Università
Ca'Foscari
Venezia

Master's Degree programme - Second Cycle
in Economics and Finance

Final Thesis

—

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Investor Sentiment in the Cryptocurrency Market

SUPERVISOR
Prof. Caterina Cruciani

CANDIDATE
Trung Dang
Student Number 877263

ACADEMIC YEAR
2019 / 2020

Acknowledgments

Foremost, I would like to express my sincere gratitude to one of the most brilliant and critical minds that I've ever had a chance to work under, my thesis supervisor - Professor Caterina Cruciani. Thank you so much for your continuous encouragement, invaluable guidance, and tremendous dedication. I appreciate every precious hour you spent reviewing and improving my work. Without you, the accomplishment of this thesis would otherwise have remained a castle in the air.

I also wish to thank my family for their unconditional love and support. Thank you for always believing in me and giving me the inner strengths to achieve even the most ambitious goals. Mom, thank you for sparking my curiosity for mathematics ever since I was a child. Without this curiosity, I would never come this far.

Contents

1 Introduction

The thesis investigates the influences of investor irrationality in the market of an alternative asset class called cryptocurrencies. In recent years, this market has captured significant attention from financial practitioners and academia due to its notorious volatility.

From the perspective of innovation management, market expectations on these digital money systems are likely to suffer from public *irrational hypes*, as they possess many characteristics of *disruptive innovations*. The challenges arise when one attempts to test this hypothesis empirically. The first issue is to quantify the *market expectations* of cryptocurrencies. This could be solved easily since we proceed to use their market prices. Another issue, which appears to be more challenging, is to find the quantitative measures for the *forces of hype* that are potentially behind the public irrationality. The field of study innovation management could not provide a solution to this problem. Thus, a shift in perspectives is a must in order for the problem to be tackled.

It turns out that researchers in finance have suspected a somewhat similar factor of emotion existing on the market scale. They called this factor *investor sentiment*. Empirical studies to test the existence and possible effects of *sentiment* in financial markets have flourished during the Dot-com bubble. This incident of excessive investor speculation also originated from a disruptive technology at that time - the Internet. In fact, if we see from the viewpoint of *Gartner Hype Cycles,* the cryptocurrency today appears at a very similar stage as the Internet during the late 1990s. This coincidence further inspires the author to delve into the studies on *sentiment* in the stock market in order to find if there is any method of measuring sentiment applicable to the case of cryptocurrencies.

After the literature exploration, the author is inspired to follow a method of measuring sentiment, referred to in the current study as the *composite* approach. The rationale behind this choice is that although the pre-mentioned approach is advantageous and used by many studies on sentiment in the stock market, it has not been adopted by any study conducted in the cryptocurrency market. Therefore, if the thesis could measure sentiment effectively through this approach, it could contribute considerably to the existing literature.

Regarding the *composite* approach, we first begin by collecting several indicators of sentiment. Those indicators could be the sentiment information extracted from millions of messages posted on social media or discussion platforms using sentiment analysis, or they could also be some popular indicators that measure market performance. Initial analysis shows that the individual indicators are strongly correlated with each other, thereby making it conceptually appealing to extract a common component that could be interpreted as the final *composite* index of sentiment (thus comes the term *composite approach*). In case the construct of the ultimate sentiment index is successful, we may proceed to further study how it interacts with the cryptocurrency prices.

Concerning the scope of the study, the current thesis chooses to examine the relationship between sentiment and cryptocurrency prices on the scale of the whole market, instead of focusing on a subset of several cryptocurrencies. The sampling data is available in daily frequency and collected for a prolonged[1] time from November 2014

[1] This duration could be considered "long" given the fact that the first cryptocurrency Bitcoin is only traded on public exchanges from 2013.

to July 2020.

The later analysis finds significant evidence that investor irrationality drives crypto investors in the short term, which is consistent with the theoretical beliefs set out in the theoretical chapters. Interestingly, the study even shows that the newly-created index is also a good predictor of cryptocurrency market returns, indicating that behavioral biases might play a significant role in the decision-making process of cryptocurrency investors.

With this in mind, the thesis is organized as follows. Chapter 2 begins our discussion from a perspective of innovation management, in which the author provides a general introduction on cryptocurrencies and the rationales why it is reasonable to suspect these digital money platforms might suffer from public irrational hypes. Chapter 3 shifts our perspective into finance and performs a comprehensive review of the literature on sentiment and its influences in both the stock market and cryptocurrency market. Chapter 4 discusses why several sentiment indicators are chosen and how the data is collected. In chapter 5, a popular signal extraction technique called *Principal Component Analysis* is employed to create the *composite* index. Time-series analysis is then performed to see how the new sentiment index interacts with the market returns of cryptocurrencies. Chapter 6 concludes the thesis by summarizing the analysis's main results and suggesting their potential implications for various parties, including the investors, the policymakers, and the developers of cryptocurrency.

