, the only currency that scored acceptable conformity when examining first digits, scores nonconformity along with three deviations that are significant on the 1 percent level (digit 2, 5 and 6).

As a conclusion of Table 12, there are some currencies that conform well with the expected distribution but are nonconforming when testing the first digits, which indicates that the distribution does not match with the expected distribution.

As shown in Table 13, when examining the first-two digits of the price data, every individual coin scores a MAD-measure of nonconformity. The highest values are measured for Binance Coin (0.01128) and Decentraland (0.01117), which were also the two coins with the highest deviation when examining first digits. In addition, when examining the second digits, Binance Coin has the highest deviation as well. The lowest deviations were recorded for Siacoin (0.00311) and MaidSafeCoin (0.00350).

The distributions formed by the price data of individual coins did not give results as expected. The total price data set conforms closely with Benford's Law, but none of the sub samples do. Also, the magnitude of the deviations is considerably large for some sub samples. This can be due to a relatively small data set regarding individual coins, the fact

Table 11. First digit distributions of subsample price data.

	1	2	3	4	5	6	7	8	9	N	X^2-test	MAD
Bytom	0.325	0.088**	0.239**	0.086	0.066	0.073	0.035	0.046	0.040	452	33,320**	0.03204
Lisk	0.303	0.247**	0.150*	0.075*	0.087	0.040**	0.058**	0.051	0.021**	941	64,190**	0.02342
Mithril	0.134**	0.259**	0.196**	0.116	0.116	0.080	0.031	0.045	0.022	224	11,127**	0.04975
Decentraland	0.500**	0.022**	0.000**	0.000**	0.012**	0.083	0.158**	0.092**	0.133**	412	156,189**	0.09845
XRP	0.158**	0.127**	0.054**	0.125**	0.129**	0.183**	0.093**	0.101**	0.030**	1,917	1,611,094**	0.06204
Aion	0.129**	0.228**	0.152	0.205**	0.094	0.055	0.039	0.063	0.034	381	37,290**	0.04773
Bitcoin	0.172**	0.189	0.095**	0.140**	0.070	0.160**	0.073**	0.056	0.046	2,015	886,802**	0.03753
Ardor	0.550**	0.196	0.080**	0.044**	0.029**	0.012**	0.005**	0.012**	0.072**	765	248,102**	0.06562
GXChain	0.266	0.331**	0.190**	0.119	0.056	0.006**	0.014**	0.006**	0.012**	496	83,126**	0.05358
Binance Coin	0.659**	0.079**	0.006**	0.011**	0.013**	0.006**	0.019**	0.049	0.157**	466	232,279**	0.10410
Nano	0.349**	0.222**	0.088**	0.052**	0.052*	0.041**	0.059	0.080**	0.058	591	31,074**	0.02989
MaidSafeCoin	0.273**	0.235**	0.155**	0.093	0.064*	0.041**	0.058	0.045	0.035*	1,650	125,781**	0.01989
Siacoin	0.241**	0.170	0.152**	0.113	0.088	0.078	0.067	0.056	0.034	1,164	38,143**	0.01714
NEM	0.443**	0.167	0.117	0.056**	0.065	0.056	0.015**	0.018**	0.063**	1,312	254,845**	0.03519
Bitcoin Gold	0.237**	0.362**	0.133	0.101	0.056	0.040*	0.043	0.024**	0.005**	376	39,947**	0.04394
EOS	0.357**	0.024**	0.016**	0.049**	0.218**	0.086	0.073	0.124**	0.051	490	149,057**	0.06849
DigiByte	0.268**	0.292**	0.183**	0.085	0.041**	0.036**	0.032**	0.028**	0.034*	1,731	499,439**	0.03877
Monero	0.319	0.144**	0.087**	0.165**	0.087	0.046**	0.036**	0.059	0.057*	1,626	228,075**	0.02510

MetaverseETP	0.209**	0.217**	0.275**	0.064**	0.072	0.047	0.021**	0.072*	0.023**	516	74,292**	0.04710
BitShares	0.250**	0.084**	0.231**	0.168**	0.102**	0.064	0.026**	0.038**	0.037	1,566	560,866**	0.04451
Dash	0.259**	0.266**	0.132	0.065**	0.037**	0.073	0.070*	0.045	0.053	1,723	267,423**	0.02705
Stratis	0.220**	0.112**	0.149*	0.098	0.114**	0.116**	0.092**	0.049	0.049	813	80,138**	0.03260
Funfair	0.358**	0.287**	0.121	0.095	0.081	0.026**	0.002**	0.010**	0.018**	494	51,232**	0.03794
Tron	0.101**	0.339**	0.214**	0.156**	0.048*	0.048	0.050	0.034	0.010**	416	75,517**	0.06914
Electroneum	0.306	0.260**	0.063**	0.027**	0.074	0.087	0.077	0.068	0.038	366	18,804**	0.03218
0x	0.342	0.149	0.041**	0.018**	0.088	0.106**	0.086**	0.083**	0.088**	444	44,651**	0.04240
Decred	0.308	0.155	0.149*	0.083	0.070	0.062	0.053	0.057	0.061*	997	16,672**	0.01172
DigixDAO	0.335*	0.095**	0.081**	0.052**	0.051**	0.052	0.088**	0.082**	0.166**	929	378,145**	0.04769
Waltonchain	0.330	0.185	0.115	0.067*	0.065	0.104**	0.051	0.058	0.025	433	8,860**	0.01815
Augur	0.312	0.211**	0.142	0.139**	0.065	0.031**	0.032**	0.030**	0.039	1,003	71,522**	0.02324
Total	0.287**	0.187**	0.128	0.098	0.074**	0.072**	0.053**	0.052	0.049**	26,709	2,045,665**	

The significance level presented with each probability refers to the result given by the Z-statistic.

