ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialization Financial Economics

Can Twitter help forecast Bitcoin's future?

An analysis into the relevance of investors' social sentiment on Bitcoin's future market movements.

Author:	W. Masquelier
Student number:	536602wm
Thesis supervisor:	Dr. J. Lemmen
Second reader:	Dr. D. Bansraj
Finish date:	1st of August 2021

Preface

Before going deeper into the subject, I must express my gratitude to my supervisor Dr. Lemmen for his support and encouragement since the initial stages of conceptualization of the research until the final draft. Furthermore, as this thesis marks the end of my academic path, I would like to thank Erasmus Universiteit Rotterdam for the inspiring and challenging MSc in Economics and Business for which I have completed both the Behavioural Economics and Financial Economics specialisations.

Keywords:
Bitcoin, Cryptocurrencies, Sentiment analysis, Twitter, Behavioural Finance, VAR analysis

JEL Classification: B55, C32, C53, D53, G17, G41, O16, O36, Z13,

Abstract

The present research aims to assess the extent to which Twitter's activity is relevant in explaining future fluctuations of Bitcoin's market movements. Twitter's activity is expected to capture the investors' social sentiment through two different indicators: the sentiment polarity (i.e. the daily ratio of positive to negative tweets) and the daily total amount of tweets posted with the hashtag '#Bitcoin', referred in this research as TBT. The forecasting power of these two social sentiment proxies is tested relative to four financial health indicators, namely Bitcoin's price, price volatility, trading volume and liquidity risk. The results of this research highlight, notably, a significant relationship between investors' sentiment proxies and Bitcoin's future price. Furthermore, similar conclusions can be drawn with regards to Bitcoin's price volatility since both sentiment indicators were proven to be significantly correlated with Bitcoin's future volatility. However, Twitter's activity revealed to be irrelevant in predicting Bitcoin's trading volume in the subsequent days. Finally, with regards to the liquidity risk associated with holding Bitcoins, this study attributes a relevant predictive power to the TBT sentiment indicator, while denying the sentiment polarity indicator. Specifically, TBT was shown to be positively correlated with Bitcoin's market breadth and negatively correlated with its tightness. Overall, an increase in TBT therefore appears to reduce Bitcoin's future liquidity risk, as its market is expected to become tighter and its breath to become thicker. In conclusion, this research highlights the relevance of Twitter's activity as a signal for Bitcoin's future performance in terms of price, volatility and liquidity.

Table of Contents

1. Chapter I: Introduction

In 2008, the individual or the group behind the pseudonym 'Nakamoto' published one of the most influential paper of the 21st century, the very first building block of a new, promising, technology known under the name of 'Blockchain'. Indeed, Nakamoto created 'Bitcoin', the first cryptocurrency, described as a peer-to-peer electronic cash system (Nakamoto, 2008). Contrary to what one might think, Bitcoin is actually not the first digital currency. In fact, digital currencies had been commonly used within the 'online gaming' environment. For instance, the famous PC game 'World of Warcraft' used 'virtual gold' as a currency within their environment. Despite being quite successful within the gaming platform, the virtual gold currency could not achieve a spill-over into the real world due to the closed design of this currency system (Glaser et al., 2014).

During the last decade, the popular debate around Bitcoin aimed at describing the financial nature of Bitcoins. As the name indicates it, Bitcoin want to propose an alternative, decentralised, digital banking system (Nakamoto, 2008). However, in practise, it is unclear whether Bitcoin holders view the cryptocurrency as a currency, as it is aimed to be, or as an investment (Bouoiyour et Selmi, 2014; Kristoufek, 2014). Furthermore, the nature of the value of Bitcoin had also been extensively discussed as the decentralised nature of Bitcoin prevents it to be viewed as 'Fiat Money' and the digital nature of Bitcoin prevents it to be viewed as 'Commodity Money' since Bitcoins are virtual unlike gold for instance (MacDonell, 2014). Nonetheless, Yermack (2013) identifies a possible severe economic problem related to Bitcoin's design. In fact, since Bitcoin's supply is fixed, Bitcoin will ultimately generate a deflationary force on the economy, preventing an optimal investment allocation of money between the different sectors of the economy (Yermack, 2013).

Throughout the academic literature on the topic, the invention of Bitcoin is commonly compared with the invention of internet (Folkinshteyn et al., 2015). Both technologies are actually similar on several aspects, the internet was initially proposed through a 'white paper' (Berners-Lee, 1989), other similarities between the technologies are their decentralised nature, the creation of an efficient and innovative protocol and the free aspect of the technology as no licensing fees are required (Folkinshteyn et al., 2015). On top of the common characteristics between the technologies, both markets seem to behave similarly. Indeed, it appears that both the companies operating through the internet and the different cryptocurrencies were subject to financial bubbles at their respective early stages of development, partially due to the corresponding short-sale restrictions on both markets during these early stages (Ofek et Richardson, 2003).

The formation of financial bubbles, at the outset of new technologies, highlights the relevance of social sentiment in setting the price of these stocks/currencies to unrealistically high levels. Social sentiment, measuring the global enthusiasm of investors around a stock/currency, was shown to be a relevant factor in the price formation in a variety of different markets (De Choudhury et Sundaram et John et Seligmann, 2008; Qiang et Shu-e, 2009; Ranco et al., 2015). Adelaar, Chang, Lacendorfer and Lee (2003) demonstrated the relevance of text messages on people's impulse buying intents and the importance of medias to transmit these intents. This result allows to collect messages posted on those platforms, such as Twitter, in order to obtain a proxy for the social sentiment at the time. De Choudhury et al. (2008) achieved to build a model predicting stock returns, based on blog communications, with approximately 87% accuracy in predicting the direction of the returns.

Concerning the case of Bitcoin, most of the social-network discussion around Bitcoin takes place on the Twitter social platform. In fact, millions of accounts follow 'crypto influencers' such as 'Justin Son' or 'John McAfee' who in turn share their feelings and thoughts over different cryptocurrencies. One of the objectives of this research is therefore to capture the investors' sentiment towards Bitcoin, through the tweets posted on Twitter. Cryptocurrencies are characterized digital decentralised currencies, partially dependent on the internet, and their respective promotion is therefore realised through different social platforms. This feature highlights the relevance of social sentiment in the pricing of those cryptocurrencies. Social sentiment could therefore be a highly relevant factor in the formation of financial pricing bubbles arising frequently within the cryptocurrency market. Twitter is the main platform used by cryptocurrency investors to express their opinion or to exchange information which makes it a great experimental field to academically observe the evolution of social sentiment about the cryptocurrencies. The present research aims at answering the following question:

"Is the future of the Bitcoin market predictable through Twitter's activity and to which extent can investors' social sentiment predict the future of its market movements?"

Several financial aspects of Bitcoin, namely its price, volatility, liquidity and daily trading volume, will be analysed in order to determine a potentially predictive influence of investors' social sentiment. Another indicator capturing Bitcoin investors' social sentiment is the daily total amount of tweets containing the '#Bitcoin' and its potential predictive power is investigated (these tweets will be referred as ‘Total Bitcoin Tweets (TBT)’ for the remaining of this paper).

The results of this research showed that the social sentiment variables are highly relevant in predicting Bitcoin’s future financial indicators. For instance, Google Trends, the daily tweets polarity and the daily total number of tweets containing the hashtag ‘#Bitcoin’ are all significant predictors of Bitcoin’s future price. Similar conclusions can be drawn for Bitcoin’s other financial indicators, at the exception of the liquidity measures which appear to be less sensitive to fluctuations in social sentiment. The relevance of using such social sentiment variables in predicting Bitcoin’s future had been enhanced by several Granger causality tests which highlighted the causative effects of these sentiment variables on Bitcoin’s future financial indicators.

The next section provides an extensive review of the literature surrounding our topic of interest, about both the Bitcoin, as a financial market, and the practise of grasping social sentiment based on social media platforms. Afterwards, the third section describes the structure of this research through 5 distinct hypotheses. The fourth section displays the methodology, the statistical procedure, employed for the realisation of the research and the fifth section exhibits the results to the specified statistical tests. The results are then discussed throughout the sixth section. The last two sections will aim to establish the limitations of this research and its results followed by some suggestions for further research, before concluding.

2. Chapter II: Literature Review

2.1. About Bitcoin

2.1.1. Bitcoin's History

Bitcoin was launched by Nakamoto in 2008. Initially the coin only valued 5 cents. We had to wait until 2012 to see a Bitcoin being traded at 1$. After that point, prices started escalating since the Bitcoin reached 230$ per unit in March 2013. It is a possibility that this rapid increase was consequent to an economic crisis which arose in Cyprus (Detrixhe, 2013). The uncertainty around banks incited investors to invest into Bitcoin to find another haven for their money (Cohan, 2013). Later that year, in October, the legitimacy of Bitcoin as a means of exchange was questioned after the black-market website 'Silk Road', using bitcoin in the trade of illegal drugs, was shut down. However, Bitcoin was still declared as a legal currency by the US Department of Justice and prices skyrocketed over 1000$, reaching 1238$ in December. Prices then stabilised for some time before starting a slow but solid increase leading to the bubble at the end of 2017, during which the Bitcoin reached approximately 20.000$. After two years of market depression (between 2018 and early 2020) the price of Bitcoin is at one of its highest price levels ever recorded, currently valued at 57.000$, on May 6th 2021.

Bitcoin's most recent bull-run, in 2021, was driven by slightly different forces in comparison with the other bull run of 2014 and 2017. In fact, crypto-influencers' tweets became extremely relevant in influencing short-term prices of cryptocurrencies. Specifically, Elon Musk became very influential, through its tweets, in 2021, and highly contributed to Bitcoin's rise and, afterwards, drop. On January 29th he added '#Bitcoin' to his Twitter bio. In reaction, Bitcoin's price shot up from 32,200$ to 37,800$ (17% increase), in less than an hour. On February 8th, Elon Musk said Tesla had bought 1.5$ billion worth of Bitcoin which consequently generated a surge of Bitcoin's price of 17% reaching a record high at 44,220$. On March 24th, Musk announced that Tesla was accepting payments in Bitcoin and that the Bitcoins paid to TESLA will not be converted into fiat currencies. On the other side, Elon Musk also contributed to its crash as he publicly disclosed that TESLA will stop accepting Bitcoins as a payment system for the purchase of their cars because of Bitcoin's extremely energy intensive transactions, on May 13th. The Bitcoin market reacted by a 17% price drop before suffering from another drop following Musk's reaction 'Indeed' to a tweet suggesting that TESLA might sell their Bitcoin holdings.

Within the academic discussion, it is common practice to discuss the similarities between the web technology, invented in 1989 by Tim Berners-Lee to the more recent blockchain technology. On top of having been launched as an open source through a white paper published by its creator, the internet stocks also exhibited exponential growth paired with investors' irrationality who contributed to the formation of a financial bubble in the early 2000s. This financial bubble was mainly inflated by technology enthusiast and financial institutions who were heavily investing at the start of the bubble (Gezcy, Musto and Reed, 2002).

2.1.2. The economics of the Bitcoin market

The last paragraph of section 2.2.1 highlighted the similarities in behaviour between internet stocks' and cryptocurrencies' early stage market. One of the main reasons why these two relatively young markets were subject to irrationally inflated prices is due to their consequent short-sales restrictions (Ofek and Richardson, 2003). Ofek and Richardson (2003) stated that short-sale restrictions tend to push the pessimistic investor out of the market thereby leaving only optimistic investors within the market. Optimistic investors altogether contribute to an irrational escalation of prices which will eventually burst at the level at which even optimistic investors can observe the overvaluation themselves.

The cryptocurrency market also suffered from short-sales restrictions before December 2017 which is the date at which the 'Chicago Board Options Exchange' introduced Bitcoin futures contracts on its platform (CBOE). Despite futures contracts not functioning with exactly the same financial mechanics as the ones behind the short-selling strategies, future contracts still leave room for the pessimist investor to bet against inflated prices and apply pressures on the prices to force them back to their equilibrium levels. Bitcoin's inflated prices in 2013-2014 and 2017-2018 might be consequent to the high concentration of technology enthusiast and highly optimistic investors within the market.

Certain types of markets are more difficult to arbitrage than others. D'Avolio (2002) observed how arbitrageurs tend to avoid some markets which they find too risky to arbitrage. Among these stocks are the ones that are either young or unprofitable or small or experiencing extreme growth (D'Avolio, 2002). Young, small and experiencing extreme growth are three adequate adjectives to describe both the cryptocurrency market the last 10 years and the internet stocks market in the early 2000s. Wurgler and Zhuravskaya (2002) determined that what makes these stocks harder to arbitrage is the high idiosyncratic volatility of their returns. Arbitrageurs who were able to perceive the mispricing at an early stage of the financial bubble might be forced to withdraw when the mispricing is the greatest (Shleifer and Vishny, 1997).

One of the most important features of Bitcoin to understand Bitcoin's price dynamics is how do investors actually perceive Bitcoin itself. Indeed, the definition of Bitcoin describes it as a (digital) currency (Nakamoto, 2008). However, the majority of transactions made in Bitcoin isn't for the exchange of goods and services by a buyer and a seller but instead consequent to the trading of cryptocurrencies on exchange platforms (Baur and Hong and Lee, 2017). This fact highlights the speculative usage of Bitcoin. Glaser et al. (2014) observed that new investors approaching the digital currency market are not directly interested by a new alternative payment system but instead view Bitcoin and other cryptocurrencies as a financial investment. The combination of findings mentioned above indicates that the main usage of Bitcoin is speculative which makes the investors' emotions highly relevant at influencing the Bitcoin's future price.

2.1.3. Bitcoin's price drivers

In order to identify the key drivers of Bitcoin's price, it is worth to try to assess what are the financial properties of Bitcoin when considered as an investment asset. It is common practise for any investor familiar with the cryptocurrency market to hear the Bitcoin being called the 'digital gold'. One may wonder whether Bitcoin's financial properties are actually similar to the ones gold is benefitting of. First of all, both Bitcoin and gold are 'extracted' through a mining process (Meech and Gu, 2014). Despite the mining processes being substantially different, the financial structure behind is relatively similar. In fact, the cost of mining is both influenced by the input prices (electricity to mine Bitcoin or oil to mine gold) but is also influenced by technological changes arising in the mining process (Meech and Gu, 2014). Bouri, Azzi and Dyhrberg (2016) compared the behaviour of Bitcoin to the financial properties of gold and observed that, in the same way as gold, Bitcoin exhibits safe-haven properties. However, it seems that these properties disappear in periods of economic downturn for the cryptocurrency market. Indeed, Bouri et al. (2016) observed that the safe-haven properties of Bitcoin disappeared after the Bitcoin market crashed in 2014.

Yermack (2013) however found a significant discrepancy between the two assets; the uncertainty around the supply. It is greatly uncertain what quantity of gold reserves remains in the ground to be potentially mined and it is also highly uncertain what proportion of gold had already been mined. The opposite holds for Bitcoin since the total supply of the coin is fixed at 21 million Bitcoins and, to this day, 18 688 393 Bitcoin already circulates among investors. This latter fact implies that if Bitcoin becomes widely adopted, the design of the currency, having a fixed supply, will exert a deflationary force on the economy and possibly prevent further investment in the economy (Yermack, 2013). Luther and White (2014), however, think that the use of bitcoin as a currency is still possible and would allow to avoid periods of hyperinflation.

Despite Bitcoin being uncorrelated to the price of gold (Yermack 2013), other indicators might be significantly correlated with Bitcoin's price. Indeed, striking correlations were found between Bitcoin's price and the number of Google search queries containing the word 'Bitcoin' (Matta and Marchesi, 2015). Bouoiyour and Selmi (2014) use Google views to measure investors' attractiveness and also found a significant relationship between the variables. Moreover, it was discovered, by Mac Donell (2014), that the Volatility Index (VIX) is negatively correlated with the Bitcoin's price. The reasoning behind such finding is that investors switch their money towards Bitcoin during periods when they fail to obtain positive excess return in the stock market due to a low level of volatility within the stock market.

The formation of inflated prices on the cryptocurrency market can be explained by a combination of social phenomena. Investors' behavioural biases have a role to play. For instance, the bias called 'representativeness' influences the investors to base their judgement about an asset or a stock based on its recent performances only, which is particularly relevant during a period of financial bubble (Baker and Ricciardi, 2014). The trend-chasing bias also pushes the investors to over-invest into already inflated stocks. Finally, the 'Herding' theory of bubble formation describes that financial bubbles tend to arise when investors behave in imitation of other investors which explain how bubbles can be sustained and afterwards suffer from drastic decrease when it bursts (Mac Donell, 2014).

2.2. Social Sentiment

2.2.1. The relevance of social sentiment on market movements

Social sentiment is defined as the global social belief, among individual investors, about future cash flows and investment risk which cannot be justified by the information and facts concerning the market available at that time (Baker and Wurgler, 2007). One market's social sentiment highly depends on its investors' emotional intelligence. Emotional Intelligence englobes a variety of personality traits relative to the individual's character or social skills (Ameriks and Wranik and Salovey, 2009). More specifically, Salovey and Mayer (1990) describe emotional intelligence as one individual's capacity to recognise and interpret emotions and his ability to integrate them for solving complex tasks. Emotions had historically been a considerable but unobservable factor of influence on financial markets. To illustrate the impact of investors' emotions on financial markets, Edmans, Garcia and Norli (2007) observed a significant correlation between soccer scores and the stock returns on the subsequent day, but only after losses. Qiang and Shu-e (2009) studied the influence of the social sentiment, measured through indirect measures such as market turnover or growth rate of accounts, on China's stock markets. The authors observed a systematic effect of the social sentiment in forming stocks' future prices. However, they found different magnitudes of impact when comparing positive fluctuations in the social sentiment index with negative fluctuations.

2.2.2. Social sentiment analysis through social media platforms

Measuring social sentiment can be a challenge considering the immense variety of techniques and platforms to extract the necessary information from. Several researches showed the relevance of using Google Trends to proxy for the social attractiveness of a stock and its statistically significant impact on future prices (Choi Hal Varian, 2009 ; Bouoiyour and Selmi's, 2014). De Choudhury et al. (2008) used discussions blogs to build a proxy for the social sentiment through an analysis of the blogs dynamics. These blogs' dynamics, measured with variables such as 'length of message', 'number of comments', 'strength of comments' or even 'number of posts', were proven to have predictive power on future stock market movements. The researchers managed to achieve a 78% accuracy in predicting the magnitude of the movement and an 87% accuracy in predicting the direction of the movement.

Antweiler and Frank (2001) conducted one of the first researches which aimed to capture social sentiment through an analysis of the messages' bullishness posted on the Yahoo Finance message board using computational linguistics methods. The authors made three interesting findings. Firstly, the number of messages posted and trading volume appeared to have a positive relationship, in both directions. This finding is more relevant for small trades than for large ones. Secondly, the number of messages posted is correlated with the stocks' volatility. The number of messages posted is more relevant to predict market volatility than the reverse relation. Thirdly, the bullishness of messages also appears to have a predictive power on trading volume. However, this relation is stronger in the opposite direction as an increasing in trading volume subsequently increase the bullishness of messages posted afterwards.

2.2.3. *Social sentiment measured on Twitter*

Among all social platforms available to our use, Twitter constitutes a suitable field to observe the investors' social sentiment. Indeed, since Twitter's purpose is to propagate information about the users' opinion (Ye and Wu, 2010), the collection of tweets concerning a certain stock may capture the social sentiment toward that same stock. In order to do so, Pak and Paroubek (2010) implemented a methodology to use Twitter as a corpus for sentiment analysis. The researcher aimed to classify tweets into three distinct categories of emotions: positive emotions, negative emotions and neutral, objective tweets. Huberman (2010) was one of the first one to successfully use the messages posted on the Twitter platforms to predict movies' future box-office revenues. Through this research, they showed that social media platforms can serve as an effective indicator for real-world future performances.

In 2011, Bollen, Mao and Zeng affirmed having built a model to predict future movement of the Dow-Jones Index based on the mood measured on the Twitter platform. Their analysis classified all tweets among six different mood states (namely calm, alert, sure, vital, kind and happy) using the 'OpinionFinder' tool. They successfully achieved to build a model having an 87.6% accuracy in predicting the daily fluctuations of the Dow-Jones Index. However, their findings were highly discussed among academics who failed to ever reproduce such results. For instance, Kuleshov (2011) tried to reproduce the findings found by Bollen et al. (2011) but only managed to build a model predicting always upwards movements, with 60% accuracy in doing so. Lachanski and Pav (2017) also tried to replicate the results found by Bollen et al. (2011) and couldn't find any predictive power of Twitter's mood analysis on future prices of the Dow-Jones Index. They concluded that the research must have suffered of one (or a combination) of the following biases: multiple-comparison bias, data snooping bias and publication bias (Lachanski and Pav, 2017).

2.3. *Bitcoin's price forecast via Twitter*

Jermain Kaminski (2016) studied whether signals can be captured on Twitter to forecast Bitcoin future prices. The researcher realised this study over a period of 104 days only (Nov. 2013 – Mar. 2014) and collected 160.000 tweets. Kaminski classified the collected tweets into three categories: positive if the tweet contained positive word(s) (e.g. happy, good, etc..), negative if it contained negative word(s) (e.g. sad, bad, etc..) or uncertain if it contained words such as 'worry' or 'fear'. The statistical analysis led to two conclusions. First, the researcher observed that social sentiment is more a reflection of the trading volume. In other words, he concludes that social sentiment will rise in response to an increase in the trading volume. Second, a Granger causality test exhibits no statistically significant predictive power to the social sentiment index on either the close price or the intra-day return.

On the other hand, Abraham, Higdon, Nelson and Ibarra (2018) found a small but significant correlation between the social sentiment measured on the Twitter platform and Bitcoin’s future price. However, they note that this correlation becomes weaker in periods of downturn for Bitcoin. They attribute this difference to the fact that tweets concerning Bitcoin are either objective and reflect the reality or highly enthusiastic despite the market falling. Their model is still to be improved considering it is a simple linear model which does not control for search volume on Google Trends or even for tweets volume.

Philippas, Rjiba, Guesmi and Goutte (2019) implemented a research relatively similar to what is being studied in the present research. They compared two different proxies for investors’ attention through media, the daily amount of search queries on Google and the daily total number of tweets posted with the hashtag ‘#Bitcoin’. They observed that Google Trends is a better predictor of Bitcoin’s future price movements than TBT. However, they still managed to establish a significant relationship between TBT and Bitcoin’s future price and volatility. Zhu, Zhang, Wu and Zheng (2021) investigated the relationship between investors’ attention, measured through social media, and Bitcoin’s future return. They observed a significant impact of investors’ attention on Bitcoin’s future price and volatility, after performing several granger causality tests. More interestingly, they found out that the explanatory power of investors’ attention on Bitcoin’s future return and volatility can last for several weeks.

Lastly, it is worth having a look at the drivers of investors’ media attention to Bitcoin through the findings provided by Urquhart (2018). The researchers discovered that Bitcoin’s future media attention is dependent of its previously realised returns and volatility levels. If the return’s volatility increases, it is likely that the future media attention for Bitcoin will increase. Similarly, Bitcoin’s media attention is observed to be proportional to the absolute value of its return.