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

MAD-score	0.00000 to 0.00600	Close conformity	0.00600 to 0.01200	Acceptable conformity
	0.01200 to 0.01500	Marginally acceptable conformity	Above 0.01500	Nonconformity

Table 12. Second digit distributions of subsample price data.

	0	1	2	3	4	5	6	7	8	9	N	X^2-test	MAD
Bytom	0.113	0.102	0.084	0.075	0.080	0.097	0.100	0.128**	0.131**	0.091	452	11,751**	0.01870
Lisk	0.095*	0.108	0.094	0.114	0.111	0.118*	0.100	0.081	0.099	0.082	941	15,308**	0.01173
Mithril	0.085	0.147	0.085	0.089	0.063	0.098	0.129	0.071	0.129*	0.103	224	4,181**	0.02613
Decentraland	0.167**	0.114	0.114	0.136*	0.092	0.090	0.056**	0.090	0.090	0.051**	412	9,986**	0.01741
XRP	0.142**	0.104	0.091**	0.108	0.101	0.090	0.084	0.099	0.094	0.087	1,917	40,606**	0.00879
Aion	0.160**	0.134	0.094	0.081	0.113	0.073	0.084	0.084	0.081	0.094	381	4,953**	0.01649
Bitcoin	0.101**	0.111	0.138**	0.138**	0.097	0.092	0.082	0.101	0.080	0.061**	2,015	123,589**	0.01473
Ardor	0.161**	0.163**	0.088	0.115	0.090	0.081	0.071*	0.065**	0.076	0.090	765	34,411**	0.02130
GXChain	0.095	0.077**	0.077*	0.123	0.115	0.115	0.099	0.115	0.093	0.093	496	10,784**	0.01888
Binance Coin	0.165**	0.084*	0.150**	0.146**	0.133*	0.094	0.069	0.069	0.036**	0.054**	466	26,389**	0.03226
Nano	0.135	0.105	0.112	0.102	0.095	0.102	0.108	0.083	0.069	0.090	591	3,667**	0.00860
MaidSafeCoin	0.127	0.092**	0.097	0.090	0.097	0.101	0.086	0.103	0.106**	0.101*	1,650	47,645**	0.01178
Siacoin	0.147**	0.114	0.106	0.096	0.097	0.093	0.085	0.082	0.095	0.085	1,164	12,815**	0.00709
NEM	0.134	0.094*	0.088*	0.098	0.108	0.124**	0.117**	0.090	0.075	0.071	1,312	48,628**	0.01479
Bitcoin Gold	0.109	0.136	0.096	0.088	0.101	0.096	0.114	0.082	0.088	0.090	376	2,131**	0.00982
EOS	0.139	0.114	0.106	0.116	0.100	0.065*	0.088	0.102	0.092	0.078	490	4,174**	0.00949
DigiByte	0.112	0.124	0.099	0.121*	0.113	0.093	0.084	0.087	0.071**	0.094	1,731	35,635**	0.00988
Monero	0.132	0.127	0.113	0.094	0.074**	0.094	0.077*	0.090	0.092	0.106**	1,626	50,012**	0.01092

MetaverseETP	0.141	0.128	0.110	0.101	0.105	0.081	0.091	0.072	0.079	0.091	516	3,604**	0.00958
BitShares	0.135	0.105	0.099	0.100	0.105	0.095	0.102	0.090	0.077	0.093	1,566	16,695**	0.00730
Dash	0.109	0.124	0.103	0.102	0.083**	0.085	0.089	0.093	0.111**	0.100*	1,723	48,382**	0.01042
Stratis	0.106	0.082**	0.095	0.117	0.098	0.123**	0.108	0.098	0.089	0.084	813	15,842**	0.01255
Funfair	0.083**	0.077**	0.079*	0.113	0.138**	0.121	0.109	0.103	0.089	0.087	494	13,938**	0.02070
Tron	0.099	0.106	0.132	0.127	0.094	0.079	0.087	0.070	0.096	0.111	416	5,492**	0.01613
Electroneum	0.137	0.142	0.148*	0.112	0.104	0.077	0.057*	0.052**	0.074	0.098	366	8,380**	0.02168
0x	0.133	0.124	0.115	0.095	0.086	0.077	0.083	0.095	0.088	0.106	444	3,216**	0.01092
Decred	0.138	0.097	0.082**	0.088	0.092	0.120**	0.118**	0.100	0.083	0.079	997	28,968**	0.01547
DigixDAO	0.209**	0.108	0.066**	0.094	0.097	0.071**	0.081	0.087	0.096	0.093	929	82,113**	0.02099
Waltonchain	0.134	0.134	0.125	0.092	0.076	0.076	0.085	0.088	0.085	0.104	433	4,510**	0.01383
Augur	0.126	0.113	0.106	0.107	0.090	0.076*	0.092	0.092	0.109*	0.092	1,003	11,827**	0.00750
Total	0.128**	0.111	0.102**	0.107	0.098	0.095	0.091	0.091	0.089	0.088	26,709	1,050,222**	0.00265

The significance level presented with each probability refers to the result given by the Z-statistic.