3. <u>Chapter III: Hypotheses</u>

The Efficient Market Hypothesis (EMH) implies that prices accurately reflect all information available to all market players. This hypothesis had always been extensively argued by economists. Nowadays, it had been proven that this hypothesis does not hold. In fact, the excess volatility in most markets could not arise within an efficient market. The same holds for financial bubbles. Investors are driven by crowd movements and they sometimes tend to follow the other investors' movement (i.e. 'Herd Theory'). As a result, mispricing on some classes of assets can last for an extended period of time. For instance, the dot-com bubble was driven by a popular excitement around the businesses working on the internet. Prices were inflated persistently due to continuous inflow of capital coming from investors recently convinced by other, more ancient, investors. The dot-com bubble showed the influence of the crowd on personal investment choices.

Therefore, the first hypothesis investigates whether the price of bitcoin is correlated with the overall sentiment around bitcoin captured by the social sentiment analysis done on Twitter. Investors social sentiment on Twitter is measured by two indicators: a linguistic analysis of the tweets and the popularity of the Bitcoin topic on Twitter (computed as the number of tweets containing '#Bitcoin'). It is likely that the price in at time t is correlated with the sentiment at the same time t. As a consequence, the research will aim to observe if investors' social sentiment, measured through Twitter, can be a predictor for future prices. As such, the investigation looks for correlation between investors' social sentiment measures at time t with the price of Bitcoin at time t+1. The volatility of these social sentiment measures might also trigger certain price movements and its potential impact over time is examined as well. The first hypothesis is stated as follow:

H1 : Bitcoin's social sentiment, measured on Twitter, is a significant predictor of the Bitcoin's future prices.

Fluctuations of the investors' sentiment towards Bitcoin probably do not only impact the price of Bitcoin but it might also affect its volatility. Indeed, as shown by Antweiler and Frank (2001), an asset's social sentiment is a relevant predictor of its future volatility. As such, if the global sentiment around Bitcoin changes drastically, we might in turn expect a corresponding decrease in the Bitcoin price. This reasoning leads us to the following hypothesis:

H2 : The volatility of the Bitcoin's return is correlated with the Bitcoin's social sentiment.

Antweiler and Frank (2001) highlighted the predictive power of the social sentiment on an asset's future trading volume. The following hypothesis therefore investigates whether these findings can be applied to the market for Bitcoins.

H3 : Bitcoin's social sentiment, measured using Twitter data, is a significant predictor of Bitcoin's future trading volume.

Akbas, Boehmer, Gnec and Petkova (2010) analysed the liquidity risk of different types of stocks and observed that growth stocks (which Bitcoin could be qualified as, to a certain extent) typically face higher liquidity risk than value stocks during periods of economic expansion. The computed social sentiment indicators might also be correlated with the investors' liquidity risk of holding Bitcoin within their portfolio. On the one hand, during a period of economic downturn, when the social sentiment around Bitcoin can be considered as relatively low, investors may be pressured to shift their wealth towards safer assets and therefore might want to liquidate their Bitcoins more aggressively, leading to a consequent higher liquidity risk. On the other hand, it could also be a possibility that a very high positive polarity of the Bitcoin social sentiment leads to a surge of investors to buy Bitcoins leading to significant variations in the relative bid-ask spreads in comparison with times when the sentiment index is more stable. The volatility of the social sentiment indicators, as well as their respective values, could be a key predictor of the liquidity risk short-term fluctuations.

Two measures of liquidity are employed in this study in order to be able to proxy for Bitcoin's liquidity risk form different perspectives. Firstly, Bitcoin's tightness liquidity aspect is investigated and is proxied through Bitcoin's bid-ask spread. Secondly, Bitcoin's resiliency is captured through the variable measuring Bitcoin's daily trading volume, in terms of BTC. Both liquidity indicators are expected to be influenced by variations in investors' social sentiment. On the one hand, a rapid increase in the social sentiment surrounding Bitcoin is likely to be reflected by a corresponding increase in Bitcoin's daily price volatility which would in turn increase Bitcoin's bid-ask spread, impacting Bitcoin's tightness. On the other hand, a high level of volatility of the social sentiment measure is likely to be correlated with relatively higher trading volumes both in terms of BTC or US Dollar, impacting Bitcoin's resiliency. Following this reasoning, it is a possibility that both of the effects mentioned above, concerning Bitcoin's tightness and resilience measures, have opposite liquidity effects. It will therefore be wise to compare them at a later stage of the analysis. The aforementioned effects can be translated into two sub-hypotheses, as such:

H4a : Bitcoin's social sentiment indicators, and their volatility, have predictive power on Bitcoin's future tightness, measured through Bitcoin's bid-ask spread.

H4b : Bitcoin's social sentiment indicators, and their volatility, have predictive power on Bitcoin's future resilience, measured through Bitcoin's trading volume (in terms of BTC).

Each one of the four hypotheses stated above can be translated into two sub-hypotheses, one of them testing the relevance of the polarity indicator (estimated through the linguistic analysis) and, the other one, evaluating the relevance of the Bitcoin hashtag's popularity on Twitter. Both indicators aim to proxy for the investor's social sentiment.

4. Chapter IV: Data

4.1. Cryptocurrencies' daily data and other controls

4.1.1. Price Data

Daily price data relative to Bitcoin were extracted from the 'Finaeon.com' online database. This database allows to collect extra variables concerning Bitcoin such as 'open price', 'closed price', '24h high price', '24h low price', 'average price', 'closed price change' and 'period change'. The data were extracted from January 2014 onwards. The data relative to both the daily trading volume in US Dollar and the daily trading volume in Bitcoin can be retrieved from the website: 'https://www.cryptocompare.com'. Additionally, Bitcoin's daily price volatility is also being considered. The website 'bitcoinity.org' provides extensive data of Bitcoin's volatility. Bitcoinity.org computes the daily price volatility as the standard deviation of all trades that occurred in the hour. Daily observations are then obtained by averaging out the 24 hourly price volatility observations of each day.

4.1.2. Measures of Liquidity

Many different measures of liquidity exist and have been analysed along the years, each one depending on core concepts and assumptions. Kyle (1985) made the distinction between three aspects of liquidity, namely depth, tightness and resiliency. Von Wyss (2004) adds 'trading time' to the list and defines them as follow. The depth is a trader's ability to buy or sell a certain quantity of an asset without having much influence on the price of this asset. The tightness measure is the ability for a market participant to buy and to sell an asset at about the same price at the same time. Lastly, the resiliency is the ability for a buyer or a seller of a certain asset to complete a large transaction with little influence on the quoted price.

With the desire of creating worldwide standards, the IMF published a working paper, in 2002, advocating for the five perspectives of liquidity, adding two different aspects of liquidity, standing as the breadth and the immediacy, to the list developed by Kyle, in 1985. The breadth liquidity indicator reflects whether the orders are numerous but also large in terms of financial value and the immediacy indicator represents the efficiency of trades, the efficiency through which a buyer can be match to a seller. It is worth nothing that the definitions and core concepts behind these liquidity measures can slightly vary among different sources and that collected data can sometimes be overlapping over several liquidity indicators.

Two of these five aspects of liquidity described by the IMF are being observed through this study, the first being the tightness measure. The tightness measure of liquidity is relative to the underlying transaction cost arising when buying or selling an asset. Such transaction cost is captured by that asset's bid-ask spread as it should be equivalent to the difference between these two prices. Bitcoin's bid-ask spread data were extracted from the website 'Bitcoinity.org'. The raw data about the bid-ask spread consists of a combination of bid-ask spread data issued from different trading platform (namely Bitfinex, Bitstamp, Btce, Cex.io, Coinbase, Gemini, Itbit, Lakebtc, Okcoin and other platforms). These different bid-ask spreads data, each dependent on one trading platform will be combined together proportionally to every platform's respective trading volume.

Furthermore, Bitcoin's liquidity will be investigated from another perspective, its breadth. The breadth measure of liquidity corresponds to the abundance, or not, of a multitude of orders both large in terms value and numerical amount. This aspect of liquidity is expected to be effectively proxied by the variable measuring Bitcoin's daily trading volume, in terms of BTC. Daily observation of the daily BTC trading volume can be retrieved from the following website: 'https://www.cryptocompare.com'.

4.1.3. Measures of social interest

The daily number of queries on Google, containing the word 'Bitcoin', highlights potential new investors' interest to invest within the cryptocurrency market. The data relative to this indicator can be freely downloaded using Google Trends. It is expected to be a relevant variable in predicting an asset future prices as confirmed by several studies such as the one conducted by Matta and Marchesi (2015). Google trends data is available on the following website: 'https://trends.google.com/'. The numerical data provided by Google Trends does not reflect anything particular in terms of absolute value. Instead, Google Trends will assign a score of 100 to the highest daily number of search queries in the requested time span and then base the other days' score in function of their relative number of queries. Google Trends' daily data accessibility is, however, not optimal as Google only provides daily data for time periods inferior to 90 days, the data being on a weekly basis otherwise. In order to overcome this issue and obtain daily data over a larger period of time, some programming was necessary. Basically, the program gathered daily data for a consequent number of 90 days periods. Since the data for every period is always comprised between 0 and 100, it was essential for the 90 days periods to be overlapping in order for the daily data to be comparable between periods. Each daily data point was then readjusted proportionally, based on the overlapping days' values in both 90 days periods. The code used to gather Google Trends daily data for the period between January 2014 and April 2021 is available in the Appendix section (Appendix 1).

4.1.4. Other controls

Many more data, relative different variables, are collected in order to control for different effects. Firstly, Mac Donell (2014) discovered a negative relationship between Bitcoin's price and the Volatility Index (VIX) of the S&P500 index. The reasoning is as follow: when the volatility of the S&P500 index is relatively low, investors tend to shift their asset towards Bitcoin in order with the aim to reach higher expected returns. Secondly, several researchers made the comparison between Bitcoin and gold financial properties and their similar behaviour (Meech and Gu, 2014; Bouri et al., 2013). As a result, it is natural to use gold price data as a control variable in the present research. Thirdly, more recent academic discussions argue that Bitcoin can constitute a hedge to the global economy and that Bitcoin prices are dependent on the world's financial health (Stensas and Nygaard, 2019). Therefore, the S&P500 index itself will be used to proxy for the world's global financial health.

4.2.Measures of Social Sentiment on Twitter

In order to obtain representative social sentiment indicators, several social sentiment measurements must be constructed and a large amount of public data must be analysed. Twitter is the place where arises most of the online talk on the topic of cryptocurrencies. The academic access to the Twitter API was granted in order to collect a very high number of historical tweets to be analysed with the aim to extract the social sentiment from this collection of tweets. Two different measures of social sentiment will be considered to compose the social sentiment index, namely the polarity of the tweets and the Total Number of Tweets posted each day with the '#Bitcoin' (TBT)

The data collection process to compute the social sentiment indicators was implemented through the programming software 'Python'. First, approximately 500 tweets were collected, daily, which must include the term 'Bitcoin' either in the tweet message or following a 'hashtag'. Over these 500 daily tweets, a selection was operated based on the language and the type of tweet. First, all tweets in another language than English were excluded from the analysis in order to facilitate the linguistic analysis of the collection of tweets. Second, the tweets which can be qualified as 'scams' or 'promotions' (such as the ones promoting free bitcoins) were also excluded from the analysis in order to minimise the chances of bias in the tweet collection. Retweets weren't excluded from the analysis as it is assumed that, on average, a Twitter user will retweet tweets reflecting his actual mood and beliefs.

4.2.1. Measure of Polarity

Once the tweets were gathered, the tone of each tweet was evaluated individually and was classified into one of these three categories: positive, negative and neutral. In order to do so, the content of each tweet was compared to two lists of words: a positive keywords list and a negative keywords list. These lists had been partially inspired by both Loughran and McDonald (2015) textual analysis in Finance (in which they provide a classification of keywords to capture the author's sentiment) paired with a collection of words proper to the cryptocurrency market, such as 'moon', for instance, which is typically used by investors to express their enthusiasm about their favourite coin. These two lists of words can be found in the appendix section (Appendix 2). The list of positive words (or 'labels') contains 96 terms or emoji for which a positive sentiment can be associated with a high degree of certainty. The list of negative labels contains 127 words and labels implying the author's negative sentiment towards Bitcoin.

In total, 15,061,208 tweets were gathered and analysed. Approximately 6,000 tweets were collected, every day, between the 1st of January 2014 and the 2nd of April 2021. The program identified 2,766,056 words belonging to the positive sentiment list and 1,224,104 words belonging to the negative sentiment list. In order for a tweet to be classified as positive (negative), it must contain strictly more positive (negative) words than the number of negative (positive) words in that same tweet. All in all, out of all these tweets analysed and all these words identified, 2,073,914 tweets were labelled as reflecting a positive sentiment and 782,771 tweets as reflecting a negative sentiment. This statistic is not surprising considering the fact that the overall tone of the tweets is bullish and reflect enthusiast investors' opinions (Antweiler et al., 2001). The polarity indicator is constructed by computing the ratio of daily positive tweets over the daily number of negative tweets. The code used to analyse the tweets is available in the appendix section (Appendix 3).

4.2.2. *The popularity of the Bitcoin hashtag*

Antweiler and al. (2001) highlighted the relevance of observing the number of queries to predict an asset's future volatility and trading volume. Total Bitcoin Tweets may therefore reflect the overall investors' enthusiasm concerning Bitcoin at a different extent than what the social sentiment polarity does. As a result, TBT should also be representative of the investors' enthusiasm towards Bitcoin. In fact, the average tone of the messages posted on social platform (such as tweets posted on Twitter) is usually bullish (Antweiler and Frank, 2001) and therefore a high number of tweets might indicate an underlying enthusiasm within the Bitcoin investors population. TBT data had been extracted from the following website: [https://bitinfocharts.com/comparison/tweets-btc.html]. Collecting these data required some basic knowledge in 'web-scraping' and the required code to obtain these data had been recently published on the 'reddit' platform. Using Google's 'inspect' function grants the access to the website's script (in java) which contains all the information available on the website, but in a raw data format. A few lines of code were then needed to extract these data, by running them directly into the website's console. These data were downloaded and transferred into a '.csv' file format using the command listed in the appendix (Appendix 4).

4.3. *Descriptive Statistics*

4.3.1. *Basic Descriptive statistics*

Before going deeper into the statistical analysis employed in this research, it is worth having a look at the descriptive statistics of the main variables which compose the studied data. Table 1 displays these statistics. The sample contains approximately 2,650 date observations, at the exceptions of the three control variables (S&P500, VIX and Gold Price) for which observations weren't recorded on the week-ends which therefore reduced the amount of data available. Concerning Bitcoin's financial indicators, the price of Bitcoin ranges from 183.01$ on the 18th of August 2015 and 61,195.30$ on the 13th of March 2021. On average, one Bitcoin is valued at 5903.63$ but its price is highly volatile and therefore has a standard deviation of 9,033$. Relative to Bitcoin's trading volume, the daily volume, measured in terms of Bitcoin, ranges between 404.38BTC on the 1st of March 2014 and 517,733BTC on 4th of November 2015, with a mean of 61,366BTC. When measured in terms of US Dollars, the maximum value was recorded recently, on the 11th of January 2021, for a total of 810 billion US Dollar of trading volume in one day! Bitcoin's bid-ask spread is on average equivalent to 0.1% of its price.

Table 1: Descriptive Statistics
Table 1 shows the basic descriptive statistics for all the variables included in the dataset. This annotation provides a brief description of the variables and of the five columns. The first column reports each variable's total count, the second represents each variable's mean, the third is relative to its standard deviation. The fourth and fifth column indicate the range of observations and are relative to each variable's minimum and maximum values, respectively. 'High' is the Bitcoin's daily highest price, 'Low' is its daily lowest price, 'Close' refers to its daily closing price, 'BidAskSpread' is computed as Bitcoin's daily bid-ask spread averaged over several platforms, 'VolumeBTC' is Bitcoin's daily trading volume measured with the BTC unit while 'Volume' is measured with the US Dollar unit, 'VIX' represent the daily observations of the S&P500's indexed volatility, 'SP500' report the index's daily price and 'PGold' reports gold's daily price, 'GoogleTrends' measures the daily relative number of search queries concerning Bitcoin, 'pos words' counts the daily number of positive words or labels identified while 'neg words' counts the daily number of negative ones, 'total positive tweets' indicates the daily number of positive tweets while 'total negative tweets' indicates the daily number of negative ones, 'tweet sum' account for the daily number of tweets analysed, 'Polarity' is the daily positive to negative tweets ratio, 'DailyReturns' reports Bitcoin's daily returns and 'TBT' its daily total number of tweets posted.

Variable	Obs	Mean	Std. Dev.	Min	Max
High	2508	6202.796	9484.997	211.03	61795.8
Low	2508	5829.022	8813.692	100	59113.7
Close	2508	6043.586	9213.756	183.01	61195.3
BidAskSpread	2508	.094	.075	.015	.635
VolumeBTC	2508	60935.769	49980.862	446.84	517733.91
Volume	2508	3.949e+08	7.224e+08	210112.72	8.101e+09
VIX	1729	17.277	7.887	9.14	82.69
SP500	1729	4994.555	1240.55	3276.04	8335.91
PGold	1731	1357.037	222.315	1049.4	2067.15
GoogleTrends	2508	257.223	355.573	33.623	4488.284
pos words	2508	1042.88	408.828	272	3611
neg words	2508	457.083	180.938	125	1783
total positive tweets	2508	781.869	312.025	205	3230
total negative tweets	2508	292.031	131.136	52	1412
tweet sum	2508	5662.051	620.771	3847	7940
Polarity	2508	2.96	1.237	.43	11.788
DailyReturns	2507	.003	.039	-.392	.268
TBT	2508	35081.23	23613.778	7300	212923

The second part of Table 1 shows the descriptive statistics relative to the social sentiment variables. First, the number of Google search queries, which can be seen as a control variable in this study as its significance had already been demonstrated by Bouoiyour and Selmi (2014), among others. The Google Trends variable reached its maximum on the 22nd of December 2017, approximately at the time when Bitcoin reached 20,000$. This statistic highlights the importance of Bitcoin's 2017-2018 bull run to attract investors attention and therefore reach a high level of visibility. The Total amount of tweets concerning Bitcoin, however, was more impacted by Bitcoin's more recent bull run as it reached its maximum on January 9th 2021 with a total of approximately 212,000 tweets within 24 hours. The polarity sentiment indicator has a mean value of 2.94 which means that on average there are 2.94 times more positive tweets identified during the course of the day, in comparison with the number of negative tweets identified. The polarity indicator is however very volatile as its standard deviation equals 1.23.

4.3.2. Graphical representation of the variables

The first figure represents the evolution of Bitcoin's financial indicators over time. The chart relative to Bitcoin's price (Graph 1) clearly highlights the two bullish periods of 2017 and 2021, the boom of 2014 also being observable. The price of Bitcoin tends to increase over time but also exhibit very high levels of volatility, unsurprisingly. The daily trading volume, measured in US Dollars, seem to follow a corresponding movement along time (Graph 2). However, we can see that Bitcoin's volume is more volatile than its price as the volume line graph experiences more radical changed between two closely related points in time. When measured in terms of Bitcoin (Graph 3), the volume exhibits an even higher level of volatility and seem to stabilise over time as Bitcoin's market integration advances. Indeed, the BTC daily trading volume reached its maximum value in 2015 and the recent bull run of 2021 only accounts for approximately half of the volume recorded on that day. Bitcoin's breadth therefore seems to stabilise as time passes by.

Bitcoin's price volatility (Graph 4) had been greatly impacted by the recent bull run which is not surprising considering the rapid escalation of prices experienced by Bitcoin. It is worth noting that we observe such distribution of observations because it is the price volatility of Bitcoin that is being observed and not the volatility of its daily returns. The price volatility of an asset is exponentially correlated to that same asset's price and this conclusion is clearly observable when comparing Graph 1 to Graph 4. On the other hand, looking at the bid-ask spread, it appears it's the liquidity indicator's volatility decreases over time and seems to stabilise as the market matures. In fact, in 2014, when Bitcoin was only valued a few hundreds of dollars, its bid-ask spread could range around 20% of Bitcoin's value. As time passes by and as the crypto market matures, Bitcoin's bid-ask spread substantially dropped over time to reach levels below 5% since 2019. Similar conclusions can be drawn, but at a lesser extent, from the distribution of the daily (BTC) trading volume variable which was most volatile before 2018.

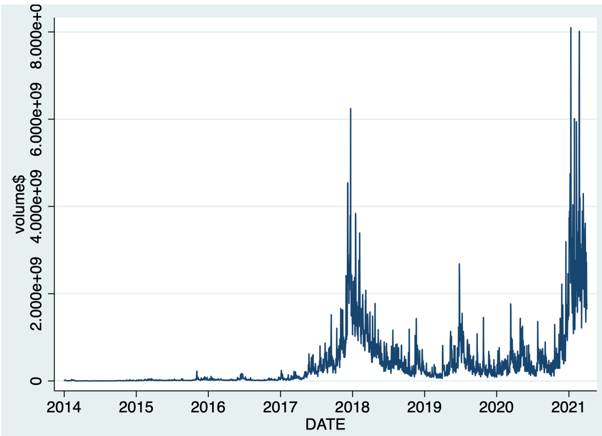
As a result, it might be possible that the two liquidity indicators, of the market's tightness and breadth, exhibit opposing liquidity implications. In fact, as it is going to be observed in the next section, both variables appear to be positively correlated meaning that when the market becomes tighter, more liquid, as the bid-ask spread decreases, it also experiences a drop in its breadth since the Trading volume is likely fall as well. Both liquidity effects will be disentangled further in the analysis.

Lastly, the daily returns seem to be relatively proportional between negative and positive returns, with a peak in positive returns during the bull run of 2017.

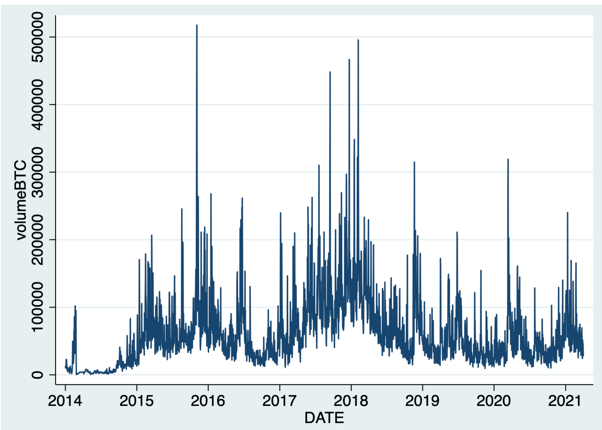
Figure 1: Graphical representation of the evolution of Bitcoin's financial indicators over time:
Graph 1 represent Bitcoin's price evolution between 2014 to 2021. Graph 2 shows the evolution over time of Bitcoin's daily trading volume, measured in US Dollars. Graph 3 represents the other liquidity measure of breadth, Bitcoin's daily trading volume, but this time in terms of BTC. Graph 4 pictures the evolution of Bitcoin's daily price volatility over time. Graph 5 is relative to Bitcoin's bid-ask spread (in proportion of its price) time evolution. Finally, Graph 6 illustrates the evolution of Bitcoin's daily return over time.