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

MAD-score	0.00000 to 0.00800	Close conformity	0.00800 to 0.01000	Acceptable conformity
	0.01000 to 0.01200	Marginally acceptable conformity	Above 0.01200	Nonconformity

that each observation in the subsample are in relation or that the price formation is not natural. The results may also indicate which coins should be more closely looked into. In order to get a more formal comparison, the volume data for each individual coin is examined.

Table 13. Deviations of first-two digits of price data.

	MAD		MAD
Bytom	0.00647	EOS	0.00819
Lisk	0.00437	DigiByte	0.00420
Mithril	0.00843	Monero	0.00412
Decentraland	0.01117	MetaverseETP	0.00606
XRP	0.00670	BitShares	0.00489
Aion	0.00650	Dash	0.00396
Bitcoin	0.00589	Stratis	0.00463
Ardor	0.00706	Funfair	0.00630
GXChain	0.00692	Tron	0.00866
Binance Coin	0.01128	Electroneum	0.00612
Nano	0.00437	0x	0.00652
MaidSafeCoin	0.00350	Decred	0.00378
Siacoin	0.00311	DigixDAO	0.00659
NEM	0.00467	Waltonchain	0.00498
Bitcoin Gold	0.00729	Augur	0.00401
MAD-Score	Above 0.00220	Nonconformity	

6.5. Volume data results for individual cryptocurrencies

Table 14 contains the results regarding the first digit distributions of volume data on an individual subsample level. The first thing to notice is that the table contains less significant deviations according to the Z-statistic than the first digit distributions of price data. One reason for this could be that none of the observations are in relation, as the trading volume of today is not determined by the trading volume of yesterday. This is not the case regarding price data, as the day's starting price is determined by yesterday's closing price. This also means that the volume data is generated by a more random and natural process. When looking at the MAD-scores, it can also be noticed that the volume data set has more subsamples that fall in the acceptable conformity range: Dash, Siacoin, Bitcoin, Nano, XRP, DigixDAO, Stratis, BitShares and NEM. The sample also contains subsamples that fall in the marginally acceptable range: Monero and Decred. The coins are listed in an increasing order in terms of the MAD-score.

None of the subsamples conform closely with Benford's distribution, but 11 conform at some level. On the other hand, the highest MAD-scores result from subsamples containing the volume data of Decentraland, GXChain, EOS, Waltonchain and 0x, which all give a score from 0.03 to 0.04. Decentraland was the 2nd and EOS the 4th worst conforming subsamples according to first digits of price data. When looking at the chi-squared results, Nano gets the lowest statistic even though digit 5 deviates from the expected on a 5 percent significance level. As indicated by the significance level, the deviation is not very large, and as it is the median of the digits, chi-squared test does not give the deviation much weight. Also, the overall values of the chi-squared statistic are relatively low compared to those of the price data. The highest statistic results from MaidSafeCoin with the value of 66,839 compared to the price data first digit distribution of XRP with the value of 1,611,094. From the table it can be concluded that the overall performance of first digit distributions concerning subsample volume data is better than that of price data, as the null hypothesis cannot be rejected for 11 of the subsamples.

Table 14. First digit distributions of subsample volume data.

	1	2	3	4	5	6	7	8	9	N	X^2-test	MAD
Bytom	0.308	0.115**	0.108	0.091	0.108*	0.086	0.060	0.075*	0.049	452	10,638**	0.01863
Lisk	0.266*	0.153	0.114	0.111	0.092	0.086*	0.071	0.063	0.045	941	20,810**	0.01577
Mithril	0.321	0.161	0.107	0.103	0.071	0.040	0.076	0.063	0.058	224	1,421**	0.01501
Decentraland	0.136**	0.170	0.172**	0.109	0.146**	0.092	0.066	0.066	0.044	412	30,748**	0.03855
XRP	0.288	0.199**	0.133	0.093	0.074	0.070	0.055	0.054	0.033**	1,772	26,036**	0.00809
Aion	0.312	0.205	0.165*	0.097	0.037**	0.037*	0.066	0.031	0.050	381	9,173**	0.02045
Bitcoin	0.302	0.161	0.115	0.115**	0.090	0.069	0.061	0.047	0.041	1,772	25,276**	0.00768
Ardor	0.385**	0.181	0.105	0.074*	0.050**	0.050	0.049	0.056	0.049	755	27,990**	0.02167
GXChain	0.432**	0.150	0.051**	0.059**	0.051*	0.069	0.063	0.079**	0.047	493	35,520**	0.03703
Binance Coin	0.204**	0.221**	0.120	0.112	0.077	0.062	0.077	0.064	0.062	466	13,302**	0.02411
Nano	0.276	0.181	0.117	0.102	0.102*	0.064	0.059	0.051	0.047	590	3,355**	0.00789
MaidSafeCoin	0.350**	0.196*	0.102**	0.082*	0.061**	0.054*	0.055	0.057	0.042	1,639	66,839**	0.01689
Siacoin	0.303	0.186	0.126	0.092	0.077	0.075	0.043*	0.042	0.056	1,121	12,273**	0.00703
NEM	0.314	0.150**	0.118	0.087	0.068	0.084**	0.061	0.066**	0.052	1,237	25,939**	0.01187
Bitcoin Gold	0.391**	0.184	0.077**	0.056**	0.059	0.048	0.059	0.066	0.061	376	11,802**	0.02859
EOS	0.222**	0.133**	0.104	0.096	0.131**	0.139**	0.076	0.061	0.039	490	36,920**	0.03358
DigiByte	0.340**	0.207**	0.133	0.092	0.061**	0.053*	0.043**	0.037**	0.034**	1,611	76,477**	0.01741
Monero	0.300	0.218**	0.145**	0.091	0.055**	0.060	0.047	0.046	0.038	1,626	68,020**	0.01375