Graph 1:

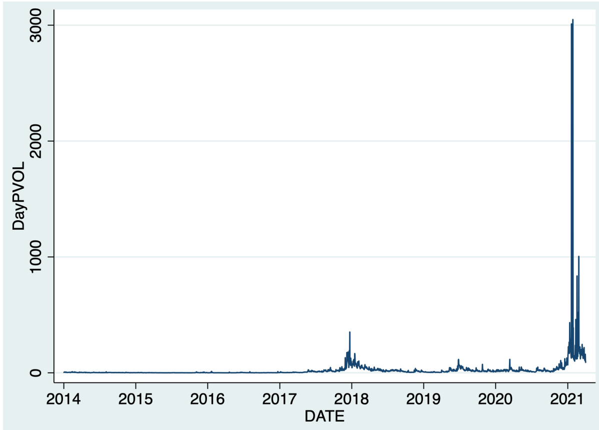
Graph 2:



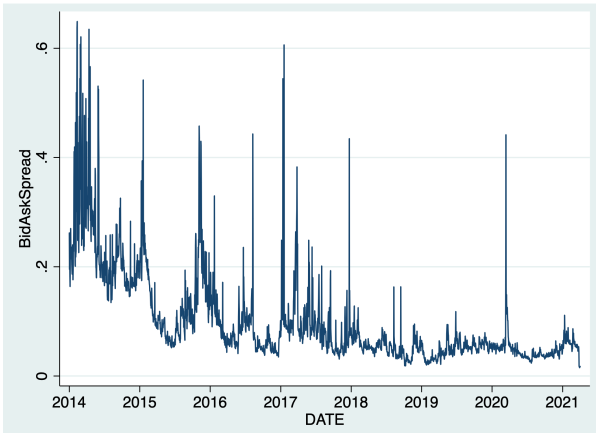
Graph 3:



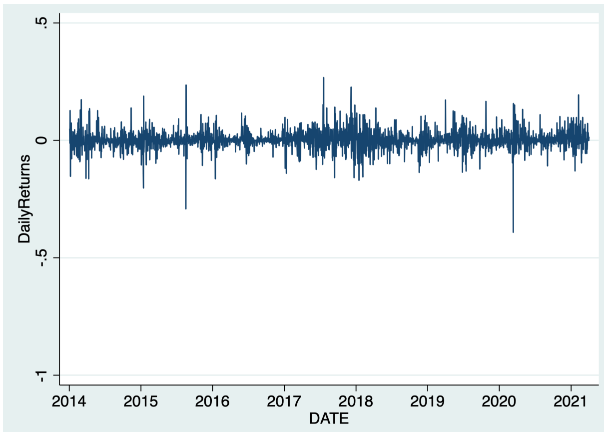
Graph 4:



Graph 5:



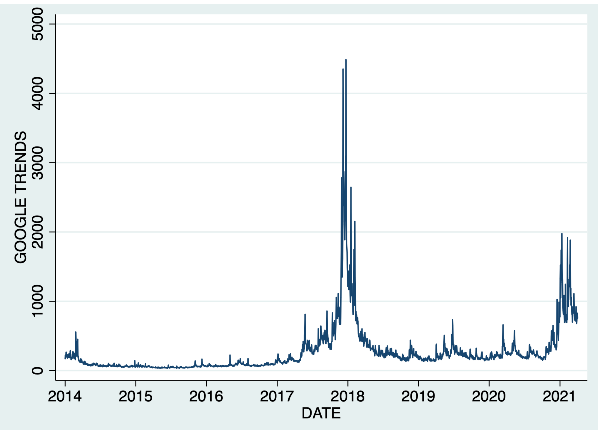
Graph 6:



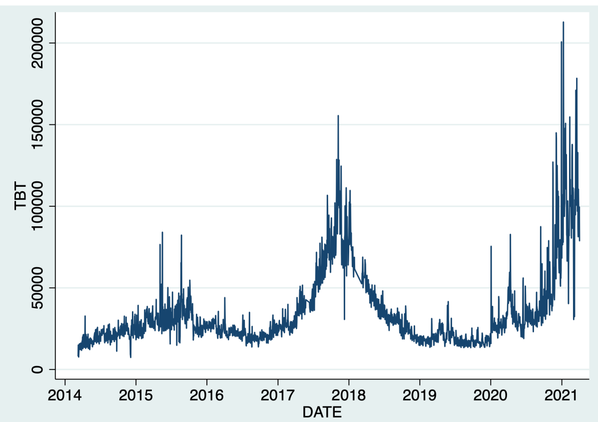
The second set of graphs illustrate the line charts of the socially related variables. There are a few comparisons and observations to be made. First, the comparison between Google Trends and TBT shows, on the one hand, that the daily total number of tweets (Graph 2) is more volatile than the daily number of Google search queries (Graph 1). On the other hand, TBT appears to follow a distribution closer to Bitcoin's price movements. Google Trends' chart clearly shows the moment at which Bitcoin attracted the world's attention and became famous, during the 2017 bull run, whereas it had been less impacted by the most recent bull run, in 2021 (probably because the world was already more aware of Bitcoin's existence). The polarity sentiment indicator (Graph 3) is highly volatile especially before 2017 when Bitcoin was not globally accepted as it had become since. In fact, the ratio fluctuates from approximately 1 up to more than 10 in only a few days, as it can be seen in 2015 or 2020, notably. Lastly, Graph 4 shows the results of the tweets' analysis by representing the daily total number of positive (blue line) and negative (red line) words identified by the algorithm.

Figure 1: Graphical representation of the evolution of Bitcoin's financial indicators over time:
Graph 1 represent Google Trend's evolution between 2014 to 2021. Graph 2 shows the evolution over time of the daily total number of tweets posted each day containing the hashtag '#Bitcoin'. Graph 3 represents the other sentiment indicator, the tweets' polarity, computed as the ratio of the total daily number positive tweets over the corresponding daily total number of negative tweets. Graph 4 pictures the daily counts of positive and negative words among all the tweets analysed, the blue line and the red line stand for the daily number of positive and negative words, respectively.

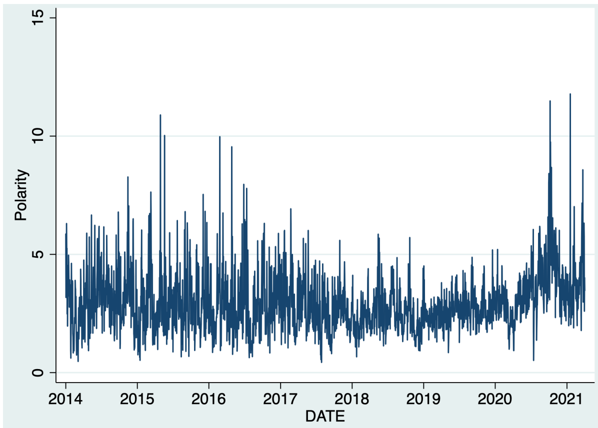
Graph 1:



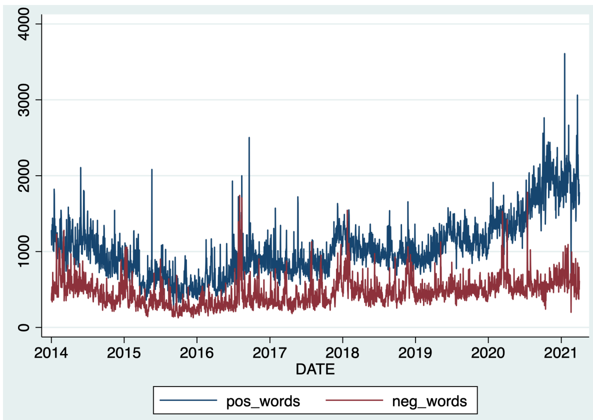
Graph 2:



Graph 3:



Graph 4:



4.3.3. *Correlation Matrix*

This last section of the descriptive statistics sub-chapter analyses the correlation matrix between the main variables of interest. This correlation matrix estimates correlations between a set of different variables at the same time t, no lags involved. The correlation matrix is displayed in the appendix section, registered as Table 2.

This section aims to observe real time correlations between the 5 variables used to test the hypotheses (Price, Volatility, Volume (in $), Volume (in BTC) and the bid-ask spread) with the different measures of social sentiment (TBT, Polarity or Google Trends). Firstly, the price variable is being observed. Most of the correlation factors are positive (at the exception of the bid-ask spread) indicating a positive relationship between Bitcoin's price and the other variables, on average. Out of all the sentiment-related variables, Google Trends, TBT, and the positivity ratio have the highest correlation factors, approximately equal to 0.6. The polarity measure, however, only has a correlation factor of 0.18. This relatively low correlation might be due to the dominance of positivism among the tweets studied and would therefore be the consequence of the investors' positive bias. The dominance of positivism might prevent the polarity indicator to positively correlate with Bitcoin's at times when the market is bearish. The price volatility variable seems to also be positively correlated with most of the variables, at the exception of the bid-ask spread measure and Bitcoin's daily return. Concerning the social sentiment indicators, TBT and Google Trends have once again the highest correlation factors of 0.366. The polarity indicator has now a correlation value below 2% as the factor equals 0.0175.

Next, we will compare the two variables quantifying the daily trading volume of Bitcoin both in terms of Bitcoins and US Dollars. On the one hand, daily trading volume, measured in US Dollars, appears to be positively correlated with all other variables at the exception of the tightness liquidity indicator, Bitcoin's bid-ask spread. Among these positive correlations, Google Trends seems to be the strongest with 82% correlation, followed by the TBT indicator with 66% correlation. The polarity indicator only has a correlation of 0.03. On the other hand, when the daily volume is measured with the BTC unit, the correlation factors change substantially. The polarity indicator (-0.17), as well as the proportion of positive tweets (-0.06), are both negatively correlated to the daily BTC trading volume. Google Trends and TBT once again report the highest (positive) correlation with correlation factors of 0.46 and 0.30 respectively.

Lastly, the bid-ask spread, which is used to proxy for the tightness of Bitcoin's market in the present research, exhibits negative relationships with all the other variables. In fact, as it was shown through the charts of the previous section, the bid-ask spread is the only variable which appears to be decreasing over time, as Bitcoin's market matures. It is therefore not a surprise to observe negative correlations between the bid-ask spread (measured in proportion of the price) and all the other variables which were shown to be increasing over time. Looking at the magnitude, it seems that the TBT variable is most negatively correlated through a coefficient of -0.206. This negative correlation implies that, if Bitcoin's bid-ask spread experiences a shock, the TBT sentiment indicator is likely to reflect that shock in the opposite direction with a magnitude of 20.6% of the shock faced by the bid-ask spread.

4.3.4. Further discussion on the data assumptions

One of the main assumptions, and conditions, when performing a regression analysis is the normality of the data. Indeed, the error term of an OLS regression are usually assumed to be normally distributed around its mean value. Normally distributed data are assumed to benefit from a higher degree of statistical power and to decrease the probability of obtaining biased results, in comparison with non-normally distributed data.

All the data relative to all the variables in the present study revealed to be non-normally distributed. Such results are not that surprising considering the large amount of observations available for each variable and the fact that the Bitcoin's market is highly volatile especially during the market's periods of financial expansion. Such events are likely to consequently generate a certain degree of skewness in the distribution of the price, volatility or volume and other financial indicators' observations. Specifically, all the distributions of both the independent variables (i.e. price, volatility, volume ($), bid-ask spread and volume (BTC)) and the social sentiment variables (TBT and Polarity) failed to be determined as normally distributed by both the skewness and Kurtosis tests. Moreover, both normality tests reported a test result of '0.0000' concerning the distributions of all the variables of interest.

At the second stage, the skewness and Kurtosis tests had been repeated but this time relative to the logarithmic transformation of each variable. The exact same results were reported by both tests at the exception of the bid-ask spread variable whose normality test result slightly increased. It can therefore be concluded that the logarithmic transformations of the variables are not efficient enough to normalize the distribution of the data observations.

Concerning the distribution of the VAR residuals, similar conclusion can be drawn. Several Jarque-Bera tests were conducted in order to assess whether the VAR residuals were normally distributed around 0. The Jarque-Bera tests relative to all five VAR models revealed that the residuals are not normally distributed. In parallel, the same VAR models, but this time involving the logarithmic transformations of every variables, were ran to observe whether using these logarithmic transformations would be associated with residuals more normally distributed. Once again, all five Jarque-Bera tests reported a value of '0.0000' indicating clear non-normal distributions of the residuals. In order to interpret the coefficients' magnitude in a more intuitive manner, the logarithmic transformations of the variables will therefore be disregarded from the statistical analysis as they failed to normalise the variables' distributions.

Despite normally distributed data being preferred, non-normally distributed can, however, still be as efficient to identify shocks. Most empirical researchers working with big data are relatively used to deal with non-normally distributed data as this is a typically reoccurring issue. It is therefore common practice to perform VAR analysis depending on non-normally distributed data. Some scientists, such as Lanne and Lutkepohl (2010), argue that, when implementing a Vector Autoregressive model, non-normally distributed data are as effective to identify shocks and impulse responses. Nonetheless, some restrictions (such as about the variables' cointegration) sometimes need to be implemented in order to obtain unbiased estimates. In conclusion, the use of the present dataset, involving non-normally distributed variables, should not be too much of a concern but the non-normality of the data distribution must be acknowledged as one of the principal limitations of this research.

5. Chapter V: Methodology

5.1. *The predictive power of investors' social sentiment*

This section will aim to describe the statistical procedure to be followed in order to assess whether or not investors' social sentiment has predictive power over Bitcoin's future indicators. We will focus on Bitcoin's future price, volatility and volume. In order to make such predictions, different autoregressive models will be implemented, one for each single indicator. More specifically, Vector Autoregressive (VAR) models will be designed and run through STATA in order to assess the relevance of these predictors. Prior to performing the statistical analysis, the optimal number of lags to be used in the VAR model must be determined. Too many lags are likely to considerably increase the degree of freedom of the model(s) and the probability to face the multicollinearity bias whereas too few lags are likely to generate specification error. The optimal number of lags is selected by taking the average optimal number suggested by three independent criteria, namely the BIC, AIC and HQIC. Further information on this topic can be found in the first section of Chapter VI. The first hypothesis investigates whether investors' social sentiment, composed of two different variables, can significantly forecast Bitcoin's future prices. The AR model regresses the price of Bitcoin at time (t) in function of the price of Bitcoin at time (t-h), the social sentiment variables, Polarity and TBT, and other control variables (i.e. Gold Price, VIX, S&P500 and Google Trends). The VAR model can be written as follow:

$$\hat{P}_t = \alpha + \sum_{h=1}^{h} \beta_{1,h} * P_{t-h} + \sum_{h=1}^{h} \beta_{2,h} * Polarity_{t-h} + \sum_{h=1}^{h} \beta_{3,h} * TBT_{t-h} + \sum_{h=1}^{h} \beta_{4,h} * Vol(Pol.)_{t-h} + \sum_{h=1}^{h} \beta_{5,h} * Vol(TBT)_{t-h} + \sum_{h=1}^{h} \beta_{\gamma,h} * Controls_{t-h} + \varepsilon_t \qquad (1)$$

Where: P_t is the price at time t, α represents the constant, h is the time-horizon (i.e. the number of lags used in the VAR model) $\beta_{1,h}$ is the corresponding coefficient for the h^{th} lag of the price, Polarity is the ratio of daily positive tweets over daily negative tweets and TBT is the Total amount of tweets concerning Bitcoin. Vol(Pol.) and Vol(TBT) are variables measuring the weekly volatility of the Polarity and TBT indicators respectively and ε is an error term.

Next, the second hypothesis aims to discover whether the two measures of social sentiment can significantly predict Bitcoin's volatility during the subsequent period. Therefore, a similar autoregressive model is implemented but this time using Bitcoin's daily price volatility as the dependant variable. The specification is as follow:

$$\hat{V}_t = \alpha + \sum_{h=1}^{h} \beta_{1,h} * V_{t-h} + \sum_{h=1}^{h} \beta_{2,h} * Polarity_{t-h} + \sum_{h=1}^{h} \beta_{3,h} * TBT_{t-h} + \sum_{h=1}^{h} \beta_{4,h} * Vol(Pol.)_{t-h} + \sum_{h=1}^{h} \beta_{5,h} * Vol(TBT)_{t-h} + \sum_{h=1}^{h} \beta_{\gamma,h} * Controls_{t-h} + \varepsilon_t \quad (2)$$

Where V_t stands for Bitcoin's volatility, along time.

Similarly, the third hypothesis investigates whether an equivalent relationship can be established between Bitcoin's daily Total Trading Volume (TTV, measured in US Dollars) at time t and the social sentiment at time t-h. The VAR model is specified below.

$$\widehat{TTV}_t = \alpha + \sum_{h=1}^{h} \beta_{1,h} * TTV_{t-h} + \sum_{h=1}^{h} \beta_{2,h} * Polarity_{t-h} + \sum_{h=1}^{h} \beta_{3,h} * TBT_{t-h} + \sum_{h=1}^{h} \beta_{4,h} * Vol(Pol.)_{t-h} + \sum_{h=1}^{h} \beta_{5,h} * Vol(TBT)_{t-h} + \sum_{h=1}^{h} \beta_{\gamma,h} * Controls_{t-h} + \varepsilon_t \quad (3)$$

5.2.Bitcoin's liquidity risk

The liquidity of Bitcoin may be affected by many exogeneous factors. Indeed, periods of high volatility typically experience bigger bid-ask spread and therefore represent a higher liquidity risk to the eyes of investors. The aim of this hypothesis is therefore to observe whether investors' social sentiment, measured through Twitter, can reveal to be a useful tool to forecast Bitcoin's future. Two different perspectives of liquidity are considered, the tightness and the breadth of the market for Bitcoins. Firstly, the bid-ask spread tightness liquidity indicator is analysed by the first VAR model, below. Secondly, the daily trading volume, expressed in BTC, captures, to a certain extent, Bitcoin's market breadth, reflecting another liquidity aspect. The model specifications are the following:

$$\widehat{BidAskSpread}_t = \alpha + \sum_{h=1}^{h} \beta_{1,h} * BidAskSpread_{t-h} + \sum_{h=1}^{h} \beta_{2,h} * Polarity_{t-h} + \sum_{h=1}^{h} \beta_{3,h} * TBT_{t-h} + \sum_{h=1}^{h} \beta_{4,h} * Vol(Pol.)_{t-h} + \sum_{h=1}^{h} \beta_{5,h} * Vol(TBT)_{t-h} + \sum_{h=1}^{h} \beta_{\gamma,h} * Controls_{t-h} + \varepsilon_t \quad (4.1)$$

$$\widehat{VolumeBTC}_t = \alpha + \sum_{h=1}^{h} \beta_{1,h} * VolumeBTC_{t-h} + \sum_{h=1}^{h} \beta_{2,h} * Polarity_{t-h} + \sum_{h=1}^{h} \beta_{3,h} * TBT_{t-h} + \sum_{h=1}^{h} \beta_{4,h} * Vol(Pol.)_{t-h} + \sum_{h=1}^{h} \beta_{5,h} * Vol(TBT)_{t-h} + \sum_{h=1}^{h} \beta_{\gamma,h} * Controls_{t-h} + \varepsilon_t \quad (4.2)$$

5.3. Granger Causality Tests

To test the causative nature of the relationships between the social sentiment variables and Bitcoin's future financial indicators, a Granger causality test is performed following every single VAR model presented above.

In order to test whether the social sentiment variables 'Granger cause' future variations in Bitcoin's price, volatility or volume, the following procedure must be applied and is computed through the 'Granger causality test'. The Granger causality tests are being implemented for each aforementioned VAR models. Firstly, a regression, called the 'restricted regression' (5), regresses Bitcoin's financial indicator in function of its own respective lagged values, excluding the social sentiment variables. From the result of the first regression, the residual sum of squares (RSS) is recorded as the 'restricted sum of squares' (RSS_R). Secondly, the same regression is repeated (6) but this time including the variable(s) relative to the social sentiment indicators, measured through Twitter. The same procedure is therefore repeated and the model's residual sum of squares is recorded as the 'unrestricted sum of squares' (RSS_U). This procedure is applied for each single relationship between Bitcoin's financial indicators and the social sentiment variables as well as their respective volatility. Equation 5 is relative to the aforementioned restricted regressions and is applied to each one of Bitcoin's financial indicators individually. Equation 6, on the other hand, is relative to the unrestricted regressions which therefore involve the social sentiment variables and their volatility.

$$BFI_t = \alpha + \sum_{i=1}^{k} \beta_j * BFI_{t-i} + \varepsilon_t \quad (5)$$

$$BFI_t = \alpha + \sum_{i=1}^{k} \beta_j * BFI_{t-i} + \sum_{i=1}^{k} \lambda_j * SSV_{t-j} + \varepsilon_t \quad (6)$$

Where BFI stands for 'Bitcoin's Financial Indicator' which could either be its price, its volatility, its daily trading volume measured either in US Dollars or in BTC or its bid-ask spread. SSV stands for 'Social Sentiment Variable(s)' representing the Polarity or TBT sentiment indicators or their respective volatility.

For each one of the relationships between Bitcoin's Financial Indicators and the Social Sentiment variables, a Granger causality test is performed using the restricted and unrestricted sum of squares obtained through their respective restricted and unrestricted regressions, involving the variables of interest to evaluate the causative effects of the relationship. These two different sums of squares measures are used to compute the F-statistic on which the Granger causality test is based. The F-statistic is computed as follow:

$$F = \frac{RSS_R - \frac{RSS_U}{k}}{\frac{RSS_U}{n - 2k - 1}} \quad (7)$$

The null hypothesis of the Granger causality test is that there exists no causative relationship between the two variables. In order to observe whether this null hypothesis can be rejected, or not, we must compare the obtained F-statistic to its critical value at a given confidence level. If the F-statistic exceeds the corresponding critical value, it can therefore be concluded that variation in the social sentiment variable ‘Granger causes’ fluctuations in the future values of Bitcoin’s financial indicator.

6. Chapter VI: Results

6.1. Determining the optimal number of lags

The first step in order to implement a vector autoregressive statistical analysis is to choose the most adequate number of lags to be used in the models. There exists a trade-off between the benefits and disadvantages of using few or many lags. On the one hand, if too few lags are being used, the model is very likely to suffer from specification errors. On the other hand, if too many lags are being used, the model will lose degrees of freedom which in turn decrease the model's propensity to observe statistically significant coefficients. Furthermore, too many lags increase also the probability for the model to suffer from multicollinearity.

As a result, the only way to minimise these potential biases is to choose the optimal number of lags using the information criterion BIC, AIC and HQIC. These criteria were computed for each one of the five models relative to the first, second, third and both fourth hypotheses. In order to be consistent all along the thesis, the number of lags used in the VAR models will stay the same among all specifications. The computed criteria are displayed in Table 3, respective to each model individually. For each one of the criteria, the optimum number of lags is the one who returned the smallest value, this optimum is indicated by the symbol ' * '.