MetaverseETP	0.244	0.136	0.114	0.136	0.109	0.089	0.070	0.064	0.039	516	16,364**	0.02557
BitShares	0.295	0.154*	0.116	0.088	0.104**	0.080*	0.064	0.056	0.043	1,566	38,961**	0.01074
Dash	0.286	0.175	0.129	0.109	0.073	0.062	0.063	0.053	0.050	1,723	12,141**	0.00612
Stratis	0.305	0.155	0.100*	0.103	0.084	0.073	0.073	0.060	0.048	813	9,446**	0.01031
Funfair	0.352**	0.154	0.087**	0.075	0.091	0.069	0.051	0.065	0.057	494	9,057**	0.01991
Tron	0.397**	0.209	0.108	0.079	0.046**	0.043	0.041	0.038	0.038	416	12,798**	0.02862
Electroneum	0.246*	0.167	0.090	0.128	0.120**	0.087	0.066	0.044	0.052	366	8,170**	0.02368
0x	0.437**	0.155	0.090*	0.074	0.050*	0.070	0.041	0.036	0.047	444	19,641**	0.03112
Decred	0.347**	0.158	0.098**	0.076*	0.079	0.066	0.062	0.052	0.060*	997	23,509**	0.01462
DigixDAO	0.280	0.188	0.132	0.083	0.089	0.067	0.063	0.054	0.044	925	5,504**	0.00810
Waltonchain	0.303	0.097	0.085	0.081	0.122	0.134	0.083	0.042	0.053	433	29,092**	0.03203
Augur	0.271	0.204*	0.155**	0.108	0.086	0.065	0.042	0.035*	0.035	897	23,304**	0.01688
Total	0.307*	0.176	0.118**	0.094	0.079	0.070	0.058	0.053	0.045	25,948	502,267**	0.00234

The significance level presented with each probability refers to the result given by the Z-statistic.

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

MAD-score	0.00000 to 0.00600	Close conformity	0.00600 to 0.01200	Acceptable conformity
	0.01200 to 0.01500	Marginally acceptable conformity	Above 0.01500	Nonconformity

Table 15. Second digit distributions of subsample volume data.

	0	1	2	3	4	5	6	7	8	9	N	X^2-test	MAD
Bytom	0.115	0.108	0.104	0.124	0.104	0.084	0.124*	0.082	0.066	0.088	452	4,526**	0.01145
Lisk	0.139	0.094	0.109	0.107	0.100	0.095	0.088	0.087	0.098	0.083	941	7,624**	0.00667
Mithril	0.112	0.112	0.080	0.080	0.116	0.076	0.138*	0.098	0.071	0.116	224	2,871**	0.01994
Decentraland	0.126	0.097	0.102	0.131	0.080	0.129*	0.075	0.100	0.097	0.063	412	6,091**	0.01678
XRP	0.128	0.113	0.102	0.098	0.108	0.087	0.086	0.098	0.097	0.082	1,772	17,071**	0.00676
Aion	0.129	0.081	0.108	0.113	0.113	0.110	0.094	0.108	0.073	0.071	381	3,203**	0.01239
Bitcoin	0.119	0.105	0.112	0.105	0.090	0.097	0.092	0.084	0.102*	0.094	1,772	17,233**	0.00544
Ardor	0.135	0.111	0.091	0.098	0.094	0.106	0.093	0.086	0.103	0.082	755	5,494**	0.00809
GXChain	0.118	0.140	0.120	0.099	0.110	0.103	0.075	0.089	0.065	0.081	493	4,445**	0.01058
Binance Coin	0.124	0.101	0.107	0.094	0.124	0.090	0.101	0.082	0.079	0.097	466	2,758**	0.00960
Nano	0.115	0.117	0.100	0.103	0.105	0.115	0.095	0.092	0.080	0.078	590	2,124**	0.00583
MaidSafeCoin	0.110	0.107	0.099	0.108	0.102	0.101	0.108*	0.100	0.087	0.077	1,639	17,043**	0.00684
Siacoin	0.113	0.112	0.095	0.112	0.104	0.084	0.093	0.105	0.092	0.088	1,121	9,173**	0.00693
NEM	0.128	0.117	0.120	0.107	0.098	0.080	0.100	0.082	0.084	0.084	1,237	9,244**	0.00629
Bitcoin Gold	0.130	0.130	0.128	0.114	0.109	0.109	0.112	0.064	0.056*	0.048**	376	6,923**	0.01907
EOS	0.139	0.090	0.112	0.084	0.108	0.116	0.106	0.069	0.092	0.084	490	5,710**	0.01341
DigiByte	0.112	0.111	0.101	0.096	0.097	0.097	0.086	0.104	0.097	0.100*	1,611	21,605**	0.00766
Monero	0.109	0.117	0.119	0.099	0.113	0.098	0.083	0.086	0.083	0.092	1,626	16,483**	0.00706