Both the first and the second model, regressing the price of Bitcoin and its volatility, are optimised by using 3 lags, if we follow the suggestion of the BIC criterion. The HQIC criterion seems to prefer the use of 5 lags while the AIC criterion suggests to use 4 lags. Relative to the third model, regressing Bitcoin's trading volume, both the AIC and the HQIC criteria are minimised by using a fifth-order VAR model while the BIC criterion suggests a fourth-order VAR. Model 4.1 has the best fit when the VAR model includes three lags, relative to the BIC criterion, and four lags if optimised using the AIC criterion. Once again, the HQIC criterion indicates that five lags are ideal. Model 4.2, regressing Bitcoin's trading BTC volume, should include four lags in order to be optimised by the BIC criterion. The HQIC criterion proposes, one more time, a fifth-order VAR model and the AIC criterion suggests a fifth-order as well.

All in all, the BIC, HQIC and AIC criteria propose different optimums among the specifications. If all these optimums are average out, we obtain an optimal number of lags of 4.26. The VAR models will therefore be implemented using 4 lags as it is the closest integer from the optimums average. The same number of lags will be used along all specifications in order to maximise the comparability of the results.

6.2. First Hypothesis: Social sentiment and Bitcoin's price

Table 4 presents the results relative to the first hypothesis. These results aim to answer the question, whether or not, Twitter's social sentiment indicators can predict Bitcoin's future prices. The two indicators of social sentiment on Twitter stand as the polarity of the tweets measured as the ratio of the daily total number of positive tweets over the daily total number of negative tweets. Both the value of these indicators and their volatility are observed in order to establish a predictive influence of the social sentiment over Bitcoin's financial indicators, in this case Bitcoin's price.

First of all, the polarity indicator reflects mixed signals vis-à-vis of the significance of its predictive influence. In fact, only the third lag is significant as the first, second and fourth have p-values exceeding the 0.1 threshold. The third lag of the polarity sentiment indicator is however significant at the 1% significance level. Unsurprisingly, its coefficient indicates a positive relationship between Bitcoin's price and the tweets polarity. On average the magnitude of the polarity's predictive power account for 55.5$ multiplicated by the polarity score, ceteris paribus. This effect is expected to be observable only 3 days following Twitter activity. The volatility of the polarity indicator, however, does not significantly influence Bitcoin's future prices as the p-values of all four lags are over the 0.1 threshold of significance.

Figure 3 displays the Impulse Response Function (IRF) of Bitcoin's price relative to a one unit increase in the tweets' sentiment polarity. Despite exhibiting a period of negative correlation around the second lagged-value, the polarity indicator seems to be positively correlated with the price's future movements, especially through its third lagged-value as observed in the VAR table of results. It can be deducted from figure 3 that a one unit increase in the daily tweets' sentiment polarity will, on average, generate slightly less than 50$ increase in Bitcoin's price within three to four days. Overall, the tweets' sentiment polarity is a meaningful indicator in order to estimate future variation of Bitcoin's price in the next few days.

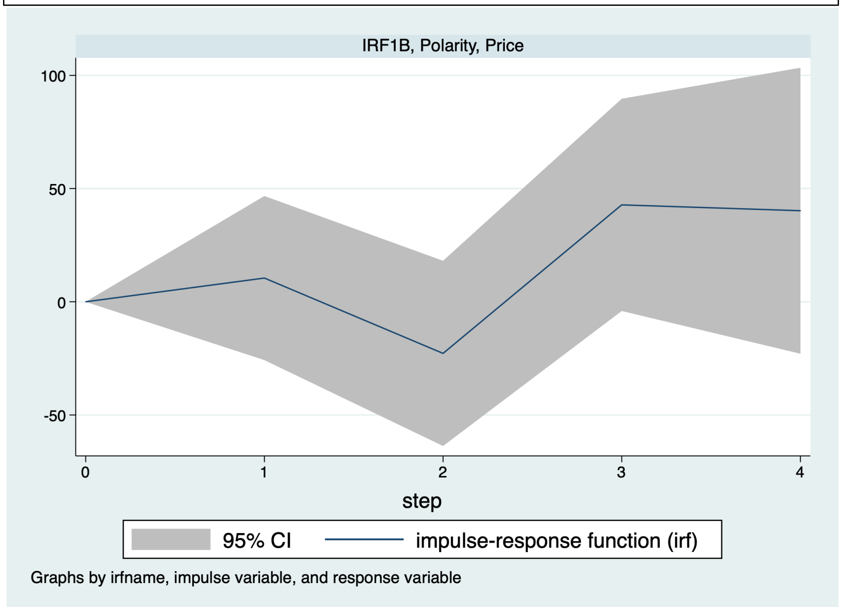
Secondly, the other sentiment indicator, accounting for the total number of tweets posted with the '#Bitcoin', also reveals promising results. In fact, the TBT sentiment indicator predict Bitcoin's future prices significantly for its three first lags. The two first lags are significant at the 1% significance level and the third lag at the 10% significance level. What could be seen as a surprise is change in the predictive effect among the different TBT lags. Indeed, the TBT sentiment indicator's first and third lags are negatively correlated with Bitcoin's price whereas the second lag is positively correlated (with the highest magnitude) with Bitcoin's price.

Table 4: VAR results for Bitcoin's Price

Table 4 exhibits the VAR results relative to the first hypothesis. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observation's average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Price						
Price						
L1.	0.700	0.044	16.010	0.000	0.615	0.786
L2.	0.011	0.061	0.180	0.855	-0.108	0.131
L3.	0.635	0.061	10.400	0.000	0.515	0.755
L4.	-0.331	0.041	-8.010	0.000	-0.412	-0.250
Polarity						
L1.	9.181	18.450	0.500	0.619	-26.979	45.342
L2.	-25.774	19.140	-1.350	0.178	-63.289	11.740
L3.	55.504	19.758	2.810	0.005	16.778	94.229
L4.	-18.085	17.053	-1.060	0.289	-51.509	15.338
TBT						
L1.	-0.006	0.002	-2.760	0.006	-0.010	-0.002
L2.	0.013	0.003	3.910	0.000	0.006	0.019
L3.	-0.008	0.004	-1.950	0.051	-0.017	0.000
L4.	-0.000	0.003	-0.030	0.976	-0.005	0.005
Polarity1weekvol						
L1.	20.578	92.056	0.220	0.823	-159.847	201.004
L2.	40.733	115.850	0.350	0.725	-186.328	267.795
L3.	-19.253	121.732	-0.160	0.874	-257.844	219.338
L4.	-25.132	99.926	-0.250	0.801	-220.984	170.720
TBT1weekvol						
L1.	-0.022	0.010	-2.260	0.024	-0.041	-0.003
L2.	0.031	0.019	1.640	0.102	-0.006	0.067
L3.	0.025	0.028	0.870	0.386	-0.031	0.080
L4.	-0.026	0.017	-1.540	0.124	-0.059	0.007
VIX						
L1.	-27.550	20.709	-1.330	0.183	-68.138	13.039
L2.	28.022	26.208	1.070	0.285	-23.345	79.388
L3.	-2.394	27.415	-0.090	0.930	-56.128	51.339
L4.	-0.819	19.150	-0.040	0.966	-38.352	36.714
S&P500						
L1.	-0.566	0.738	-0.770	0.443	-2.013	0.880
L2.	0.413	0.903	0.460	0.647	-1.357	2.184
L3.	0.311	0.782	0.400	0.690	-1.221	1.843
L4.	-0.168	0.532	-0.320	0.752	-1.210	0.874
Google Trends						
L1.	0.558	0.138	4.040	0.000	0.288	0.829
L2.	-0.437	0.257	-1.700	0.089	-0.941	0.067
L3.	-0.338	0.425	-0.790	0.427	-1.172	0.496
L4.	0.072	0.319	0.220	0.823	-0.554	0.697
P(Gold)						
L1.	1.807	1.295	1.390	0.163	-0.732	4.346
L2.	-2.017	2.006	-1.010	0.315	-5.949	1.915
L3.	-0.902	2.034	-0.440	0.658	-4.888	3.085
L4.	1.197	1.459	0.820	0.412	-1.664	4.057
_cons	-87.632	161.049	-0.540	0.586	-403.282	228.018

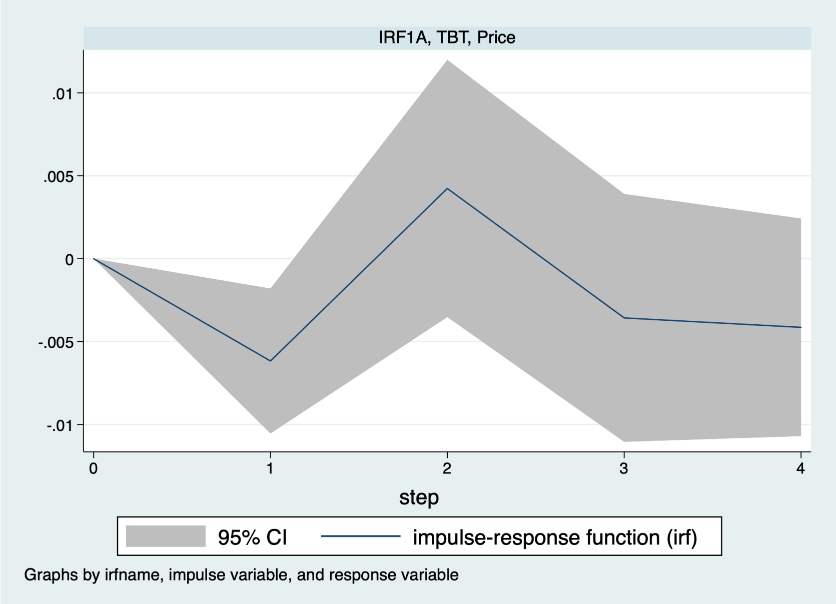
Figure 3: Impulse Response Function of Bitcoin's Price (1)
Figure 3 graphs the expected future fluctuations of Bitcoin's price following a one-unit increase of the tweets' polarity. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the polarity indicator. The y-axis reports the variation of Bitcoin's price in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



On average, Bitcoin's price decreases, ceteris paribus, by 6$ per thousand tweets posted one day before but increases by 13$ per thousand tweets posted two days earlier before being decreased again by 8$ per thousand tweets posted three days preceding. In other words, if 100.000 tweets were to be posted today, the ceteris paribus effect of these tweets, predicted by the VAR model below, would decrease Bitcoin's price by 600$ on the first day, increase Bitcoin's price by 1300$ on the second day and finally decrease it again by 800$, on the third day. TBT's one lagged weekly volatility also significantly predicts Bitcoin's future price, at the 5% significance level. If the weekly volatility of daily tweet posted increases, Bitcoin's future price (one lag) decreases.

Similarly to what was described through Figure 3, Figure 4 exhibits the Impulse Response Function of Bitcoin's price relative to the TBT sentiment indicator. The IRF reflects, first, a strictly negative effect of TBT over Bitcoin's future price on the subsequent day following the one-unit increase of the TBT indicator. On the second day following the increase in the daily number of tweets (TBT), Bitcoin's price experiences a substantial rise and indicate a positive correlation. TBT's positive effect on Bitcoin's future price appears, however, to be ephemeral as the IRF converges back to the negative area, predicting instead a negative correlation between the variables when looking at a longer time-horizon (observing the third and fourth lagged-values). Overall, despite the fact that TBT shows a clear positive effect over Bitcoin's future price on the second subsequent day following TBT's one-unit increase, this positive correlation disappears when looking at TBT's influence in the longer-term. Numerically, an increase of 1000 tweets, from one day to the other, will first generate a drop of approximately 5$ in Bitcoin's price during the first subsequent day while this drop is going to be entirely recovered on the second day and will even bounce back to almost 5$ above Bitcoin's initial price. Nonetheless, it is expected that this upward movement is not going to sustain and Bitcoin's price is likely to then converge back to slightly less than 5$ below its initial price.

Figure 4: Impulse Response Function of Bitcoin's Price (2)
Figure 4 graphs the expected future fluctuations of Bitcoin's price following a one-unit increase of the TBT indicator. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the TBT sentiment indicator. The y-axis reports the variation of Bitcoin's price in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval



It could be possible that the effects described in the previous paragraph are the consequence of reverse causality. In order to assess the causality between the variables, a 'Granger causality test' is being performed. The Granger causality test's null hypothesis is that the sentiment indicators do not 'Granger cause' variations in Bitcoin future price. The Granger causality test's output is displayed through Table 5.

The Granger causality test indicates a p-value of 0.088 when testing the hypothesis that the polarity indicator does not Granger cause fluctuation in Bitcoin's price. This hypothesis can therefore be rejected, at the 10% significance level, implying that the polarity indicator does Granger causes variations in Bitcoin's future price. This finding suggests that bitcoin's price can be explained, at least partially, by the polarity sentiment indicator.

Table 5: Granger causality Wald test for the specification relative to Bitcoin's Price
Table 5 exhibits the result of the Granger causality test. The Granger causality test estimates whether excluding a variable from the equation significantly decrease the model's predictive power. The Granger causality test tests the null hypothesis stating that there is no causative effect of a variable on Bitcoin's future price. A p-value below the 0.1, 0.05 and 0.01 threshold means that this null hypothesis can be rejected at the 10%, 5% and 1% confidence levels. The first two columns 'indicates the two variables between which causality is being tested. The third column displays the estimated 'Chi-Square' statistic for each causality test, the fourth column indicates the test's corresponding degree of freedom and the last column report the test's p-value.

Equation	Excluded	chi2	df	Prob>Chi2
Price	Polarity	8.093	4	0.088
Price	TBT	18.717	4	0.001
Price	Polarity vol	0.762	4	0.943
Price	TBT vol	18.018	4	0.001
Price	VIX	2.792	4	0.593
Price	S&P500	0.733	4	0.947
Price	Google Trends	22.316	4	0.000
Price	P(Gold)	2.581	4	0.630
Price	ALL	100.120	32	0.000

Concerning the TBT sentiment indicator, the null hypotheses stating that TBT does not Granger cause changes in Bitcoin's future price can also be rejected at all significance level as its p-value equals 0.001. Both indicators have therefore, to a certain degree, predictive power over Bitcoin's future price, despite the TBT sentiment indicator being more relevant as its volatility also exhibits explanatory power, which is not the case for the polarity indicator.

6.3. Second Hypothesis: Social Sentiment and Bitcoin's Volatility

The second VAR model investigates whether the lags of the sentiment indicators can reveal to be significantly correlated with Bitcoin's volatility and whether these indicators can explain, to a certain degree, variations in the doily volatility of Bitcoin. The results are displayed in Table 6.

Similarly to what had been observed through the first model, the polarity indicator is again a significant predictor of Bitcoin's volatility but only when regressing using the indicator's third lag. This result isn't surprising considering that polarity's third lag was already significantly correlated to Bitcoin's future price but one might wonder why only the third lag had a predictive influence and not the first or second lag.

Table 6: VAR results for Bitcoin's Price Volatility

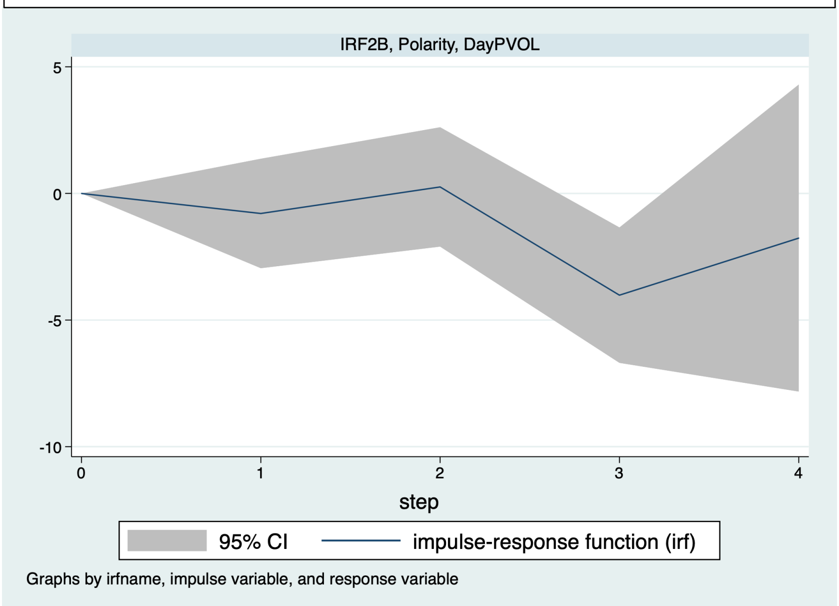
Table 6 exhibits the VAR results relative to the second hypothesis. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observation's average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Daily Price Volatility						
Daily Price Volatility						
L1.	-0.353	0.066	-5.320	0.000	-0.483	-0.223
L2.	0.065	0.008	8.370	0.000	0.050	0.081
L3.	2.406	0.093	25.860	0.000	2.224	2.589
L4.	-0.557	0.115	-4.840	0.000	-0.782	-0.331
Polarity						
L1.	-0.963	1.110	-0.870	0.386	-3.138	1.212
L2.	1.040	1.158	0.900	0.369	-1.229	3.309
L3.	-3.615	1.186	-3.050	0.002	-5.939	-1.292
L4.	1.009	1.027	0.980	0.326	-1.003	3.022
TBT (per 1,000 tweets)						
L1.	0.145	0.137	1.060	0.289	-0.123	0.413
L2.	-0.579	0.214	-2.710	0.007	-0.998	-0.160
L3.	0.928	0.280	3.310	0.001	0.379	1.477
L4.	-0.173	0.163	-1.060	0.288	-0.493	0.147
Polarity1weekvol						
L1.	-7.183	5.587	-1.290	0.199	-18.133	3.768
L2.	5.778	7.000	0.830	0.409	-7.943	19.498
L3.	-3.389	7.336	-0.460	0.644	-17.767	10.988
L4.	3.279	6.042	0.540	0.587	-8.564	15.121
TBT1weekvol						
L1.	-0.003	0.001	-5.650	0.000	-0.005	-0.002
L2.	0.001	0.001	1.280	0.202	-0.001	0.004
L3.	0.000	0.002	0.070	0.940	-0.003	0.003
L4.	-0.000	0.001	-0.310	0.754	-0.002	0.002
VIX						
L1.	4.096	1.254	3.260	0.001	1.637	6.554
L2.	-4.332	1.770	-2.450	0.014	-7.802	-0.862
L3.	-0.721	1.774	-0.410	0.685	-4.199	2.757
L4.	1.227	1.155	1.060	0.288	-1.036	3.490
S&P500						
L1.	0.073	0.045	1.610	0.106	-0.016	0.161
L2.	-0.115	0.058	-2.000	0.045	-0.228	-0.003
L3.	0.023	0.049	0.470	0.635	-0.072	0.119
L4.	0.015	0.032	0.460	0.647	-0.048	0.078
Google Trends						
L1.	0.065	0.008	8.160	0.000	0.049	0.081
L2.	0.017	0.016	1.070	0.284	-0.014	0.048
L3.	-0.032	0.033	-0.960	0.336	-0.096	0.033
L4.	-0.076	0.025	-3.030	0.002	-0.125	-0.027
P(Gold)						
L1.	-0.089	0.077	-1.160	0.246	-0.240	0.062
L2.	-0.036	0.120	-0.300	0.766	-0.272	0.200
L3.	0.050	0.123	0.410	0.684	-0.192	0.292
L4.	0.086	0.088	0.980	0.328	-0.086	0.258
_cons	10.461	8.197	1.280	0.202	-5.605	26.527

Similarly to what had been observed through the first model, the polarity indicator is again a significant predictor of Bitcoin's volatility but only when regressing using the indicator's third lag. This result isn't surprising considering that polarity's third lag was already significantly correlated to Bitcoin's future price but one might wonder why only the third lag had a predictive influence and not the first or second lag. Polarity's third lag is significantly (at the 1% significance level) and negatively correlated to Bitcoin's price volatility with a p-value of 0.002 and a coefficient of -3.615. The polarity indicator's weekly volatility, however, fails to provide explanatory power to the model and therefore does not predict Bitcoin's future price volatility.

Figure 5 shows Bitcoin's price volatility impulse response consequent to a one-unit increase of the polarity indicator. While it can be concluded that the polarity indicator has no or a very low impact on Bitcoin's future price during the two first days following the sentiment increase, the third subsequent day exhibits a clear and significantly negative correlation. Indeed, both the IRF and its 95% confidence interval lie below the '0 line' and the sentiment polarity's one-unit increase is likely to consequently diminish Bitcoin's price volatility of approximately 4$ on the third subsequent day. Overall, the polarity indicator's impact on Bitcoin's price volatility can be estimated to be negative, especially through its third lagged-value. The IRF result confirms the intuition retrieved from the VAR output analysis.

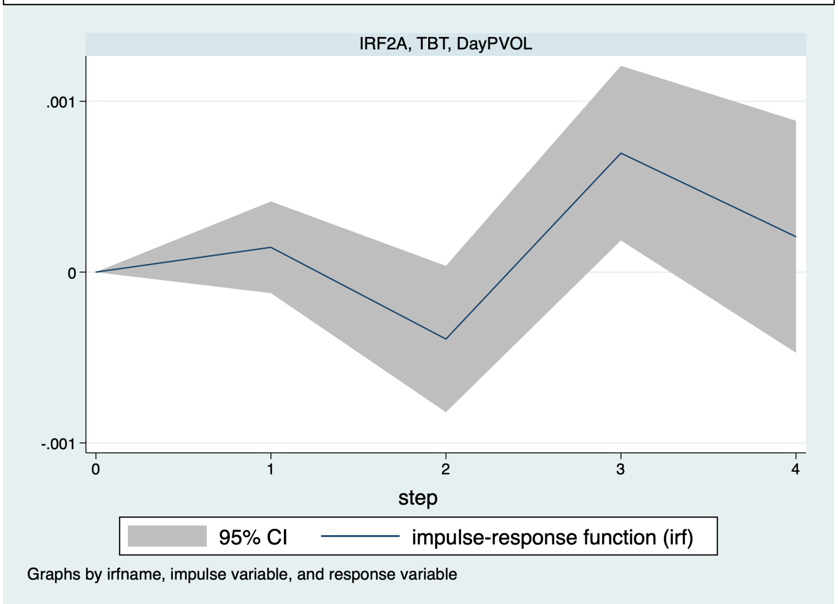
Figure 5: Impulse Response Function of Bitcoin's Price Volatility (1)
Figure 5 graphs the expected future fluctuations of Bitcoin's price volatility following a one-unit increase of the tweets' polarity. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the polarity indicator. The y-axis reports the variation of Bitcoin's price volatility in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



Furthermore, the TBT indicator, representing the daily total discussions about Bitcoin, is significantly correlated with Bitcoin's price volatility but only for its second and third lag. TBT's first lag, which was significant in the previous VAR model, does not have any explanatory power anymore. Both TBT's second and third lag explain variation in Bitcoin's price volatility significantly, at the 1% significance level with coefficients of 0.001. The only difference is that, for the second lag, the coefficient is negative whereas is positive for the third lag. All in all, this finding implies that an increase in the TBT sentiment indicator at time t will generate a decrease in Bitcoin's price volatility at time t+2 but will afterwards generate an approximatively corresponding increase in Bitcoin's price volatility at time t+3, ceteris paribus. The magnitude of the effects can be estimated as follow: two days following an increase of 1,000 tweets, Bitcoin's volatility will decrease by 0.579 but will increase by 0.928 on the third, ceteris paribus. Considering both effects, the overall effect of an increase of the TBT indicator will generate an increase in Bitcoin's volatility in the subsequent periods. While the TBT indicator's first lag failed to provide explanatory power to the model, the first lag of the TBT's weekly volatility does. Indeed, with a p-value of 0.000, TBT's weekly volatility significantly explains variations in Bitcoin's subsequent period price volatility, at the maximum significance level. Since its coefficient is negative, it is expected that an increase in TBT's weekly volatility will consequently generate a subsequent decrease in Bitcoin's price volatility, ceteris paribus.