MetaverseETP	0.124	0.097	0.110	0.093	0.126	0.078	0.093	0.087	0.097	0.095	516	4,414**	0.01019
BitShares	0.133	0.114	0.102	0.114	0.101	0.095	0.094	0.082	0.081	0.084	1,566	9,902**	0.00484
Dash	0.127	0.110	0.098	0.108	0.094	0.089	0.091	0.097	0.095	0.092	1,723	13,856**	0.00639
Stratis	0.114	0.121	0.098	0.119	0.113	0.087	0.073*	0.096	0.098	0.080	813	8,554**	0.01018
Funfair	0.134	0.113	0.101	0.119	0.103	0.085	0.095	0.093	0.075	0.081	494	1,946**	0.00730
Tron	0.135	0.156**	0.089	0.079	0.082	0.072	0.079	0.113	0.115	0.079	416	9,331**	0.02155
Electroneum	0.112	0.104	0.126	0.120	0.096	0.087	0.098	0.085	0.085	0.087	366	1,112**	0.00803
0x	0.108	0.115	0.108	0.110	0.115	0.110	0.099	0.068	0.086	0.081	444	2,336**	0.00819
Decred	0.124	0.136*	0.110	0.124*	0.092	0.082	0.073*	0.094	0.084	0.078	997	16,392**	0.01054
DigixDAO	0.124	0.092*	0.101	0.107	0.116	0.102	0.093	0.084	0.084	0.097	925	8,591**	0.00799
Waltonchain	0.139	0.109	0.097	0.113	0.106	0.088	0.088	0.092	0.074	0.095	433	1,886**	0.00907
Augur	0.107	0.123	0.105	0.113	0.099	0.094	0.103	0.094	0.076	0.088	897	4,535**	0.00651
Total	0.122	0.112	0.105	0.106	0.103	0.094	0.092	0.091	0.088	0.086	25,948	244,417**	0.00184

The significance level presented with each probability refers to the result given by the Z-statistic.

** Represents a significance level of 1% (p<0.01).

* Represents a significance level of 5% (p<0.05).

MAD-score	0.00000 to 0.00800	Close conformity	0.00800 to 0.01000	Acceptable conformity
	0.01000 to 0.01200	Marginally acceptable conformity	Above 0.01200	Nonconformity

The distribution formed by second digits of the subsamples of volume data show small deviation when observing significance based on the Z-statistic, as shown in Table 15. The table contains only two digits from different subsamples that are significant on a 1 percent level (Bitcoin Gold: 9 and Tron: 1). Also, when observing the MAD-measure, 13 subsamples fall in the close conformity range and another 5 are in both the acceptable conformity range and marginally acceptable range. This leaves the data set with only 7 subsamples that result in nonconformity. This is a considerably good result compared to the price data distribution of second digits. Also, the results are in line with other comparisons between price data and volume data. The volume data seem to follow Benford's distributions more closely.

The worst performing subsample is Tron, with a MAD-score of 0.02155. Tron gave a result of nonconformity also in all the previous tests. Next in line, in terms of worst MAD-scores, are Mithril, Bitcoin Gold and Decentraland, which have also previously in this research resulted in nonconformity. As the majority of the subsamples result in close or acceptable conformity, there is no added value in comparing which is the absolute best. From Table 15, a conclusion can be drawn that the null hypothesis is rejected only in 7 subsamples regarding second digit distributions of the volume data.

The first-two digits of subsample volume data follow the same theme as the first and second digit distributions, as they conform more closely with Benford's distribution. Each subsample of price data resulted in nonconformity, but volume data has two subsamples in the marginally acceptable range: Dash and XRP with MAD-measures of 0.00186 and 0.00205, respectively, as indicated in Table 16. The highest deviations are recorded in Decentraland's distribution, which results in a measure of 0.00630. As the distribution of the first-two digits can be assumed to be the combined distribution of the first and second digits, it is expected that the worst performing subsamples according to first or/and second digit distributions are also the worst performing according to the distribution formed by the first-two digits.

Now that all the results are presented for subsamples and total samples of the price and volume data, it is obvious that the price data deviates more significantly from Benford's Law. This can be at least partly explained by the irregularities of the pricing process, randomness of observations and relations between observations. The pricing process can be affected by the initial offer price, psychological aspects affecting investors and plain manipulation of the price. There is no initial trading volume, no psychological drivers for a specific trading volume and the volume can be affected by manipulation attempts. The randomness and relation between price observations go hand-in-hand as the starting price of today is the closing price of yesterday, breaking both rules simultaneously. Volume data does not suffer from either one. However, price data is affected by this problem only for a short period of time, as the lagging "bond" is broken after each trading day. Tomorrows starting price is not affected by yesterday's closing price. In addition, the cryptocurrency markets are running seven days a week and 24 hours a day, which means that there is no closing price in the cryptocurrency markets. There is only the price of the last moment of the day, which changes a second later.