Figure 6 represents the IRF relative to TBT's one-unit increase effect on Bitcoin's price volatility over the next four days following the shock. The IRF highlights a response composed of multiple oscillations around zero. In fact, despite the first step's response staying close to the '0 line', the second and third steps, representing the only two significant coefficients of the VAR model output, exhibit consequent derivations from zero. The second day following TBT's shock, Bitcoin's daily price volatility is likely to drop below its initial level, but this effect is not statistically significant as the 95% confidence interval fails to exclude zero. Nonetheless, the third lag highlights probably the biggest response as the IRF peaks to its maximum value and, this time, significantly lies above zero. However, this remarkable increase in price volatility is likely to converge back towards zero, on the fourth day following TBT's shock. All in all, the IRF shows mixed-results as it clearly exhibits oscillations both above and below zero. However, only the oscillation above 0 appears to be statistically significant which tends to indicate an overall positive effect of the TBT sentiment indicator on Bitcoin's future price volatility.

Figure 6: Impulse Response Function of Bitcoin's Price Volatility (2)
Figure 6 graphs the expected future fluctuations of Bitcoin's price volatility following a one-unit increase of the TBT indicator. The x-axis, labelled as steps, indicates the number of days following the one-unit shock of the TBT indicator. The y-axis reports the variation of Bitcoin's price volatility in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



In order to assess the causality between the variables, a Granger causality test is performed. The output of the test is displayed by Table 7 on the next page.

Firstly, it appears that we can reject the null hypothesis stating that the polarity indicator does not Granger cause variations in Bitcoin's daily price volatility, at the 5% significance level. It can therefore be concluded that the polarity indicator does Granger cause fluctuations in Bitcoin's daily price volatility but only with 3 days interval. Relative to the other sentiment indicator, both TBT and its volatility can be considered as having a causative effect on Bitcoin's daily price volatility since their corresponding p-values (0.001 and 0.000, respectively) at the Granger causality test are both significant at the 1% significance level.

To conclude, it appears that both sentiment indicators have explanatory power on the future variations in Bitcoin's daily price volatility. The polarity indicator is slightly less powerful at predicting Bitcoin's future volatility since only its third lagged value is relevant in this prediction. On the other hand, the TBT sentiment indicator exhibits great predictive power as both its value and its volatility help predicting Bitcoin's price volatility fluctuations for all three first lags.

Table 7: Granger causality Wald test for the specification relative to Bitcoin's Price Volatility

Table 5 exhibits the result of the Granger causality test. The Granger causality test estimates whether excluding a variable from the equation significantly decrease the model's predictive power. The Granger causality test tests the null hypothesis stating that there is no causative effect of a variable on Bitcoin's future price volatility. A p-value below the 0.1, 0.05 and 0.01 threshold means that this null hypothesis can be rejected at the 10%, 5% and 1% confidence levels. The first two columns 'indicates the two variables between which causality is being tested. The third column displays the estimated 'Chi-Square' statistic for each causality test, the fourth column indicates the test's corresponding degree of freedom and the last column report the test's p-value.

Equation	Excluded	chi2	df	Prob>Chi2
Daily Price Volatility	Polarity	10.601	4	0.031
Daily Price Volatility	TBT	18.740	4	0.001
Daily Price Volatility	Polarity vol	1.996	4	0.737
Daily Price Volatility	TBT vol	43.636	4	0.000
Daily Price Volatility	VIX	13.522	4	0.009
Daily Price Volatility	S&P500	11.578	4	0.021
Daily Price Volatility	Google Trends	137.940	4	0.000
Daily Price Volatility	P(Gold)	5.505	4	0.239
Daily Price Volatility	ALL	322.040	32	0.000

Firstly, it appears that we can reject the null hypothesis stating that the polarity indicator does not Granger cause variations in Bitcoin's daily price volatility, at the 5% significance level. It can therefore be concluded that the polarity indicator does Granger cause fluctuations in Bitcoin's daily price volatility but only with 3 days interval. Relative to the other sentiment indicator, both TBT and its volatility can be considered as having a causative effect on Bitcoin's daily price volatility since their corresponding p-values (0.001 and 0.000, respectively) at the Granger causality test are both significant at the 1% significance level.

To conclude, it appears that both sentiment indicators have explanatory power on the future variations in Bitcoin's daily price volatility. The polarity indicator is slightly less powerful at predicting Bitcoin's future volatility since only its third lagged value is relevant in this prediction. On the other hand, the TBT sentiment indicator exhibits great predictive power as both its value and its volatility help predicting Bitcoin's price volatility fluctuations for all three first lags.

6.4. Third Hypothesis: Social Sentiment and Bitcoin's Trading Volume

The following section aims at providing an answer to the third hypothesis. The third hypothesis investigates whether the findings made over the two last models in section 6.2. and 6.3. can be reiterated when looking at another financial indicator: Bitcoin's trading volume (measured in US Dollar). Table 8, on the next page, presents the results of the VAR model investigating this hypothesis.

The polarity sentiment indicator exhibits no significant correlation with Bitcoin's trading volume for any of its for lags since all four p-values lie above 0.3. Its volatility cannot provide explanatory power to the model neither. As a result, the polarity indicator is therefore uncorrelated with Bitcoin's trading volume and none of its coefficient results can be interpreted with confidence. Figure 7 represents the IRF of Bitcoin's trading volume relative to a one-unit shock in the polarity indicator. The Impulse Response Function confirms the conclusion drawn while looking at the VAR output (Table 8) and seem to indicate no particular correlation between investors' sentiment polarity and future trading volume since the IRF is almost flat.

The other sentiment measure, TBT, only sees its second lag to be correlated with Bitcoin's trading volume. This effect, however, is significant at the 10% significance level only which might simply be a consequence of randomness considering the relatively high p-value (0.072) and the non-significance of its other lags. The coefficient of TBT's second lag predicts a negative relationship between the variables. As such, if the total number of tweets increase, it is expected that the trading volume will decrease, slightly, in the subsequent 2 periods. TBT's weekly volatility is however a much better predictor than TBT's value itself. Indeed, TBT's weekly volatility significantly explains variations in Bitcoin's trading volume for the three first lags. The first lag is significant at the 10% significance level, the second at the 1% significance level and the third lag is significant at the 5% level. The first lag is expected to have a negative impact on the trading volume while the second lag is expected to have a positive effect and the third one a negative effect again. In other words, increasing TBT's weekly volatility by one unit leads to decrease Bitcoin's trading volume by 18,400$ in the subsequent period while increasing it by 66,600$ on the second day before pushing it down again of 55,100$ three days after, ceteris paribus.

Table 8: VAR results for Bitcoin's trading volume, in US Dollars.

Table 8 exhibits the VAR results relative to the third hypothesis. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observations' average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Volume						
Volume						
L1.	0.036	0.111	0.320	0.750	-0.183	0.254
L2.	0.375	0.113	3.330	0.001	0.154	0.595
L3.	0.379	0.106	3.570	0.000	0.171	0.588
L4.	-0.046	0.081	-0.570	0.571	-0.206	0.114
Polarity						
L1.	-1.85e+07	1.78e+07	-1.040	0.300	-5.33e+07	1.64e+07
L2.	1.51e+07	1.87e+07	0.810	0.419	-2.15e+07	5.17e+07
L3.	-1.16e+07	1.91e+07	-0.610	0.542	-4.90e+07	2.58e+07
L4.	-8881244	1.64e+07	-0.540	0.589	-4.11e+07	2.33e+07
TBT						
L1.	1786.831	2125.718	0.840	0.401	-2379.499	5953.162
L2.	-6005.413	3332.807	-1.800	0.072	-1.25e+04	526.769
L3.	3850.584	4476.967	0.860	0.390	-4924.111	12625.280
L4.	954.676	2506.640	0.380	0.703	-3958.248	5867.600
Polarity1weekvol						
L1.	6.54e+07	8.98e+07	0.730	0.466	-1.11e+08	2.41e+08
L2.	-1.17e+08	1.13e+08	-1.030	0.301	-3.39e+08	1.05e+08
L3.	1.07e+08	1.18e+08	0.910	0.364	-1.25e+08	3.39e+08
L4.	-6.17e+07	9.72e+07	-0.630	0.526	-2.52e+08	1.29e+08
TBT1weekvol						
L1.	-1.84e+04	9941.252	-1.860	0.063	-3.79e+04	1035.308
L2.	66602.850	18273.840	3.640	0.000	30786.790	1.02e+05
L3.	-5.51e+04	27775.720	-1.980	0.047	-1.10e+05	-662.894
L4.	5540.543	16542.140	0.330	0.738	-2.69e+04	37962.540
VIX						
L1.	-4.75e+07	1.95e+07	-2.430	0.015	-8.57e+07	-9186397
L2.	1.45e+08	2.44e+07	5.960	0.000	9.76e+07	1.93e+08
L3.	-1.12e+08	2.68e+07	-4.180	0.000	-1.64e+08	-5.93e+07
L4.	1.86e+07	1.86e+07	1.000	0.318	-1.79e+07	5.50e+07
S&P500						
L1.	-1558538	7.10e+05	-2.190	0.028	-2951065	-1.66e+05
L2.	2762576	8.70e+05	3.170	0.002	1056917	4468234
L3.	-1534327	7.69e+05	-1.990	0.046	-3042458	-2.62e+04
L4.	3.13e+05	5.20e+05	0.600	0.547	-7.06e+05	1332412
Google Trends						
L1.	7.44e+05	1.81e+05	4.110	0.000	3.89e+05	1098476
L2.	2.29e+05	2.96e+05	0.770	0.441	-352400	8.10e+05
L3.	-1220257	553666	-2.200	0.028	-2305422	-1.35e+05
L4.	8.71e+05	4.15e+05	2.100	0.036	58546.690	1683834
P(Gold)						
L1.	-1.07e+05	1238432	-0.090	0.931	-2534717	2319849
L2.	5.23e+05	1935664	0.270	0.787	-3271137	4316527
L3.	-1141013	1979305	-0.580	0.564	-5020379	2738353
L4.	7.72e+05	1412960	0.550	0.585	-1997100	3541600
_cons	-3.61e+07	1.37e+08	-0.260	0.793	-3.06e+08	2.33e+08

Figure 7: Impulse Response Function of Bitcoin's Trading Volume (1)
Figure 7 graphs the expected future fluctuations of Bitcoin's trading volume ($) following a one-unit increase of the tweets' polarity. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the polarity indicator. The y-axis reports the variation of Bitcoin's trading volume, in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.

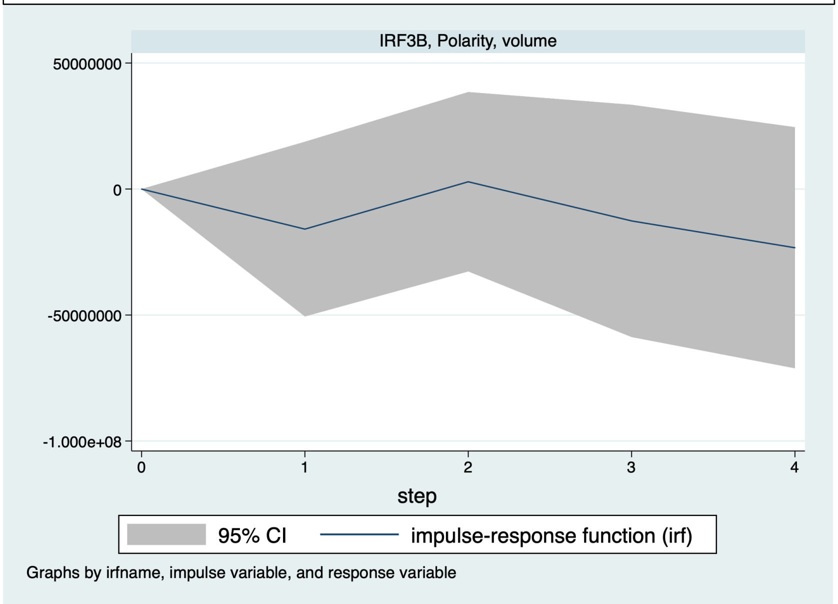


Figure 8: Impulse Response Function of Bitcoin's Trading Volume (2)
Figure 8 graphs the expected future fluctuations of Bitcoin's trading volume ($) following a one-unit increase in TBT. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the TBT indicator. The y-axis reports the variation of Bitcoin's trading volume, in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.

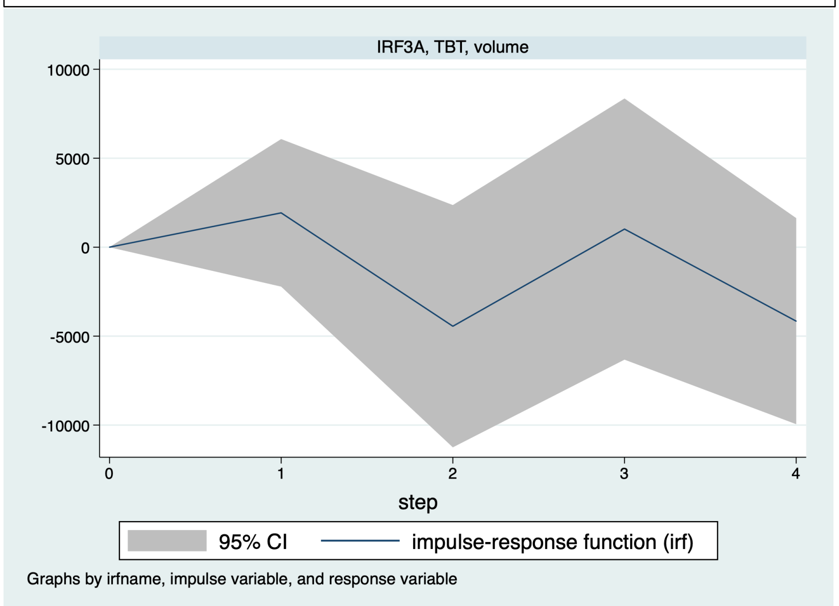


Figure 8 graphically represents the aforementioned effects of TBT's first four lags on Bitcoin's trading volume. Overall, it appears that TBT's shocks have a negative effect on Bitcoin's future trading volume, despite the IRF never significantly lying below zero. Indeed, it appears that the IRF function progressively decrease as time passes by. The IRF estimates that a 1000 tweets increase will correspondingly generate a drop of slightly less that 500,000$ in Bitcoin's future daily trading volume on the second and fourth day following TBT's shock, compared to the volume's initial level.

The next step of this analysis would be to determine the causality between the variables which is being done through the following Granger causality test (Table 9):

Table 9: Granger causality Wald test for the specification relative to Bitcoin's Volume
Table 9 exhibits the result of the Granger causality test. The Granger causality test estimates whether excluding a variable from the equation significantly decrease the model's predictive power. The Granger causality test tests the null hypothesis stating that there is no causative effect of a variable on Bitcoin's future trading volume. A p-value below the 0.1, 0.05 and 0.01 threshold means that this null hypothesis can be rejected at the 10%, 5% and 1% confidence levels. The first two columns 'indicates the two variables between which causality is being tested. The third column displays the estimated 'Chi-Square' statistic for each causality test, the fourth column indicates the test's corresponding degree of freedom and the last column report the test's p-value.

Equation	Excluded	chi2	df	Prob>Chi2
volume	Polarity	2.052	4	0.726
volume	TBT	3.912	4	0.418
volume	Polarity vol	1.343	4	0.854
volume	TBT vol	16.669	4	0.002
volume	VIX	44.258	4	0.000
volume	S&P500	13.081	4	0.011
volume	Google Trends	71.760	4	0.000
volume	P(Gold)	0.502	4	0.973
volume	ALL	179	32	0.000

Since the polarity sentiment indicator and its volatility both failed to be relevant in explaining variations of Bitcoin's daily trading volume, there is no need to observe and/or comment their results to the Granger causality test. Instead, the Granger causality test will help us determine whether TBT's explanatory power has a causative effect on Bitcoin's trading volume. Firstly, the TBT variable itself shows a p-value of 0.418 which deny TBT of any causative influence on Bitcoin's trading volume. As a consequence, TBT's significant second lag can be disregarded. Nonetheless, TBT's weekly volatility highlights a p-value of 0.002.

This finding implies that the null hypothesis denying the volatility variable of any causative effect on the trading volume can be rejected at the 1% significance level.

To conclude, despite both sentiment indicators being irrelevant in predicting Bitcoin's future trading volume, the volatility of the TBT sentiment indicator displayed a highly significant explanatory power on the Bitcoin's future volume. The effect relative to the total number of tweets posted each day is expected to last for the three following days, at least.

6.5. Fourth Hypothesis: Social sentiment and liquidity risk

Since investors' social sentiment is a relevant factor in explaining the variation of Bitcoin's price, volatility and trading volume, the same might hold to predict variations in Bitcoin's liquidity. For this research, Bitcoin's liquidity risk is proxied through its bid-ask spread and its daily volume, measured with the BTC unit.

6.5.1. Bitcoin's tightness, proxied by its bid-ask spread

The first model regresses the bid-ask spread tightness indicator and the VAR output is displayed in Table 10.

On the one hand, the polarity indicator's first lag is correlated to Bitcoin's bid-ask spread in the subsequent period. This finding appears to confirm the intuition emitted in the hypothesis section stating: a positive increase in the polarity is likely to provoke a surge of investors to buy Bitcoin which in turn is likely to narrow the bid-ask spread. The VAR output shows that Bitcoin's market becomes tighter as the social sentiment's polarity gets more positive. All in all, the polarity indicator therefore seems to decrease Bitcoin's liquidity risk when the proportion of positive tweets, relative to the proportion of negative tweets, increases. One unit increase in the ratio of positive tweets to the negative ones is expected to decrease Bitcoin's bid-ask spread by 0.4% of Bitcoin's monetary value. Nevertheless, this effect is only significant at the 10% level (p-value = 0.07). On the other hand, the variable representing the polarity indicator's volatility fails to significantly explain variations in Bitcoin's bid-ask spread. It was expected that an increase in the volatility of the sentiment would have generated a corresponding increase in Bitcoin's bid-ask spread.

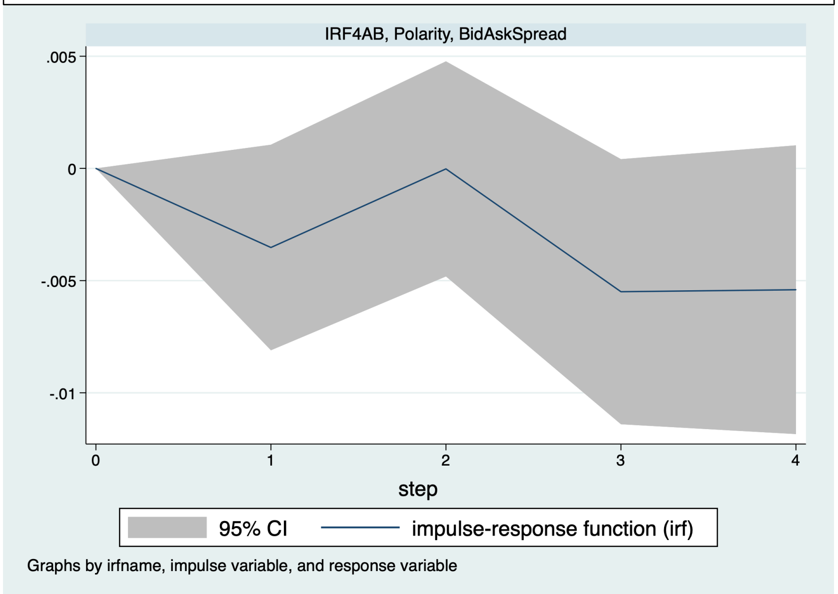
Table 10: VAR results for Bitcoin's bid-ask spread

Table 10 exhibits the VAR results relative to the fourth hypothesis, sub-hypothesis 'a'. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observation's average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
BidAskSpread						
BidAskSpread						
L1.	0.543	0.065	8.320	0.000	0.415	0.671
L2.	0.172	0.109	1.580	0.115	-0.042	0.385
L3.	-0.058	0.094	-0.620	0.536	-0.242	0.126
L4.	0.324	0.088	3.670	0.000	0.151	0.498
Polarity						
L1.	-0.004	0.002	-1.810	0.070	-0.009	0.000
L2.	0.003	0.002	1.240	0.214	-0.002	0.008
L3.	-0.003	0.002	-1.070	0.283	-0.007	0.002
L4.	-0.002	0.002	-0.850	0.395	-0.006	0.002
TBT (per 100k tweets)						
L1.	0.014	0.028	0.510	0.613	-0.040	0.068
L2.	-0.104	0.042	-2.470	0.014	-0.187	-0.021
L3.	0.122	0.055	2.240	0.025	0.015	0.229
L4.	-0.038	0.032	-1.180	0.236	-0.101	0.025
Polarity1weekvol						
L1.	0.008	0.012	0.660	0.507	-0.015	0.031
L2.	-0.019	0.015	-1.320	0.188	-0.048	0.009
L3.	0.024	0.015	1.550	0.121	-0.006	0.054
L4.	-0.016	0.013	-1.280	0.200	-0.041	0.009
TBT1weekvol						
L1.	0.000	0.000	0.280	0.777	-0.000	0.000
L2.	0.000	0.000	1.200	0.231	-0.000	0.000
L3.	-0.000	0.000	-2.030	0.042	-0.000	-0.000
L4.	0.000	0.000	1.930	0.054	-0.000	0.000
VIX						
L1.	0.003	0.003	1.270	0.205	-0.002	0.008
L2.	-0.001	0.003	-0.230	0.816	-0.007	0.005
L3.	-0.004	0.003	-1.030	0.305	-0.010	0.003
L4.	0.001	0.002	0.620	0.536	-0.003	0.006
SP500						
L1.	-0.000	0.000	-0.400	0.690	-0.000	0.000
L2.	0.000	0.000	0.410	0.683	-0.000	0.000
L3.	0.000	0.000	0.110	0.916	-0.000	0.000
L4.	-0.000	0.000	-0.430	0.668	-0.000	0.000
Google Trends						
L1.	-0.000	0.000	-0.630	0.529	-0.000	0.000
L2.	0.000	0.000	2.390	0.017	0.000	0.000
L3.	-0.000	0.000	-0.200	0.841	-0.000	0.000
L4.	-0.000	0.000	-0.700	0.484	-0.000	0.000
P(Gold)						
L1.	-0.000	0.000	-2.200	0.028	-0.001	-0.000
L2.	0.000	0.000	1.920	0.054	-0.000	0.001
L3.	-0.000	0.000	-1.760	0.079	-0.001	0.000
L4.	0.000	0.000	1.870	0.061	-0.000	0.001
_cons	0.028	0.019	1.490	0.136	-0.009	0.065

Figure 9, representing the bid-ask spread's impulse response function to a one-unit shock in the investors' sentiment polarity, confirms the observation made while analyzing the VAR output (Table 10). The IRF clearly illustrates how a positive shock in investors' sentiment polarity is likely to consequently decrease Bitcoin's market tightness and therefore reduces the liquidity risk associated with holding Bitcoins. The IRF never estimates an increase in Bitcoin's bid-ask spread level, for any of the four days following the shock in sentiment polarity, in comparison with the bid-ask spread's initial level. Instead, the IRF appears to converge around or even below -0.005 as the steps increase. The magnitude of the effect can be interpreted as follow: Bitcoin's bid-ask spread is likely to decrease of approximately 0.5% of Bitcoin's value, by the third to the fourth day following the shock in sentiment polarity.