Table 16. Deviations of first-two digits of volume data.

	MAD		MAD
Bytom	0.00371	EOS	0.00457
Lisk	0.00284	DigiByte	0.00251
Mithril	0.00532	Monero	0.00247
Decentraland	0.00630	MetaverseETP	0.00424
XRP	0.00205	BitShares	0.00224
Aion	0.00426	Dash	0.00186
Bitcoin	0.00229	Stratis	0.00336
Ardor	0.00325	Funfair	0.00368
GXChain	0.00538	Tron	0.00491
Binance Coin	0.00411	Electroneum	0.00404
Nano	0.00305	0x	0.00452
MaidSafeCoin	0.00268	Decred	0.00283
Siacoin	0.00243	DigixDAO	0.00285
NEM	0.00258	Waltonchain	0.00455
Bitcoin Gold	0.00514	Augur	0.00312
MAD-score	0.00180 to 0.00220	Marginally acceptable conformity	
	Above 0.00220	Nonconformity	

6.6. Summary of individual subsample tests

It is interesting to notice that the global realization of cryptocurrencies' potential, which resulted in the total market capitalization to increase from around 15 billion in the beginning of 2017 to over 800 billion in the beginning of 2018, does not cause the distribution of the price and volume data set to deviate from Benford's Law. In their research, Riccioni and Cerqueti (2018) mention that the deviations recorded in the price and volume data of the global financial markets can be caused by break through events. The latter half of year 2017 can be thought as a break through event for the cryptocurrency markets. Break through events can also be checked on an individual subsample level, but one would think that subsamples of both data sets would suffer from the same effect. However, this is not the case as the volume data conforms considerably more closely with Benford's Law.

It can be hard to conceive the results regarding each subsample and each test simultaneously, thus a simple summation table is presented below. This makes the interpretation of the results more convenient. The results are given as fail (F) or pass (P). Failing a test means that the subsample does not even marginally conform with Benford's distribution and passing a test means that the subsample conforms on at least one of the three levels of conformity. None of the subsamples pass more than one test when looking at price data. Volume data has two subsamples that pass every test: XRP and Dash. In addition, there is eight subsamples that pass two tests. The table aims to provide a brief analysis of the results, which can then be more closely observed in the previous tables. The table also sums up the result section.

Table 17. Performance of subsamples' distributions.

	Price data			Volume data		
	First	Second	First-two	First	Second	First-two
Bytom	F	F	F	F	P	F
Lisk	F	P	F	F	P	F
Mithril	F	F	F	F	F	F
Decentraland	F	F	F	F	F	F
XRP	F	P	F	P	P	P

Aion	F	F	F	F	F	F
Bitcoin	F	F	F	P	P	F
Ardor	F	F	F	F	P	F
GXChain	F	F	F	F	P	F
Binance Coin	F	F	F	F	P	F
Nano	F	P	F	P	P	F
MaidSafeCoin	F	P	F	F	P	F
Siacoin	F	P	F	P	P	F
NEM	F	F	F	P	P	F
Bitcoin Gold	F	P	F	F	F	F
EOS	F	P	F	F	F	F
DigiByte	F	P	F	F	P	F
Monero	F	P	F	P	P	F
MetaverseETP	F	P	F	F	P	F
BitShares	F	P	F	P	P	F
Dash	F	P	F	P	P	P
Stratis	F	F	F	P	P	F
Funfair	F	F	F	F	P	F
Tron	F	F	F	F	F	F
Electroneum	F	F	F	F	P	F
0x	F	P	F	F	P	F
Decred	P	F	F	P	P	F
DigixDAO	F	F	F	P	P	F
Waltonchain	F	F	F	F	P	F
Augur	F	P	F	F	P	F

P stands for passing the test.

F stands for failing the test.

7. CONCLUSION

In this study Benford's Law is used as an indicator of validity and quality of the cryptocurrency market data, more precisely price- and volume data. Firstly, three tests are conducted on the entire price and volume data set. Further assessment is provided as all the 30 subsamples are tested separately. The Chi squared-test and the z-statistic are used as statistical tools for goodness-of-fit and presented as reference points. However, as even small deviations result in high scores in both previously mentioned tests, a simple measure of the mean average deviation is used to determine the conformity. The MAD measure is far more "forgiving" than the statistical tests, thus gives a more reasonable result when testing unregulated and young markets.

The results show that when the entire price and volume data sets are tested, both conform closely with BL. Further assessments, however, show that only two subsamples of the volume data conform with BL at least on some level. Every subsample of the price data fails at least one of the three tests conducted.

The results also show that the volume data tends to conform more closely with BL than the price data. This is visible both when testing the entire data set and subsamples individually. An explanation may be that, volume data is generated by a more unbiased process, thus resulting in a more natural distribution of numbers. The price data can be more easily affected by psychologically influenced behaviours of investors and schemes reducing market efficiency, which both affect the data.