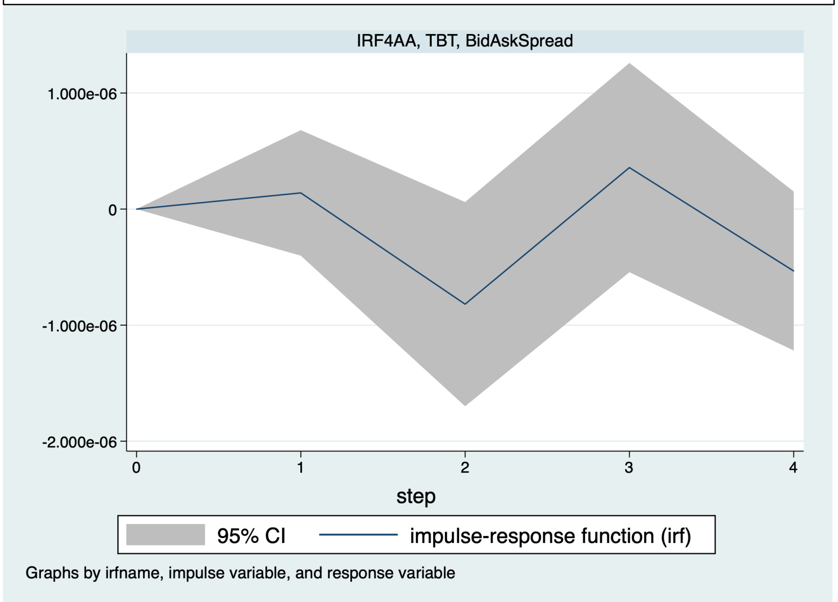
Figure 9: Impulse Response Function of Bitcoin's Bid-Ask Spread (1)
Figure 9 graphs the expected future fluctuations of Bitcoin's bid-ask spread following a one-unit increase in the tweets' polarity. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the polarity indicator. The y-axis reports the variation of Bitcoin's bid-ask spread, in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



The TBT sentiment indicator provides more accurate explanatory power to the VAR model through two lagged values, the second and third lags, both significant at the 5% significance level. TBT's second lagged value negatively impacts Bitcoin's bid-ask spread while the third lagged value positively correlates with the bid-ask spread. The coefficients help to provide an estimation to quantify the effects. Ceteris paribus, a rise of 10,000 tweets from one day to the other will generate, on the following second day, an approximately 10% decrease of Bitcoin's spread before rebounding with a 12% increase on the third day following the TBT increase. The volatility of the TBT also clearly exhibits significant correlations with Bitcoin's bid-ask spread, through its third and fourth lagged values, indicating a longer-term impact of the indicator. Specifically, it appears that the third lag of TBT's weekly volatility is negatively correlated while its fourth lag is positively correlated to Bitcoin's spread. These findings are significant at the 5% and 10% significance levels, respectively.

Figure 10 represents Bitcoin's bid-ask spread IRF and illustrates the aforementioned findings. Despite the IRF never significantly lying above or below zero, the second lagged-value's negative drop and the third one's positive bounce are clearly observable and represents the IRF's minimum and maximum values, respectively. On average, the IRF appears to converge towards zero through the described double oscillation on its second and third steps. In conclusion, TBT's shock tightens Bitcoin's market, reducing its associated liquidity risk, on the second day following the tweets' increase while that same shock is likely to have the opposite effect on the third day following the shock, when compared to its initial bid-ask spread's level.

Figure 10: Impulse Response Function of Bitcoin's Bid-Ask Spread (2)
Figure 10 graphs the expected future fluctuations of Bitcoin's bid-ask spread following a one-unit increase in the TBT indicator. The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the TBT indicator. The y-axis reports the variation of Bitcoin's bid-ask spread, in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



The Granger causality test was performed in order to confirm or to refute the causality of the sentiment variables on Bitcoin's bid-ask spread. The result of the Granger test can be found in Table 11, on the next page.

Table 11: Granger causality Wald test for the specification relative to Bitcoin's bid-ask spread

Table 11 exhibits the result of the Granger causality test. The Granger causality test estimates whether excluding a variable from the equation significantly decrease the model's predictive power. The Granger causality test tests the null hypothesis stating that there is no causative effect of a variable on Bitcoin's future bid-ask spread. A p-value below the 0.1, 0.05 and 0.01 threshold means that this null hypothesis can be rejected at the 10%, 5% and 1% confidence levels. The first two columns 'indicates the two variables between which causality is being tested. The third column displays the estimated 'Chi-Square' statistic for each causality test, the fourth column indicates the test's corresponding degree of freedom and the last column report the test's p-value.

Equation	Excluded	chi2	df	Prob>Chi2
BidAsk	Polarity	5.997	4	0.199
BidAsk	TBT	7.996	4	0.092
BidAsk	Polarityvol	3.192	4	0.526
BidAsk	TBTvol	4.690	4	0.321
BidAsk	VIX	3.527	4	0.474
BidAsk	S&P500	4.458	4	0.348
BidAsk	Google Trends	18.008	4	0.001
BidAsk	P(Gold)	11.204	4	0.024
BidAsk	ALL	89.762	32	0.000

Only the TBT indicator barely lies below the 0.1 p-value threshold. TBT's causative effect is therefore confirmed by the Granger causality test but this result should be taken with a grain of salt considering its p-value (0.092). The volatility of the TBT indicator is also denied of all causative effect on the bid-ask spread since the Granger causality test failed to reject the null hypothesis. Both the polarity indicator and its volatility are denied as well of causative impact on the dependent variable. The effect described in the previous paragraph, observing a negative explanatory power of the polarity indicator's first lag on the spread, is rejected based on the Granger causality test's disappointing result. The analysis of the first liquidity indicator, Bitcoin's tightness, proxied by Bitcoin's bid-ask spread, delivered mixed results which makes it even more interesting to observe another liquidity indicator, Bitcoin's breadth, measured through Bitcoin's daily trading volume, to assess whether the social sentiment has potentially a causative effect on Bitcoin's liquidity.

6.5.2. Bitcoin's breadth, proxied by its daily BTC trading Volume

Table 12, on the next page, displays the results of the VAR model relative to the BTC daily trading volume in function of the sentiment indicators and other controls.

Table 12: VAR results for Bitcoin's daily trading volume, in BTC

Table 12 exhibits the VAR results relative to the fourth hypothesis, sub-hypothesis 'b'. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observation's average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

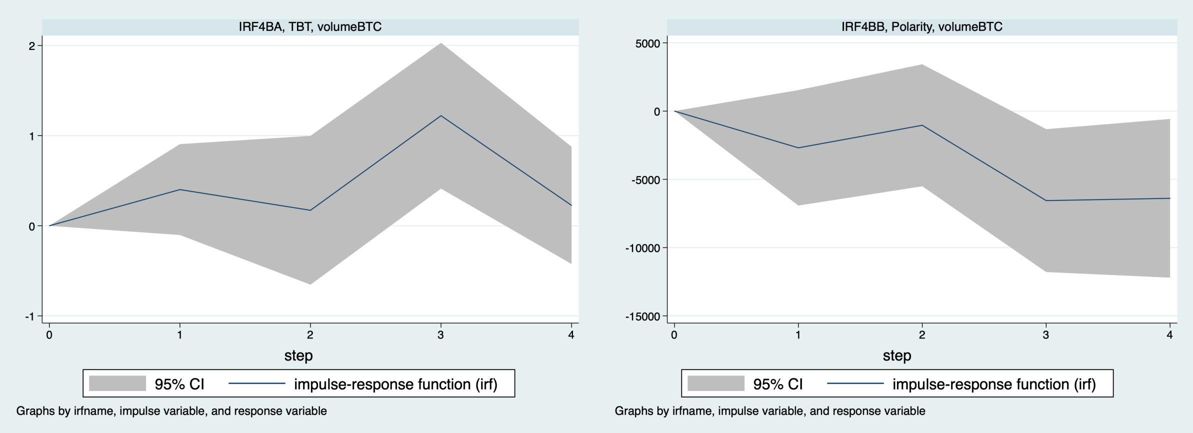
	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
VolumeBTC						
VolumeBTC						
L1.	0.525	0.070	7.470	0.000	0.387	0.663
L2.	-0.081	0.074	-1.090	0.277	-0.227	0.065
L3.	0.169	0.091	1.860	0.063	-0.009	0.346
L4.	0.113	0.088	1.280	0.199	-0.060	0.286
Polarity						
L1.	-3407.633	2232.422	-1.530	0.127	-7783.100	967.834
L2.	1774.144	2319.690	0.760	0.444	-2772.365	6320.654
L3.	-4360.053	2376.320	-1.830	0.067	-9017.554	297.448
L4.	-976.050	2054.762	-0.480	0.635	-5003.310	3051.211
TBT						
L1.	0.370	0.266	1.390	0.164	-0.151	0.891
L2.	-0.471	0.400	-1.180	0.240	-1.255	0.314
L3.	1.057	0.516	2.050	0.041	0.045	2.069
L4.	-0.417	0.310	-1.350	0.178	-1.024	0.190
Polarity1weekvol						
L1.	-139.828	11275.760	-0.010	0.990	-2.22e+04	21960.250
L2.	-9863.166	14145.620	-0.700	0.486	-3.76e+04	17861.750
L3.	16980.260	14774.290	1.150	0.250	-1.20e+04	45937.330
L4.	-4145.974	12149.580	-0.340	0.733	-2.80e+04	19666.760
TBT1weekvol						
L1.	0.155	1.169	0.130	0.894	-2.137	2.447
L2.	2.476	2.128	1.160	0.244	-1.694	6.647
L3.	-3.639	3.313	-1.100	0.272	-10.133	2.855
L4.	-0.109	2.021	-0.050	0.957	-4.071	3.853
VIX						
L1.	2036.975	2446.559	0.830	0.405	-2758.192	6832.142
L2.	3330.159	3047.696	1.090	0.275	-2643.215	9303.532
L3.	-6707.311	3330.315	-2.010	0.044	-1.32e+04	-180.013
L4.	1765.020	2333.832	0.760	0.449	-2809.208	6339.247
S&P500						
L1.	50.891	89.225	0.570	0.568	-123.987	225.769
L2.	-0.770	109.461	-0.010	0.994	-215.310	213.770
L3.	-44.636	95.619	-0.470	0.641	-232.045	142.774
L4.	-6.374	64.583	-0.100	0.921	-132.955	120.207
Google Trends						
L1.	-7.822	15.458	-0.510	0.613	-38.118	22.474
L2.	100.222	30.870	3.250	0.001	39.718	160.726
L3.	-78.073	53.781	-1.450	0.147	-183.482	27.337
L4.	9.549	41.140	0.230	0.816	-71.084	90.181
P(Gold)						
L1.	-173.734	154.868	-1.120	0.262	-477.270	129.802
L2.	168.277	242.405	0.690	0.488	-306.827	643.381
L3.	-541.470	248.214	-2.180	0.029	-1027.961	-54.979
L4.	518.727	176.594	2.940	0.003	172.610	864.845
_cons	54263.490	17515.130	3.100	0.002	19934.460	88592.520

The VAR output displayed above shows very similar results to the ones observed looking at Bitcoin's spread. Both sentiment indicators' third lag appears to significantly explain variations of the daily Bitcoin volume. These correlations are significant at the 10% significance level (p-value = 0.067) for the polarity indicator while being significant at the 5% level relative to the TBT indicator (p-value 0.041). The sign of the correlation factor is however confusing as the two indicators appear to move in opposite directions. In fact, the polarity indicator predicts a negative correlation while the TBT indicator predicts a positive one.

Concerning the magnitude of these effects, the polarity indicator reports a coefficient of -4,360 for its third lag, meaning that for a one unit increase in the polarity ratio, the daily trading volume is expected to decrease by 4,360BTC on the third day following the ratio increase, ceteris paribus. TBT's coefficient indicates that when the daily total number of tweets increases by 1000, it is expected that the daily BTC trading volume is going to increase by 1057BTC consequently, on the third day following the TBT increase, ceteris paribus. The volatility of both sentiment indicators is uncorrelated with the future variation of the BTC trading volume, for all four lagged values.

The VAR results of both sentiment indicators are represented, once again, through their respective IRF described through Figure 11. First of all, the IRFs exhibits the opposite effect of the sentiment indicators on Bitcoin's liquidity measure of breadth. On the one hand, TBT's third lagged-value shows a significantly positive effect on Bitcoin's trading volume, in comparison with its initial trading volume. Assuming an daily increase of 1000 tweets, it is estimated, by the IRF, that the daily trading volume will increase by approximately 1000BTC (technically slightly more) on the following third day. Therefore, a positive shock in TBT will probably decrease the liquidity risk associated with holding Bitcoins through a rise in Bitcoin's market breadth. The IRF, however, converges back to 0 on TBT's fourth and last lagged-value. On the other hand, a one-unit positive shock in the tweets' sentiment polarity is projected to have an exclusively negative impact on Bitcoin's future trading volume, measured in terms of BTC. In fact, the IRF (right panel) appears to be almost constantly decreasing along time. It is expected that Bitcoin's daily trading volume will decrease by more or less 6,000BTC, compared to its initial volume, by the third to fourth day following the sentiment indicator's shock. As a result, a positive shock in the tweets' sentiment polarity is statistically estimated to consequently decrease Bitcoin's market breadth and therefore increase its liquidity risk.

Figure 11: Impulse Response Function of Bitcoin's Daily Trading Volume (BTC)
Figure 11 graphs the expected future fluctuations of Bitcoin's trading volume following a one-unit increase in the TBT indicator (left) and in the tweets' polarity (right). The x-axis, labelled as steps, indicates the number of days following the one-unit shock in the sentiment indicators. The y-axis reports the variation of Bitcoin's daily trading volume, in comparison with its initial level. The blue line represents the IRF and the grey area its 95% confidence interval.



The next step of this analysis is to perform a Granger causality test in order to discover what social sentiment effect is predominant between the positively correlated TBT sentiment indicator and the negatively correlated polarity indicator. The result of the Granger causality test is displayed in Table 13.

Table 13: Granger causality Wald test for the specification relative to Bitcoin's Trading Volume, in BTC
Table 13 exhibits the result of the Granger causality test. The Granger causality test estimates whether excluding a variable from the equation significantly decrease the model's predictive power. The Granger causality test tests the null hypothesis stating that there is no causative effect of a variable on Bitcoin's future trading volume. A p-value below the 0.1, 0.05 and 0.01 threshold means that this null hypothesis can be rejected at the 10%, 5% and 1% confidence levels. The first two columns 'indicates the two variables between which causality is being tested. The third column displays the estimated 'Chi-Square' statistic for each causality test, the fourth column indicates the test's corresponding degree of freedom and the last column report the test's p-value.

Equation	Excluded	chi2	df	Prob>Chi2
volumeBTC	Polarity	7.550	4	0.110
volumeBTC	TBT	9.690	4	0.046
volumeBTC	Polarity vol	2.399	4	0.663
volumeBTC	TBT vol	6.666	4	0.155
volumeBTC	VIX	7.202	4	0.126
volumeBTC	SP500	1.187	4	0.880
volumeBTC	Google Trends	21.546	4	0.000
volumeBTC	P(Gold)	11.890	4	0.018
volumeBTC	ALL	91.295	32	0.000

The Granger causality test that TBT is the only variable having a causative effect on the daily BTC trading volume. Indeed, the Granger causality test failed to reject the null hypothesis, assuming that the polarity indicator has no causative effect on the BTC volume, despite the test delivering a p-value very close to the 0.1 threshold (p-value = 0.11). The TBT indicator, however, saw his causative effect on the future BTC volume being confirmed by the Granger causality test, at the 5% significance level (p-value = 0.046). Going back to the discussion on the effect of social sentiment on Bitcoin's breadth, since only the TBT indicator was estimated to have a causative effect on the dependent variable, it is expected that an increase in investors' social sentiment is going to consequently increase the subsequent BTC trading volume which in turn decreases the liquidity risk associated with holding Bitcoins. This result is in accord with the findings relative to the polarity indicator, discovered in the tightness analysis.

7. Chapter VII: Discussion of the results

7.1. Conclusion over the Hypotheses

Table 14: Summary of the findings relative to each hypothesis

Table 14 exhibits a combination of the main findings and results observed in this research, relative to each individual hypothesis. The first column enunciates each one of the five hypotheses. The second column reports the sentiment variable over which conclusions are being drawn. The third column briefly describes the results provided by the Vector Autoregressive model (VAR). The fourth column reports the results of the Granger Causality test, performed for each individual VAR model. Finally, the last column transcribes the main findings provided by the Impulse Response Functions, describing Bitcoin's expected future financial movements following a shock in the sentiment variables. Colour Code: Green indicates that the hypothesis is accepted for both sentiment indicators, Yellow indicates that only one of the two sentiment indicators can confirm the hypothesis, Red indicates that the hypothesis can be rejected.

HYPOTHESIS	TYPE	VAR	GRANGER	IRF
H1: Social Sentiment impacts Bitcoin's future price	TBT	TBT significantly predicts Bitcoin's future price the first three days. TBT's volatility also negatively influences Bitcoin's price on the subsequent day.	The Granger Causality test CONFIRMS TBT's causative effect.	The IRF shows a relatively negative effect of TBT over time despite peaking high on the second subsequent day after the shock.
	Polarity	A shock in the sentiment polarity significantly and positively impacts Bitcoin's future price on the following third day.	The Granger Causality test CONFIRMS the polarity indicator's causative effect.	The IRF shows that Bitcoin's price will tend to continuously move upward, following a positive shock in the indicator.
H2: Social Sentiment impacts Bitcoin's future price volatility	TBT	TBT is a significant predictor of Bitcoin's future volatility with a two to three days time-horizon. TBT's volatility is significantly and negatively correlated with Bitcoin's volatility on the next day.	The Granger Causality test CONFIRMS TBT's causative effect.	Bitcoin's volatility is expected to oscillate several times around 0. Its volatility reaches its minimum value after two days and its maximum one after three days.
	Polarity	The sentiment polarity's third lag is negatively correlated with Bitcoin's future volatility.	The Granger Causality test CONFIRMS the polarity indicator's causative effect.	Bitcoin's price volatility is expected to decrease following the polarity indicator's positive shock.
H3: Social Sentiment impacts Bitcoin's future trading volume (in US Dollars)	TBT	TBT's second lagged value and its volatility's first three lagged values are all significant predictors of Bitcoin's future trading volume.	The Granger Causality test REFUTES TBT's causative effect.	The IRF predicts a relatively negative effect over time of TBT's shock on Bitcoin's future trading volume.
	Polarity	The polarity indicator is irrelevant in predicting Bitcoin's future volume.	The Granger Causality test REFUTES the polarity indicator's causative effect.	The IRF highlights the polarity indicator's poor explanatory power on Bitcoin's future trading volume (in $).
H4A: Social Sentiment impacts Bitcoin's future liquidity through its market tightness	TBT	A shock in the TBT indicator generates significant fluctuations on the second and third days following that shock. Similarly, TBT's volatility predicts future tightness after three to four days.	The Granger Causality test CONFIRMS TBT's causative effect.	TBT's one-unit shock is likely to generate multiple oscillations around Bitcoin's initial tightness level.
	Polarity	A shock in the sentiment polarity positively impacts Bitcoin's market tightness on the subsequent day.	The Granger Causality test REFUTES the polarity indicator's causative effect.	The IRF shows a clear negative effect over time of the polarity shock over Bitcoin's future bid-ask spread.
H4B: Social Sentiment impacts Bitcoin's future liquidity through its market breadth	TBT	Only TBT's third lag positively and significantly correlates with Bitcoin's future trading volume, in terms of BTC.	The Granger Causality test CONFIRMS TBT's causative effect.	TBT exhibits a positive influence on the future trading volume through a peak on the third day following TBT's shock.
	Polarity	A shock in the polarity indicator is likely to have a negative effect on Bitcoin's trading volume, but only three days after that shock.	The Granger Causality test REFUTES the polarity indicator's causative effect.	A positive shock in the sentiment polarity appears to consequently decrease Bitcoin's BTC trading volume over time.

Table 14 aims to gather altogether the findings described in the result sections relative to each one of the social sentiment indicators. Concerning the First Hypothesis, the tweets' polarity helps explaining variations in Bitcoin's future prices but only through its third lag. Specifically, the tweets' polarity is positively correlated to Bitcoin's price three days following the change in polarity. If the polarity ratio increases on one day, it is expected that a corresponding increase in Bitcoin's price will occur three days following the ratio change. Overall, the tweets' polarity tends to significantly and positively impact Bitcoin's future price trend, as confirmed by its Impulse Response Function. As a result, the sentiment polarity analysis appears to confirm the intuition lying behind the First Hypothesis. Additionally, a rise in the TBT sentiment indicator (i.e. an increase in the daily total number of tweets) Granger causes a drop in Bitcoin's price on the first day followed by a larger increase on the second day before suffering from another drop on the third day. This pattern was adequately described by the VAR's corresponding Impulse Response Function, predicting considerable price oscillations, on the following second and third days, consequent to TBT's one-unit shock. Considering both sentiment indicators' significance, the First Hypothesis can be accepted without much doubt.

Considering the Second Hypothesis, the second VAR model demonstrated that an upward change in the polarity ratio is likely to granger cause a decrease in Bitcoin's future volatility, on the third day following the change in polarity ratio. The relationship between the polarity indicator and Bitcoin's future volatility is this time negative, in contrast with what had been observed relative to the First Hypothesis. Furthermore, TBT's effect on Bitcoin's future volatility is also significant since a rise of the daily number of tweets posted is expected to significantly reduce Bitcoin's volatility on the second day, before increasing it on the third day following the tweets' number increase. It is hard to assess with certainty the overall impact of the TBT sentiment indicator since both effects have similar magnitudes and its corresponding IRF reports multiple oscillations around 0. However, it is certain that TBT has, to a certain extent, a significant influence on Bitcoin's future volatility. In conclusion, the Second Hypothesis can be accepted through the paired significance of both sentiment indicators in predicting Bitcoin's future volatility movements.