Finally, the subsample testing gives direction for further research. A comparison between the best and worst conforming subsamples could further explain the differences in the data formation processes. Previous research also indicate that deviations of such magnitude should all be more closely looked into. In addition, conducting similar testing on the entire available cryptocurrency market data is required. This would give a more complete view of the markets, as this study examines only 30 of the 100 largest currencies based on market capitalization. Finally, as the cryptocurrency markets are relatively young, similar studies should be conducted with more trading day data available, so the results could be compared. At that point, some regulation and more academic research may be available, which can be used to answer the question: are the deviations caused by low market efficiency, psychologically influenced behaviours of investors or intentional market manipulation?

REFERENCES

Abend, Gabriel (2008). The Meaning of Theory. *Sociological Theory* 26:2, 173-199.

Aggarwal, Rajesh K. & Guojun Wu (2006). Stock Market Manipulations. *The Journal of Business* 79:4, 1915-1953.

Alexander, Gordon J. & Mark A. Peterson (2007). An Analysis of Trade-Size Clustering and its Relation to Stealth Trading. *Journal of Financial Economics* 84, 435-471.

Allen, Franklin, Lubomir Litov & Jianping Mei (2006). Large Investors, Price Manipulation, and Limits to Arbitrage: An Anatomy of Market Corners. *Review of Finance* 10:4, 645-693.

Ausloos, Marcel, Rosella Castellano & Roy Cerqueti (2016). Regularities and Discrepancies of Credit Default Swaps: A Data Science Approach Through Benford's Law. *Chaos, Solitons and Fractals* 90, 8-17.

Azad, A.S.M Sohel, Saad Azmat, Piyadasa Edirisuriya & Victor Fang (2014). Unchecked Manipulations, Price-Volume Relationship and Market Efficiency: Evidence from Emerging Markets. *Research in International Business and Finance* 30, 51-71.

Banerjee, Abhijit V. (1993). The Economics of Rumours. *Review of Economic Studies* 60:2, 309-327.

Benford, Frank (1938). The Law of Anomalous Numbers. *American Philosophical Society* 78:4, 551-572.

Bertin, William J., Khalil M. Torabzadeh & Terry L. Zivney (1996). Overreaction to Takeover Speculation. *The Quarterly Review of Economics and Finance* 36:1, 89-115.

Bjerg, Ole (2016). How is Bitcoin Money? *Theory, Culture and Society* 33:1, 53-72.

Brown, Steven D. (2016). Cryptocurrency and Criminality: The Bitcoin Opportunity. *Police Journal: Theory, Practice and Principles* 89:4, 327-339.

Brähler, Gernot, Stefan Engel, Max Göttsche & Bernhard Rauch (2011). Fact and Fiction in EU-Governmental Economic Data. *German Economic Review* 12:3: 243.255.

Buckner, H. Taylor (1965). A Theory of Rumor Transmission. *The Public Opinion Quarterly* 29:1, 54-70.

Carrera, César (2015). Tracking Exchange Rate Management in Latin America. *Review of Financial Economics* 25:1, 35-41.

Carslaw, Charles A. P. N. (1988). Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour. *The Accounting Review* 63:2, 321-327.

Chino, Theo & Ramesh Subramanian (2015). The State of Cryptocurrencies, Their Issues and Policy Interactions. *Journal of International Technology and Information Management* 24:3, Article 2. Available from World Wide Web: http://scholarworks.lib.csusb.edu/jitim/vol24/iss3/2

Cho, Wendy K. T., Brian J. Gaines (2012). Breaking the (Benford) Law. *The American Statistician* 61:3, 218-223.

CoinMarketCap [online] (2018a). Global Charts – Total Market Capitalization [Cited on 16.4.2018]. Available from World Wide Web: https://coinmarketcap.com/charts/

CoinMarketCap [online] (2018b). Top 100 Cryptocurrency Exchanges by Trading Volume [Cited on 17.12.2018]. Available from World Wide Web: https://coinmarketcap.com/rankings/exchanges/

CoinMarketCap [online] (2018c). Top 100 Cryptocurrencies by Market Capitalization [Cited on 25.10.2018]. Available from World Wide Web: https://coinmarketcap.com/

Cumming, Douglas, Sofia Johan & Dan Li (2011). Exchange Trading Rules and Stock Market Liquidity. *Journal of Financial Economics* 99:3, 651-671.

Daniel, Kent, David Hirshleifer & Avanidhar Subrahmanyam (1998). Investor Psychology and Security Market Under- and Over-Reactions. *Journal of Finance* 53, 1839-1886.

Davis, Gerald F., Michael Lounsbury & Klaus Weber (2009). Policy as Myth and Ceremony? The Global Spread of Stock Exchanges, 1980-2005. *Academy of Management Journal* 52:6, 1319-1347.

De Bondt, Werner F. M. & Richard H. Thaler (1987). Further Evidence on Investor Overreaction and Stock Market Seasonality. *The Journal of Finance* 42:3, 557-581.

Dostov, Victor & Pavel Shust (2014). Cryptocurrencies: An Unconvetional Challenge to the AML/CFT regulators? *Journal of Financial Crime* 21:3, 249-263.

Durtschi, Cindy, William Hillison & Carl Pacini (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting* 1524-5586:5, 17-34.

European Central Bank (2012). Virtual Currency Schemes. Available from World Wide Web:https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf

European Central Bank (2015). Virtual Currency Schemes – A Further Analysis. Available from the World Wide Web: https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemesen.pdf.

Fama, Eugene F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* 25:2, 383-417.