The third hypothesis, relative to Bitcoin's daily trading volume measured in US Dollars, can be rejected when considering all the information available about Twitter's social sentiment explanatory power on Bitcoin's future volume. The different Granger causality tests rejected the causative impact of the sentiment variables despite TBT and its volatility having significant coefficients when observing in the VAR results. As a result, this research could not establish a significant relationship between the sentiment indicators and Bitcoin's future trading volume.

Relative to the liquidity risk, analysed through hypothesis 4a and 4b, the polarity sentiment indicator exhibited a negative effect on Bitcoin's bid-ask spread, through its first lag, and a negative influence as well on the BTC trading volume, but this time through its third lagged value. In other words, a rise in the social sentiment polarity will likely impact Bitcoin's market in two different ways: the market will become tighter but it will experience a decrease in its breadth as well. However, both of these effects on the liquidity risk associated with holding Bitcoin were refuted by the Granger causality test which failed to prove causative influence of the polarity indicator on the future liquidity measures. Nonetheless, the IRF still shows significant impact of the sentiment polarity indicator in the longer-term. In fact, the IRF shows that both Bitcoin's bid-ask spread and its daily BTC trading volume will significantly lie below their initial levels three to four days following the positive shock in sentiment polarity. The volatility of the polarity ratio did not provide any significant explanatory power in predicting any of the Bitcoin's financial indicators.

TBT also predicts variation in Bitcoin's future liquidity risk. Indeed, Bitcoin's breadth, proxied by the daily BTC trading volume, is positively correlated with TBT's third lag. In other words, an increase in the daily number of tweets talking about Bitcoin is likely to increase the daily BTC trading volume on the third day following the increase of the number of tweets. Relative to Bitcoin's tightness, proxied by its bid-ask spread, the total number of tweets predict a strongly significant decrease of the bid-ask spread on the second day after the increase of the number of tweets followed by a similar, but less significant, increase of the spread on the third day following the rise of tweets. The predictive explanatory power of the TBT measure was revealed to 'Granger cause' the fluctuations of both liquidity proxies. To summarise TBT's explanatory power on Bitcoin's future liquidity, a significant rise of the number tweets posted daily will improve Bitcoin's liquidity risk since it is likely to tighten Bitcoin's market and to thickened its breadth as well. All in all, as the social sentiment increases, the liquidity risk associated with holding Bitcoins decreases and vice versa. Both of these effects can clearly be observed through their respective IRFs. As a result, we can therefore accept both liquidity hypothesis but only for the TBT sentiment indicator. On the other hand, the polarity indicator failed to be granted the support of the Granger causality test, despite showing significant movements along time through their IRFs.

This research provides support to the findings made by Philippas et al. (2019), who analysed TBT's predictive as well on Bitcoin's future short-term returns. An increase in the daily number of tweets with the hashtag '#Bitcoin' significantly increase Bitcoin's short-term return and volatility. This thesis' contribution is to extent the significance of Twitter's influence on Bitcoin's future by confirming the significance of the proxy for tweets' sentiment polarity on its future volatility and returns. The latter observation is in accordance with what had been observed by Abraham et al (2018) who also pointed out the relevance of social sentiment measures, based on a linguistic analysis, on Bitcoin's future. In contrast to the study conducted by Kaminksy (2016), the present research did not establish any notable relationship between the social sentiment variables and Bitcoin's trading volume, measured in terms of US Dollars. Lastly, concerning Bitcoin's future liquidity, no previous articles ever mentioned, to the best of my knowledge, to have successfully used the TBT measure to predict Bitcoin's future liquidity risk. The outstanding aspect of this thesis is to describe through which mechanism is Bitcoin's liquidity impacted by the Twitter's social environment and popularity, as described above.

In conclusion, the TBT sentiment indicator appears to be a much more reliable predictor of future market movements in comparison with the tweets' polarity sentiment indicator. Indeed, either the TBT indicator itself or its volatility provide explanatory power to each one of the VAR models. Both indicators, however, provide great insights on how the social sentiment affects Bitcoin's price and volatility.

7.2. The predictive impact of the control variables

This section will discuss Bitcoin's financial behaviour in comparison with the three control variables, S&P500, VIX and Gold Price which aim to capture the worldwide financial situation. For instance, despite the sentiment indicators being very relevant in predicting Bitcoin's future prices, none of the control variables mentioned above provided any significant explanatory power over Bitcoin's future price movements. However, two interesting observations can be made when observing the impact of the control variables.

Firstly, both the VIX and the S&P500 index appear to be highly correlated with Bitcoin's future volume and volatility. These findings confirm the ones made by Mac Donell (2014) who found a negative relationship between the VIX and Bitcoin's price. In this study, no significant relationships were established between Bitcoin's price and the VIX but the latter variable clearly has a role to play on Bitcoin's future trading activity. In fact, as Mac Donell hypothesised, the activity on the Bitcoin market appears to be enhanced in periods of low overall volatility on the stock market. Investors therefore partially switch their assets towards Bitcoin in order to increase the expected return of their portfolio during periods of low volatility within the stock market. The effect is reflected in Bitcoin's daily trading volume and daily price volatility which appear to be both negatively correlated to the VIX, on average. The VIX was demonstrated to Granger cause fluctuation in both Bitcoin's trading volume (in US Dollar) and its daily price volatility, at the 1% significance level. The S&P500 index, itself, can also explain future variations in Bitcoin's market activity. Similarly, an increase in the value of the index appears to be correlated with a lower future activity on Bitcoin's market through a lower daily price volatility and a lower trading volume (in US Dollar). The reasoning to explain this phenomenon can be put in parallel with the VIX effect. During periods when returns decrease on the stock market, investors tend to switch their assets towards Bitcoin or the other way around when stock returns increase. The significance of this effect is however less strong, in comparison with the VIX effect, since most coefficients and the Granger causality tests are significant at the 5% level. Both variables are however insignificant in explaining variations in Bitcoin's liquidity measures.

Secondly, the following paragraph will discuss the explanatory power of Gold's Price over Bitcoin's financial indicators. Gold Price is used in the VAR model to control for the behaviour of the most common safe-haven asset, with which Bitcoin is usually compared. Gold price appears to be uncorrelated to Bitcoin's classical market movement measures (i.e. Price, Volatility and trading Volume). However, Gold price becomes the strongest predictor of Bitcoin's future liquidity. Indeed, all four lags of the Gold price variable are significant in predicting Bitcoin's future bid-ask spread. Furthermore, Gold price is also a relevant factor in predicting Bitcoin's future trading volume (measured in BTC). Both effects were confirmed by two Granger causality tests through which it was possible to reject the null hypothesis assuming that the price of gold does not have any causative impact on the liquidity measures. The reason(s) why the price of gold is correlated with the future liquidity measures is (are) unclear but it is a fact that Gold price affect Bitcoin's liquidity risk. Further research on Bitcoin's liquidity could investigate this interesting correlation.

7.3.Google Trends: Another measure of social sentiment

Thorough this study, Google Trends was used as a control variable since its significance in predicting Bitcoin's (or any stock's) future prices had already been established in previous researches (Choi Hal Varian, 2009 ; Bouoiyour et Selmi's, 2014). It is however still possible to describe its predictive influence estimated by the VAR models described in this study.

Out of all the variables observed, Google Trends exhibits the highest predictive power about Bitcoin's market movements since the Google Trends variable showed significant coefficients in all five VAR model specifications. Moreover, Google Trends had been demonstrated to Granger cause variations in Bitcoin's future financial indicators variables at the 1% significance level, for all five model specifications.

Specifically, the first specification of the VAR model, estimating the levels of correlation between Google Trends' four lags and the price of Bitcoin at time t, highlighted an overall positive correlation between the two variables. Indeed, without much surprise, an increase in the number of search queries concerning Bitcoin will generate a rise in Bitcoin's price consequently. Only the first lagged value of the Google Trends variable predicts such effect, at the 1% significance level. It is worth noting that the second lag is negatively correlated with Bitcoin's future price but with an inferior coefficient, in absolute value, which is only significant at the 10% significance level. As a result, it can be concluded that there exists a positive relationship between Google Trends and Bitcoin's future price, in the subsequent period.

Furthermore, Google Trends is also relevant in predicting Bitcoin's future volume and volatility but this time its overall effect is harder to distinguish. In fact, an increase in the total number of search queries significantly (at the 1% significance level) increases both Bitcoin's volume and volatility during the subsequent day. However, on the third and fourth day following the increase in search queries, both Bitcoin's volume and volatility are suffering from a relatively similar decrease suggesting a short-term effect of the Google Trends variable on these financial indicators.

Google Trends was shown to also significantly impact Bitcoin's future liquidity but Bitcoin's liquidity reaction following an increase in its search queries is more difficult to assess. On the one hand, the Bitcoin's bid-ask spread is positively correlated with the second lagged value of the Google Trends variable. In other words, an increase in the number of search queries concerning Bitcoin is likely to reduce, in the following two days, Bitcoin's tightness and therefore increasing its liquidity risk. On other hand, Google Trends is also positively correlated with Bitcoin's future trading volume (measured in BTC). It means that the same increase of search queries is likely to increase Bitcoin's breadth within two days and therefore improve Bitcoin's liquidity. Google Trends therefore exhibits opposing effects on Bitcoin's liquidity risk and no conclusions can be made relative to the relationship between these two variables.

In conclusion, the behaviour of the Google Trends variable is somewhat similar to what was observed in the analysis of TBT's predictive power.

7.4. Robustness check using Bitcoin's daily returns

In order to verify the relevance of the sentiment variables to be used as predictors for Bitcoin's future price, a robustness check had been implemented regressing Bitcoin's daily return instead of daily price. The results shown in Table 14 are going to be put in parallel with the results relative to Bitcoin's price, in section 6.2.

The first lagged value of the polarity indicator appears to be positively correlated to Bitcoin's daily return. This correlation is significant at the 5% significance level. The implication of this result for Bitcoin's price can be translated as follow: if the ratio of positive tweets to negative tweets increases by one unit, Bitcoin's return, on the subsequent day, will increase by 0.5%, ceteris paribus. None of the following three lags significantly impact Bitcoin's daily returns. The polarity indicator's volatility also exhibits explanatory power on Bitcoin's future daily returns through its first and second lagged values. Specifically, if the volatility of the investors' sentiment polarity increases by one unit, Bitcoin's daily return is likely to decrease by 0.23% on the subsequent day before showing a reversing effect on the second day through a 0.25% increase in Bitcoin's daily return, ceteris paribus. These effects are significant at the 5% significance level and at the 10% significance level, respectively.

With regards to the TBT sentiment indicator, it appears that there exists a positive relationship between the two variables. In fact, TBT's second lagged valued is highly significant, at the 1% level, with a coefficient of 0.138. The interpretation of the coefficient can be done as follow: if the daily total number of tweets posted (with the '#Bitcoin' mention) increases by 100,000 from one day to the other, it is expected that Bitcoin's daily return will increase by 13.8%, ceteris paribus, on the second day following the TBT increase. Despite TBT being positively correlated to Bitcoin's daily future returns, TBT's weekly volatility failed to significantly explain future variation in Bitcoin's daily returns.

Both sentiment indicators, TBT and polarity, showed their relevance in predicting Bitcoin's future price through the realised robustness check. The same conclusion cannot be made for the Google Trends variable, for instance, which was relevant in the first specification but failed to significantly explain variations in this latter model.

Table 14: VAR results for Bitcoin's daily returns

Table 14 exhibits the VAR results relative to the robustness check. Every variable's impact is defined through four lags, labelled as L1, L2, L3 and L4. The first column 'Coef.' reports each lagged-variable's estimated coefficient on Bitcoin's future price. The second column, 'Std.Err.", indicates the data observation's average standard deviation from the estimated coefficient. The third column 'z' reports all coefficient's 'z' statistic while the fourth column indicates its corresponding p-values. The last two columns represent the estimated coefficient's '95% confidence interval' based on the z statistic.

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
DailyReturns						
DailyReturns						
L1.	-0.167	0.048	-3.510	0.000	-0.260	-0.074
L2.	-0.052	0.054	-0.960	0.339	-0.158	0.054
L3.	0.057	0.055	1.040	0.300	-0.051	0.164
L4.	0.036	0.061	0.590	0.556	-0.083	0.155
Polarity						
L1.	0.005	0.002	2.030	0.042	0.000	0.009
L2.	-0.001	0.002	-0.380	0.704	-0.006	0.004
L3.	-0.003	0.002	-1.320	0.186	-0.008	0.002
L4.	-0.001	0.002	-0.420	0.673	-0.005	0.003
TBT (per 100k tweets)						
L1.	-0.040	0.027	-1.450	0.146	-0.093	0.014
L2.	0.138	0.041	3.360	0.001	0.057	0.218
L3.	-0.083	0.053	-1.570	0.116	-0.187	0.020
L4.	-0.029	0.031	-0.930	0.352	-0.091	0.032
Polarity1weekvol						
L1.	-0.023	0.011	-2.030	0.043	-0.046	-0.001
L2.	0.025	0.014	1.730	0.084	-0.003	0.053
L3.	0.002	0.015	0.110	0.913	-0.028	0.031
L4.	-0.003	0.012	-0.250	0.801	-0.028	0.021
TBT1weekvol						
L1.	0.000	0.000	0.360	0.722	-0.000	0.000
L2.	0.000	0.000	0.470	0.637	-0.000	0.000
L3.	0.000	0.000	0.710	0.480	-0.000	0.000
L4.	-0.000	0.000	-1.250	0.210	-0.000	0.000
VIX						
L1.	-0.002	0.003	-0.920	0.359	-0.007	0.003
L2.	0.002	0.003	0.690	0.487	-0.004	0.008
L3.	0.001	0.003	0.150	0.883	-0.006	0.007
L4.	-0.000	0.002	-0.160	0.874	-0.005	0.004
SP500						
L1.	-0.000	0.000	-1.570	0.116	-0.000	0.000
L2.	0.000	0.000	1.050	0.294	-0.000	0.000
L3.	0.000	0.000	1.010	0.312	-0.000	0.000
L4.	-0.000	0.000	-1.040	0.298	-0.000	0.000
Google Trends						
L1.	0.000	0.000	1.020	0.308	-0.000	0.000
L2.	0.000	0.000	0.820	0.412	-0.000	0.000
L3.	-0.000	0.000	-0.510	0.612	-0.000	0.000
L4.	-0.000	0.000	-0.760	0.445	-0.000	0.000
PGold						
L1.	-0.000	0.000	-0.140	0.892	-0.000	0.000
L2.	0.000	0.000	0.050	0.958	-0.000	0.001
L3.	-0.000	0.000	-0.450	0.651	-0.001	0.000
L4.	0.000	0.000	0.670	0.501	-0.000	0.000
_cons	-0.005	0.016	-0.310	0.759	-0.036	0.026

8. Chapter VIII: Limitations

Every research comes with its limitations. First of all, the sentiment variables can eventually be subject to endogeneity. Indeed, only a few variables are used as control variables and there are probably a multitude of variable which truly have an impact on Bitcoin's future financial indicators. It is therefore a possibility that the significant test results obtained, confirming the predictive power of one or several sentiment indicator(s), is the result of an unobserved variable, highly correlated with the sentiment variables. The consequent endogeneity of the unobserved variable on the sentiment indicators is likely to considerably increase the probability to face the 'omitted variable' bias.

Second, the VAR models analysed in this study are insensitive to policy change or to drastic changes in the cryptocurrency market. These macroeconomic changes generate instability and this macroeconomic instability is hardly captured and/or understood by the VAR model. As a consequence, the absence of control for these macroeconomic changes in the economy increases the probability to face misspecification which might in turn bias the coefficients of each variables' lagged values.

Next, VAR models are not able to observe within-period shocks. While this limitation makes sense when the VAR model uses weekly, monthly or quarterly data, it should be less of a concern in the study at hand as daily observations are being used. However, considering the fact that markets can respond quite fast to social media activity, there might exists some unobservable correlation between the social sentiment at time t and Bitcoin's financial indicators an hour later.

Furthermore, as mentioned in the first section of the result chapter, the number of lags used in the VAR model can lead to biases in the statistical analysis. Specifically, using four lags decreased the degrees of freedom making significant correlation, from one day to the other, more difficult to observe statistically and might also generate multicollinearity among the variables. On the other hand, using such few lags may create room for specification error as mentioned previously. This limitation was however considered when choosing the optimal number of lags using the three measures: BIC, HQIC and AIC.

Nevertheless, the proxy capturing investors' social sentiment polarity is also highly subject to biases through its the linguistic analysis. In fact, the current program used in this research associates a sentiment to different dictionaries of words and base the tweet's sentiment rating based on the presence or the absence of such words. Therefore, the program is unable to distinguish different forms of humour, such as irony, which will even confuse the algorithm and influence it in the wrong direction. It is also worth noting as well that the vocabulary used to talk about Bitcoin drastically changed along the years. As a result, the algorithm might be more or less prone to identify a tweet's sentiment during the dataset's earlier times in comparison with its later times. Moreover, the algorithm used records a relatively high rate of neutrality among the analysed tweets. It is the consequence of a relatively high proportion of tweets which do not contain any words from neither of the two sentiment lists or eventually, but at a lesser extent, an equal amount of words issued from each one of the dictionaries.

Last but not least, as mentioned in the Data chapter, all variables appeared to be highly skewed as both their natural values and their corresponding logarithmic transformation failed to report a p-value higher than 0.000 when the normality of their data distribution was investigated by the Skewness and Kurtosis test. In addition, the residuals of the VAR models themselves also appear to be non-normally distributed as the Jarque-Bera tests indicated, for all five models, a clear non-normal distribution of the regression's residuals. The normal distribution of the data is a common, but highly discussed, assumption of OLS models. The Vector Autoregressive model can be considered as an OLS model, to a certain extent, but should still be able to identify shocks despite non-normally distributed data, if we believe the findings provided by Lanne and Lutkepohl (2010).

In conclusion, the research and the results presented through this thesis can also be subject to different biases. However, the identified issues aforementioned do not obligatorily lead to statistical biases. In fact, The VAR model was implemented observing daily fluctuations and was optimised through its lag selection following the suggestions provided by three different indicators. Despite the sentiment classification algorithm reporting a high rate of neutrality, it can still be assumed that changes in the real social sentiment will still be reflected by changes in the sentiment polarity indicator as, on average, the bias is the same for both the daily number of positive and negative tweets which compose the polarity ratio. Additionally, the dictionaries should on average be representative of the post's global sentiment as it was built upon the relatively famous lexicon composed by Loughran and Mc-Donald.

9. Chapter IX: Suggestions for future research

First of all, further research could aim at building a predictive model, using either a VAR or ARIMA/ARIMAX model to forecast Bitcoin's financial indicators based on a set of different variables. This forecasting model should include proxies for the global state of the world economy (VIX, S&P500 and other indexes for other geographical regions), proxies for other safe-haven assets such as Gold (which has a clear influence on Bitcoin's liquidity) and most importantly must include a sentiment index (composed of different sentiment measures such as Google Trends, TBT and other linguistic analysis to determine the overall social positivity concerning Bitcoin). Several models should be developed and afterwards compared through their respective RSME.

Other techniques to grasp social sentiment should also be investigated. For instance, it could be wise to build a social sentiment index based on the positivity/negativity of the most influent Twitter accounts, such as Justin Son, Mc Afee or even Elon Musk (whose Twitter activity was an important driver of Bitcoin's most recent bull-run as well as a strong contributor to burst the recent cryptocurrency bubble). This new type of social sentiment index should be compared first to other typical social sentiment index, such as the one constructed in this study, in order to assess whether these 'influencers' exhibits similar sentiment behaviour in comparison with the population. Afterwards, it should be analysed whether the new sentiment index is a better predictor, or not, of Bitcoin's future than the classical sentiment index.

The present research highlighted interesting correlations between Bitcoin's future and different economic variables. Among them, the relation between the stock market and Bitcoin's future volatility and daily trading volume (measured in US Dollar). Both the VIX and the S&P500 index appear to be significant predictors of Bitcoin's future market activity. But which mechanism is driving these correlations. Is it due to the stock market investors looking to increase the expected return of their portfolio or is it more the consequence of a behavioural phenomenon? Furthermore, the impact of Gold prices on Bitcoin's future liquidity is a relatively surprising and interesting result. Is this unexpected result due to specification error or can gold influence Bitcoin's future liquidity? If so, what mechanism is triggering such effect?

10. Chapter X: Conclusion

Predicting stocks' movement through social media activity had always constituted a point of high interest among academics. Past researches achieved to predict future stock movements using blog dynamics, Google search queries or even through sentiment analysis, usually conducted via the social platform Twitter. Twitter had been increasingly used by investors, politics, corporations, etc. to share their opinion and beliefs. The study at hand aimed at investigating whether Twitter can be used as a forecasting tool to predict Bitcoin's future market movements. On the one hand, two indicators of social sentiment were retrieved from the social platform. First, the total number of tweets posted each day with the hashtag '#Bitcoin', referred in this research as TBT, and second, the polarity of the Twitter discussion, computed as the ratio of the total daily positive tweets to the total amount of negative tweets on that same day. On the other hand, four financial indicators of Bitcoin were considered in the analysis, namely Bitcoin's price, daily price volatility, trading volume and liquidity. Both social sentiment variables' last four lags were regressed into a VAR model in order to predict Bitcoin's price, volatility, volume or liquidity at time 't'. Once significant correlations were observed between the variables, a Granger causality test was implemented in order to confirm or refute the causative effect of the social sentiment variable on Bitcoin's market movements.

This research's main findings are the following. Firstly, Twitter appears to have a strong forecasting ability on Bitcoin's future price. The polarity of Twitter's discussion appears to be positively correlated to Bitcoin's future price through its third lagged-value, this result is significant at the 1% level and had been confirm by the Granger causality test. The other social sentiment variable, TBT, also exhibits a significant correlation with Bitcoin's future price through its first three lagged-values. However, its effect is this time more complex to determine since TBT's first lag is negatively correlated to Bitcoin's price, its second lag is positively correlated, with a coefficient twice as big, followed by the third lag which is negatively correlated again. Another interesting finding concerning the TBT indicator is that its volatility also helps predicting Bitcoin's future price variation. There exist a negative relationship between the variables meaning that an increase in the TBT indicator's volatility is likely to decrease Bitcoin's price in turn.

Secondly, the social sentiment variables also seem to explain future variations in Bitcoin's future price volatility. The polarity indicator's third lag is this time negatively correlated to Bitcoin's future price. Furthermore, the TBT indicator indicates an overall positive correlation with Bitcoin's future price volatility while the indicator's volatility exhibits a negative correlation with Bitcoin's future price volatility. Thirdly, both the polarity of Twitter's discussion and TBT are becoming meaningless in explaining the future fluctuation of Bitcoin's trading volume, measured in terms of US Dollar. However, the TBT sentiment indicator's weekly volatility is significantly and negatively correlated with Bitcoin's future trading volume. In other words, if the daily number of tweets get more volatile, Bitcoin's trading is likely to decline as a consequence.

Fourthly, the two measures of liquidity, Bitcoin's tightness and breadth, lead to mixed results. While the polarity indicator was shown to be significantly and negatively correlated to both liquidity measures, its causative impact was denied by the Granger causality test, twice. TBT, on the other hand, was proven to be a relevant factor in predicting Bitcoin's future liquidity risk. An increase in the daily total discussion about Bitcoin on Twitter is likely to decrease Bitcoin's bid-ask spread and to increase its trading volume, measured in BTC. Overall, it can be concluded that TBT is positively correlated with Bitcoin's future liquidity levels since the market tightens and its breadth gets thicker. Both the polarity and TBT indicators' volatility failed to Granger cause any variation in Bitcoin's future liquidity level.

Furthermore, a robustness check was conducted regressing Bitcoin's daily returns in function of the social sentiment variables. This procedure was implemented in order to provide support to the findings related to Bitcoin's price. It was shown that, once again, both sentiment indicators were significantly and positively correlated with Bitcoin's future returns. Google Trends had also been highly relevant to proxy for investors' social sentiment and to predict Bitcoin's market movements. It was interesting to observe how proxies for the stock market's activity (S&P500 and VIX) are relevant in predicting the future volatility and volume within the Bitcoin market. Moreover, the price of gold, despite being insignificant in most models, is highly relevant when assessing Bitcoin's future liquidity. Further investigations are required to disentangle both of these effects and offer a better explanation for these phenomena.

In conclusion, the formation of financial bubble, price trend and sudden drop are partially the cause of behavioural phenomena arising through the social media platform Twitter. Twitter is therefore a useful tool to predict Bitcoin's future and further research should aim at investigating the impact of the few most influential public figures (in the cryptocurrency world) in order to understand to which extent an individual or a group of individuals can influence and control a market as big as the Bitcoin market.

11. References

Abraham, J., Higdon, D., Nelson, J. and Ibarra, J. (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review*, 1(3).

Akbas, F., Boehmer, E., Genc, E., Petkova, R. (2010). The Time-Varying Liquidity Risk of Value and Growth Stocks. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1572763.

Ameriks, J., Wranik, T. and Salovey, P. (2009). Emotional Intelligence and Investor Behavior. *The Research Foundation of CFA Institute*.

Antweiler, W. and Frank, M. (2001). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance,* 59(3), 1259-1294.

Baker, M. and Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2).

Baur, D., Lee, A. and Hong, K. (2017). Bitcoin: Medium of Exchange or Speculative Assets?. *Journal of International Financial Markets Institutions and Money*, 18(2).

Berners-Lee, T. (1989). Information management: A proposal.

Bollen, J., Huina M., and Xiaojun Z. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science* 2(1): 1–8.

Bouoiyour, J., Selmi, R. (2014). What Bitcoin looks like?. *Munich Personal Repec*, *Archive Paper N°58091*.

Bouri, E.; Azzi, G.; Haubo Dyhrberg, A. (2016). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics Discussion Papers*, 41.

Choi, H.; Varian, H. (2009). Predicting the Present with Google Trends. *Economic Record,* 88.

Cohan, P. (2013). Are Bitcoins Safer Than Cyprus?. *Forbes Magazine*. Available online : https://www.forbes.com/sites/petercohan/2013/04/02/are-bitcoins-safer-than-cyprus/?sh=20d0094552f2.

D'Avolio, G. (2002). The Market for Borrowing Stock. *Journal of Financial Economics,* 66(2–3), 271–306.

De Choudhury, M., Sundaram, H., John, A. (2008). Can blog communication dynamics be correlated with stock market activity?. *Proceedings of the nineteenth conference of the Association for Computing Machinery on Hypertext and hypermedia*, 55–60.

Detrixhe, J. (2013). Cyprus Rating Cut Two Steps to B by Fitch on Bank-Bailout Costs. Available online: https://www.bloomberg.com/news/articles/2013-01-25/cyprus-rating-cut-two-steps-to-b-by-fitch-on-bank-bailout-costs.

Edmans, A., Garcia, D., Norli, Ø. (2007). Sports sentiment and stock returns*. The Journal of Finance*, 62(4), 1967-1998.

Folkinshteyn, D., Lennon, M. (2015). A Tail of Twin: Bitcoin and the WWW. *ISSN*, *10(2).*

Geczy, C., Musto, D., Reed, A. (2002). Stocks are special too: An analysis of the equity lending market, *Journal of Financial Economics,* 66(4).

Glaser, F., Zimmerman, K., Haferkorn, M. and Siering, M. (2014). Bitcoin: Asset or Currency? Revealing users' hidden intentions. *Twenty Second European Conference on Information Systems.* Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425247

Huberman, B., Asur, S. (2010). Predicting the Future with Social Media. *2010 International conference on web intelligence and intelligent agent technology*.

Kaminski, J. (2016). Nowcasting the Bitcoin Market with Twitter Signals. *Working paper, MIT Media Lab.*

Kent Baker, H. and Ricciardi, V. (2014). How Biases Affect Investor Behaviour. *The European Financial Review, February-March 2014, 7-10.*

Korajczyk, R. A.,Sadka, R. (2008). Pricing the Commonality Across Alternative Measures of Liquidity, *Journal of Financial Economics* 87, 45-72.

Kristoufek, L. (2014). What are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. *Plos One,* 10(4).

Kuleshov, V. (2011). Can Twitter predict the stock market?. Available online: http://cs229.stanford.edu/proj2011/Kuleshov-CanTwitterPredictTheStockMarket.pdf

Kyle, A.S. (1985) Continuous auctions and insider trading. *Econometrica*. 53(6), 1315–1336.

Lachanski, M. and Pav, S. (2017). Shy of the Character Limit: "Twitter Mood Predicts the Stock Market" Revisited. *Econ Journal Watch*, 14(3).

Lancendorfer, K., Lee, B., Chang, S. and Adelaar, T. (2003). Effects of Media Formats on Emotions and Impulse Buying Intent. *Journal of Information Technology*, 11.

Loughran, T. and McDonald, B. (2015). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 17(2).

Matta, M. and Marchesi, M. (2015). Bitcoin Spread Prediction Using Social And Web Search Media. *UMAP Workshops.*

Meech, J and R. Gu. (2014). Bitcoin - the "New Gold" for a "Safe-Haven" Investment?. *Sustainable Industrial Processing Summit.* 1(1), 215-230.

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

Ofek, E. and Richardson, M. (2003). DotCom Mania: The Rise and Fall of Internet Stock Prices. *The Journal of Finance*, 58(3).

Pak, A., Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation,* 17-23.

Philippas, D., Rjiba, H., Guesmi, K., & Goutte, S. (2019). Media attention and Bitcoin prices. *Finance Research Letters*, *30*, 37-43.

Qiang, Z. and She-u, Y. (2009). Noise Trading, Investor Sentiment Volatility, and Stock Returns. *Chinese language journal*, 29.

Salovey, P., and J.D. Mayer. (1990). Emotional Intelligence. *Imagination, Cognition and Personality*, 9(3) ; 185–211.

Shaozhi, Y., Wu, F. (2010). Measuring Message Propagation and Social Influence on Twitter. *Proceedings of the Second international conference on Social informatics*. 216–231.

Shleifer, A., Vishny, R. (1997). “The Limits of Arbitrage.” *Journal of Finance,* 52(1): 35–55.

Subramaniam, S., & Chakraborty, M. (2020). Investor attention and cryptocurrency returns: Evidence from quantile causality approach. *Journal of Behavioral Finance*, *21*(1), 103-115.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *Plos one* 10(9). doi:10.1371

Urquhart, A. (2018). What causes the attention of Bitcoin?. *Economics Letters*, *166*, 40-44.

von Wyss, R. (2004). Measuring and Predicting Liquidity in the Stock Market. *University of St. Gallen: Dissertation* N°2899.

Wurgler, J., Zhuravskaya, E. (2002). Does Arbitrage Flatten Demand Curves for Stocks?. *Journal of Business,* 75(4), 583-609.

Yermack, M. (2013). Is Bitcoin a Real Currency?. *NBER Working Paper*, 19747.

Zhu, P., Zhang, X., Wu, Y., Zheng, H., & Zhang, Y. (2021). Investor attention and cryptocurrency: Evidence from the Bitcoin market. *Plos one*, *16*(2).

Zięba, D., Kokoszczyński, R., & Śledziewska, K. (2019). Shock transmission in the cryptocurrency market: Is Bitcoin the most influential?. *International Review of Financial Analysis*, *64*, 102-125.

12. Appendix

Table 2: Matrix of correlations
Table 2 shows the correlation factors between the 11 variables cited below

Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Price	1.000										
(2) Bid-Ask Spread	-0.319	1.000									
(3) VolumeBTC	0.064	-0.004	1.000								
(4) Volume$	0.800	-0.200	0.417	1.000							
(5) Daily Returns	0.054	-0.051	-0.002	0.019	1.000						
(6) Daily Price Volatility	0.446	-0.075	0.134	0.499	-0.024	1.000					
(7) Google Trends	0.591	-0.163	0.459	0.830	0.049	0.366	1.000				
(8) Polarity	0.179	-0.041	-0.170	0.033	0.110	0.018	-0.043	1.000			
(9) TBT	0.643	-0.215	0.308	0.660	0.049	0.367	0.648	0.080	1.000		
(10) positive ratio	0.638	-0.375	-0.069	0.454	0.046	0.229	0.351	0.402	0.294	1.000	
(11) negative ratio	0.289	-0.214	0.163	0.322	-0.107	0.159	0.329	-0.536	0.141	0.400	1.000

Table 3: Lag selection: Selection-order criteria AIC, HQIC and BIC　　Number of observations = 2474

Model 1:

lag	LL	LR	df	p	FPE	AIC	HQIC	BIC
1	6140.510	35896	36	0.000	0.000	-4.930	-4.894	-4.831
2	6304.010	327	36	0.000	0.000	-5.033	-4.967	-4.850
3	6452.170	296.310	36	0.000	0.000	-5.124	-5.027	-4.856*
4	6557.550	210.750	36	0.000	0.000	-5.180*	-5.052	-4.827
5	6633.740	152.39*	36	0.000	2.2e-10	-5.112	-5.054*	-4.775

Model 2:

lag	LL	LR	df	p	FPE	AIC	HQIC	BIC
1	341.443	24080	36	0.000	0.000	-0.241	-0.206	-0.143
2	647.456	612.030	36	0.000	0.000	-0.459	-0.393	-0.276
3	862.648	430.380	36	0.000	0.000	-0.604	-0.506	-0.336*
4	993.883	262.470	36	0.000	0.000	-0.680*	-0.553	-0.329
5	1093.980	200.19*	36	0.000	1.9e-08	-0.632	-0.574*	-0.296

Model 3:

lag	LL	LR	df	p	FPE	AIC	HQIC	BIC
1	387.310	26147	36	0.000	0.000	-0.278	-0.243	-0.180
2	690.886	607.150	36	0.000	0.000	-0.494	-0.428	-0.311
3	963.859	545.950	36	0.000	0.000	-0.685	-0.588	-0.418
4	1144.190	360.660	36	0.000	0.000	-0.801	-0.674	-0.450*
5	1251.380	214.38*	36	0.000	1.7e-08	-0.859*	-0.700*	-0.423

Table 3: continuation

Model 4.1:

lag	LL	LR	df	p	FPE	AIC	HQIC	BIC
1	2084.220	26924	36	0.000	0.000	-1.646	-1.611	-1.548
2	2307.690	446.950	36	0.000	0.000	-1.797	-1.731	-1.615
3	2513.860	412.320	36	0.000	0.000	-1.935	-1.838	-1.667*
4	2651.750	275.790	36	0.000	0.000	-2.017*	-1.889	-1.665
5	2743.440	183.37*	36	0.000	5.1e-09	-2.002	-1.903*	-1.626

Model 4.2:

lag	LL	LR	df	p	FPE	AIC	HQIC	BIC
1	445.274	25527	36	0.000	0.000	-0.325	-0.289	-0.227
2	738.876	587.200	36	0.000	0.000	-0.533	-0.466	-0.350
3	985.987	494.220	36	0.000	0.000	-0.703	-0.606	-0.436
4	1156.650	341.330	36	0.000	0.000	-0.811	-0.684	-0.460*
5	1259.050	204.79*	36	0.000	1.7e-08	-0.865*	-0.707*	-0.429

Appendix 1: Extracting daily Google Trends data:

AVAILABLE ON THE FOLLOWING GITHUB LINK ADDRESS:
https://github.com/PhilippeCodes/Getting-Long-Term-Daily-Google-Trends-Data.ipynb/blob/master/Getting%20Long-Term%20Daily%20Google%20Trends%20Data.ipynb?fbclid=IwAR1vVMhzgSAvQA7nSpmSeONaN0UkR75ky7P1T_Zks2Mz0GW9AVYdZpOllLU

```python
# importing dependencies
import pytrends
from pytrends.request import TrendReq
import pandas as pd
from datetime import date, datetime, timedelta
start_date= date(2014, 1, 1) # specify your start date
end_date= date(2021, 4, 30) # specify your end date
key_word = 'Bitcoin' # use one key word
_cat = 0 # Category to narrow down your results
_geo = '' # Two letter country abbreviation/default: ALL
_gprop = '' # What Google property to filter to (e.g 'images')
_hl = 'en' # Specify Language and Region
_tz = 360 # specify your time-zone
def perdelta(start, end, delta):
  curr = start
  while curr < end:
    yield curr
    curr += delta
dates=[]
for res in perdelta(start_date, end_date, timedelta(days=90)):
  dates.append(res)
dates.append(end_date)
appended_data = []
for i in range(len(dates)-1):
  try:
    _timeframe = str(dates[i]) + ' ' + str(dates[i+1])
    totalTrend = TrendReq(hl=_hl, tz=_tz)
    totalTrend.build_payload([key_word], cat=_cat, timeframe=_timeframe, geo=_geo, gprop=_gprop)
    totalTrend = totalTrend.interest_over_time()
    appended_data.append(totalTrend)
  except KeyError:
    print('Please specify the Parameters (e.g. Keyword)')
    break
for i in range(len(appended_data)-1):
  x = appended_data[i][key_word].tail(1).values
  y = appended_data[i+1][key_word].head(1).values
  if x == 0 and y == 0:
    factor = 1
  elif x == 0:
    factor = 0.5/y
  elif y == 0:
    factor = x/0.5
  else:
    factor = x/y
  appended_data[i+1][key_word] = appended_data[i+1][key_word] * factor
appended_df = pd.concat(appended_data, axis=0)
appended_df = appended_df[~appended_df.index.duplicated(keep='first')]
appended_df.to_csv('daily_gtrends.csv')
```

Appendix 2: Lists of positive and negative words

```
positive_labels = (
        '🚀',
        'moon',
        'whale',
        '100.000',
        '100000',
        '1.000.000',
        '1000000',
        'buy',
        'Buy',
        '🌔',
        'pump',
        '🔥',
        '📈',
        'bullish',
        'hold',
        'vision',
        'million',
        'rebound',
        'dream',
        'grow',
        'enthusiast',
        'profit',
        'gold',
        'miracle',
        'mooning',
        'bought',
        'fan',
        'boost',
        'mining',
        'rise',
        'mining',
        'boost',
        'auction',
        'ATM',
        'push',
        'adopt',
        'accept',
        'succes',
        'succeed',
        'long',
        'luv',
        'love',
        'Luv',
        'Love',
        'LOVE',
        'like',
        'likes',
        'get started',
        'interesting',
        'interested',
```

```
    'launch',
    'appreciate',
    'gain',
    'miner',
    'payment option',
    'go',
    'GO',
    'Go',
    'benefit',
    'support',
    'payment',
    'reinvent',
    'great',
    'amazing',
    'amazed',
    'future',
    'rising',
    ':)',
    ':-)',
    ';)',
    ';-)',
    'strong',
    'achieve',
    'accomplish',
    'assure',
    'attract',
    'benefit',
    'compliment',
    'delight',
    'empower',
    'enhance',
    'exceptional',
    'excite',
    'good',
    'high',
    'innovat',
    'outperform',
    'perfect',
    'popular',
    'positive',
    'prestige',
    'progress',
    'strength',
    'success',
    'worthy',
    'valuable'
)

negative_labels = (
    'crash',
    'scandal',
    'damage',
    'bubble',
    'burst',
    'short',
```

'sell',
'trash',
'dump',
'fuck',
'negative',
'scam',
'low',
'fall',
'hack',
'fell',
'ban',
'sink',
'revers',
'decline',
'bearish',
'risk',
'global',
'warming',
'tax',
'drop',
'📉',
'crisis',
'rob',
'stolen',
'sale',
'stop',
'bottom',
'attack',
'warning',
'shit',
'hate',
'dislike',
'crime',
'criminal',
'dishonnest',
'illegal',
'laundry',
'laundering',
'inconvenient',
'rules',
'regulations',
'rift',
'unsafe',
'not safe',
'dissappoint',
'disappoint',
'danger',
':/',
':-/',
':(',
':-(',
'weak',
'worse',
'worst',
'worry',

'wrong',
'worthless',
'vulnerable',
'unsustainable',
'unstable',
'unreliable',
'unsuccessful',
'unsure',
'unsafe',
'unprofitable',
'unnecessary',
'unfair',
'tragic',
'threat',
'stress',
'spam',
'scam',
'shrink',
'shut',
'severe',
'shock',
'seize',
'risk',
'reject',
'refuse',
'provoke',
'punish',
'protest',
'problem',
'poor',
'peril',
'penalty',
'panic',
'overvalue',
'overestimate',
'neglect',
'negligence',
'mistake',
'misprice',
'mislead',
'los',
'lag',
'lack',
'inefficient',
'incorrect',
'inconsistent',
'illicit',
'fraud',
'flaw',
'forbid',
'fail',
'downturn',
'doubt',
'discredit',
'delist',

```
        'deficit',
        'default',
        'decline',
        'dead',
        'dangerous',
        'critical',
        'criminal',
        'corrupt',
        'broke',
        'bankrupt',
        'abuse',
        'abandon'
    )
```

Appendix 3: Code relative to the tweets' linguistic analysis

```
import pandas as pd
import timeit

def remove_tweet(df, exclude_words, batch_no, chunksize):

    texts = df['text']

    i = batch_no*chunksize
    exclude_index = []
    for text in texts:
        words = str(text).split(" ")

        for word in words:
            if word in exclude_words:
                exclude_index.append(i)
                break
        i=i+1

    df = df.drop(exclude_index, axis = 0)

    df = df.drop(columns="#")
    #dff.to_csv('df.csv')
    return df


def count_occurency(df, positive_labels, negative_labels):

    texts = df['text']

    nbr_pos_lab_list = []
    nbr_neg_lab_list = []
    tweet_pos_lab_list = []
    tweet_neg_lab_list = []

    pos_tweet_counter=0
    neg_tweet_counter=0

    pos_lab_counter=0
    neg_lab_counter=0
```

```
	for text in texts:
		words = str(text).split(" ")
		curr_pos_lab_counter=0
		curr_neg_lab_counter=0
		for word in words:
			if word in positive_labels:
				curr_pos_lab_counter+=1

			if word in negative_labels:
				curr_neg_lab_counter+=1

		if curr_pos_lab_counter > curr_neg_lab_counter:
			pos_tweet_counter+=1
			tweet_pos_lab_list.append(1)
			tweet_neg_lab_list.append(0)
		elif curr_pos_lab_counter < curr_neg_lab_counter:
			tweet_pos_lab_list.append(0)
			tweet_neg_lab_list.append(1)
			neg_tweet_counter+=1
		else:
			tweet_pos_lab_list.append(0)
			tweet_neg_lab_list.append(0)


		nbr_pos_lab_list.append(curr_pos_lab_counter)
		nbr_neg_lab_list.append(curr_neg_lab_counter)
		pos_lab_counter += curr_pos_lab_counter
		neg_lab_counter += curr_neg_lab_counter

	df = df.drop(columns="text")
	df = df.assign(positive_words = nbr_pos_lab_list)
	df = df.assign(negative_words = nbr_neg_lab_list)
	df = df.assign(pos_sentiment = tweet_pos_lab_list)
	df = df.assign(neg_sentiment = tweet_neg_lab_list)

	#print(f'Number of positive tweet : {pos_tweet_counter}')
	#print(f'Number of negative tweet : {neg_tweet_counter}')
	#print(f'Number of positive words : {pos_lab_counter}')
	#print(f'Number of negative word : {neg_lab_counter}')

	return df


def global_count(df_):

	df = df_.groupby('date').sum()

	df = df.rename(columns={"retweet": "total_retweet", "like": "total_like", "positive_words":
"total_positive_words", "negative_words": "total_negative_words", "pos_sentiment":
"total_pos_sentiment", "neg_sentiment": "total_neg_sentiment" })

	df.to_csv("total_tweet.csv")

def remove_wrong_date(df):
```

```
        dates = df['date']
        exclude_index = []

        i=0
        for date in dates:

                date_array = str(date).split("-")
                if len(date_array) != 3:
                        exclude_index.append(i)
                elif len(date_array[0]) != 4 or len(date_array[1]) != 2 or len(date_array[2]) != 2:
                        exclude_index.append(i)
                i=i+1

        df = df.drop(exclude_index, axis = 0)

        return df

if __name__ == '__main__':

        positive_labels = (
                APPENDIX 2
          )

        negative_labels = (
                APPENDIX 2
          )

        exclude_words = (
                'giveway',
                'GIVEWAY',
                'Giveway',
                'giving away',
                'Giving away',
                'GIVING AWAY',
                'free',
                'Free',
                'FREE',
                'Collect',
                'collect',
                'receive',
                'easiest way to get Bitcoin',


        )

        start = timeit.default_timer()

        chunksize = 100
        batch_no = 0
        frames = []


        for chunk in pd.read_csv('mining.csv', sep='\t', chunksize=chunksize):
                df = remove_tweet(chunk, exclude_words, batch_no, chunksize)
```

```
        df = count_occurency(df, positive_labels, negative_labels)
        frames.append(df)

        batch_no+=1
        if batch_no >161:
            break

    print(batch_no*chunksize)

    df = pd.concat(frames, ignore_index=True)

    df = remove_wrong_date(df)

    global_count(df)



    stop = timeit.default_timer()

    print('Time: ', stop - start)



    global_count(df)
```

Appendix 4: TBT webscraping code :

https://www.reddit.com/r/datasets/comments/mnf66v/bitcoin_tweets_chart_data_extraction_for/

```
let csvcont = "data:text/csv;charset=utf-8,"; d.rawData_.forEach(function(rowArray) { let dateVal = new Date(rowArray[0]); let tweetVal = rowArray[1]; csvcont += (dateVal.getDate() + '/' + dateVal.getMonth() + '/' + dateVal.getFullYear() + ', ' + tweetVal) + "\r\n"; }); let encodedUri = encodeURI(csvcont); window.open(encodedUri);
```