Farmer, J. Doyne & Andrew W. Lo (1999). Frontiers of Finance: Evolution and Efficient Markets. *Proceeding of the National Academy of Science of the United States of America* 96:18, 9991-9992.

Gandal, Neil, JT Hamrick, Tyler Moore & Tali Oberman (2017). Price Manipulation in the Bitcoin Ecosystem. *Journal of Monetary Economics* 000, 1-11.

Gârleanu, Nicolae & Lasse Heje Pedersen (2018). Efficiently Inefficient Markets for Assets and Asset Management. *The Journal of Finance* 73:4, 1663-1712.

Garvais, Simon, Ron Kaniel & Dan H. Mingelgrin (2001). The High-Volume Return Premium. *Journal of Finance* 56:3, 877-919.

Glaeser, Edward, Simon Johnson & Andrei Shleifner (2001). Coase Versus the Coasians. *The Quarterly Journal of Economics* 116:3, 853-899.

Glaser, Florian, Martin Haferkorn, Michael Siering, Moritz Christian Weber & Kai Zimmermann (2014). Bitcoin – Asset or Currency? Revealing Users' Hidden

Intentions. *ECIS 2014 (Tel Aviv).* Available from World Wide Web: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425247

Goldberg, Dror (2012). The Tax-foundation Theory of Fiat Money. *Economic Theory* 50:2, 489-497.

Harris, Lawrence (1991). Stock Price Clustering and Discreteness. *The Review of Financial Studies* 4:3, 389-415.

HE 167/2018 (2018). Hallituksen Esitys Eduskunnalle Laiksi Pankki- ja Maksutilien Valvontajärjestelmästä ja Eräiksi Siihen Liittyviksi Laeiksi. Available from World Wide Web: http://finlex.fi/fi/esitykset/he/2018/20180167

Hill, Theodore P. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science* 10:4, 354-363.

Ikenberry, David L. & James P. Weston (2008). Clustering in US Stock Prices after Decimalisation. *European Financial Management* 14:1, 30-54.

Keloharju, Matti & Jyrki Niskanen (2010). Earnings Cosmetics in a Tax-driven Accounting Environment: Evidence from Finnish Public Firms. *European Accounting Review* 9:3, 443-452.

Khwaja, Asim I. & Atif Mian (2005). Unchecked Intermediaries: Price Manipulation in an Emerging Stock Market. *Journal of Financial Economics* 78, 203-241.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer & Robert W. Vishny (1998). Law and Finance. *Journal of Political Economy* 106:6, 1113-1155.

Ley, Eduardo (1996). On the Peculiar Distribution of the U.S. Stock Indexes' Digits. *The American Statistician* 50:4, 311-313.

Li, Xin & Chong A. Wang (2016). The Technology and Economic Determinants of Cryptocurrency Exchange Rates: The Case of Bitcoin. *Decision Support System* 95, 49-60.

McCann, Craig (1999). Churning. *Journal of Legal Economics* 9:1, 49-68.

Morck, Randall, Bernard Yeung & Wayne Yu (2000). The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements? *Journal of Financial Economics* 58, 215-260.

Muth, John F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica* 29:3, 315-335.

Nakamoto, Satoshi (2009). Bitcoin: A Peer-to-Peer Electronic Cash System (Bitcoin whitepaper). Available from World Wide Web: https://bitcoin.org/bitcoin.pdf

Niederhoffer, Victor & M. F. M. Osborne (1966). Market Making and Reversal on the Stock Exchange. *Journal of the American Statistical Association* 61:316, 897-916.

Nigrini, Mark & Joseph T. Wells (2012). *Benford's Law – Application for Forensic Accounting, Auditing, and Fraud Detection.* Hoboken, New Jersey: John Wiley & Sons.

Newcomb, Simon (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics* 4:1, 39-40.

Pandey, Vivek & Chen Wu (2014). The Value of Bitcoin in Enhancing the Efficiency of an Investment's Portfolio. *Journal of Financial Planning* 27:9, 44-52.

Plassaras, Nicholas A. (2013). Regulating Digital Currencies: Bringin Bitcoin within the Reach of the IMF. Chicago Journal of International Law 14:1, Article 12. Available from World Wide Web: http://chicagounbound.uchicago.edu/cjil/vol14/iss1/12

Riccioni, Jessica & Roy Cerqueti (2018). Regular paths in financial markets: Investigating the Benford's law. *Chaos, Solitons and Fractals* 107, 186-194.

Thomas, Jacob. K (1989). Unusual Patterns in Reported Earnings. *The Accounting Review* 64:4, 773-787.

Verohallinto (2018). Virtuaalivaluuttojen Verotus. Available online from the World Wide Web: https://www.vero.fi/syventavat-vero-ohjeet/ohje-hakusivu/48411/virtuaalivaluuttojen-verotus/

Wallace, Wanda A. (2002). Assessing the Quality of data Used for Benchmarking and Decision-making. *The Journal of Government Financial Management* 51:3, 16-22.

Zarowin, Paul (1990). Size, Seasonality, and Stock Market Overreaction. *The Journal of Financial and Quantitative Analysis* 25:1, 113-125.

Zimmer, Zac (2017). Bitcoin and Potosí Silver: Historical Perspective on Cryptocurrency. *Technology and Culture* 58:2, 307-334.