2 Background And Motivation

2.1 Premise

This chapter presents the motivation and theoretical background of the study. In section 2.2, we first explore the technology behind cryptocurrencies and how they resemble the term *disruptive innovation*, first defined and analyzed by Bower and Christensen (1995). Section 2.3 introduced the *Innovation Hype Cycle* theory assuming that disruptive technologies often have their *market expectations* mainly driven by *irrational hypes* during the very early stages of their diffusion process. The section then sets forward a similar assumption in the context of the cryptocurrency market and further discusses the difficulties encountered during the empirical testing process. Among those difficulties, the one considered most problematic is to find a plausible measure for the *market hypes* that are potentially behind the public's unreasonable expectations. The chapter concludes by emphasizing the need to shift our perspective from innovation management into finance in order to solve the pre-mentioned challenge.

2.2 An Introduction Of Bitcoin And Cryptocurrencies

2.2.1 The Birth Of Bitcoin

On October 3, 2008, in the midst of the financial crisis, the *Emergency Economic Stabilization Act* was finally passed by the 110^{th} U.S. Congress and signed into law by the U.S. President George W. Bush. The act created a *Troubled Asset Relief Program* to rescue the country's financial system from its collapse, with an initial cost estimated at $700 billion of taxpayers' money. A few weeks later, on October 31, an author under the pseudonym[2] *Satoshi Nakamoto* (2008) released the technical paper of a digital money platform called *Bitcoin*. The platform is clearly not the first digital currency in the history since several other systems had already existed for a long time before it, since at least the 1990s, for instance, eCash (original idea published in the paper by David Chaum in 1983; officially launched in 1995), or E-gold (introduced by Gold & Silver Reserve Inc. in 1996). However, Bitcoin was designed in a very unique way that essentially separated it from all the other previous digital currencies. Because of its unique principles of design, the system was later widely considered as one of the breakthroughs in the making of true Internet money. But what exactly makes Bitcoin so special?

To answer this question, it is necessary to take a step back and define a major issue of digital payment systems, which is the so-called *double-spending problem*. The problem, by the name itself, is a potential flaw in a payment scheme whereby the same amount of money can be spent more than once. It turns out that Bitcoin and its precursory digital systems solved this problem in two essentially different ways. The common solution to double-spending, adopted by traditional payment systems and all digital currencies that pre-dated Bitcoin, is to introduce a trusted party who keeps a record of all transactions. This central party is responsible for checking every single transaction for double-spending. In the case of several transactions originated from one issuer pointing at the same amount of money, whichever transaction announced by the authorized party as the first to arrive will become the valid one, while the rest are

[2] A fictitious name used by an author to conceal his or her real identity.

considered invalid and rejected. While this solution worked for most cases, it still suffers from the inherent weaknesses of a Trusted Third Party based model. Due to their centralized nature, the participants of these systems are prone to the failures that do not originate from themselves, but rather from the central third party. These failures are either caused by the central party unintentionally (since the party could have several deficiencies in its essence) or intentionally (as a result of conflicts of interest between the central party and other users). Furthermore, the fate of the entire money system depends on the well-functioning of the party that is in charge. If that authorized party is prevented from operating, the whole system will be shut down. Typical examples could be mentioned as eCash's dissolution as its operator DigiCash declared bankruptcy (Pitta, 1999), or how E-gold, once second only to PayPal in the online payment industry, was ultimately shut down along with its mother company by the U.S. government (Zetter, 2009).

Seeing the intrinsic fragility in the centralized systems, Satoshi Nakomoto stated his pure desire to build a decentralized version of Internet money in one of his emails to Dustin D. Trammel, an initial contributor of the Bitcoin project (Franco, 2014):

> *"I think there were a lot more people interested in the 90s, but after more than a decade of failed Trusted Third Party based systems (Digicash, etc.), they see it as a lost cause. I hope they can make the distinction, that this is the first time I know of that we're trying a non-trust-based system."*

In a digital money system without a central party, to prevent double-spending, transactions must be completely announced among all network participants (Nakamoto, 2008). In other words, instead of a central party keeping the entire record of transactions, each participant has to keep a copy of the record and acts as a validator by himself. The tricky part is that how such a network could ensure participants reach a state of full consensus[3] on a single record of transactions. The records are easy to tamper with and if each participant has his own version of the record, the entire network becomes unusable. This problem did not have a solution until Satoshi Nakomoto proposed the first answer in the introductory paper of Bitcoin. In the most fundamental way, Bitcoin was designed so that transactions happening on its network could be validated using cryptographic proofs instead of trust (Nakamoto, 2008). In order to validate a number of transactions, several participants of the Bitcoin network will join a competition of decoding a very complicated (but feasible) cryptographic puzzle that incurs significant expenses (in the case of Bitcoin, computing power and electricity). The competitor who first cracks the puzzle and uploads his proofs to the network will become the only validator of those transactions. This mechanism ensures that the system is secure and harmonized as long as honest participants cumulatively control more computing power than any cooperating group of deceptive participants. Moreover, the Bitcoin system rewards newly created coins to the decoding competition's winner in exchange for the costs incurred by the process. This acts as incentives for those who participate to stay honest. Imagine a situation where a specific participant collected more computing power than all the honest participants (which incurred very significant expenses), he would have to choose between using it to keep collecting more new coins rewarded by the systems than everyone else combined, or to defraud people through double-spending which eventually undermines the network and the value of his previous coins. Between these two options, the participant ought to

[3] A state of absolute agreement among the network participants.

find it more profitable to stay honest and play by the rules. As a result, for the first time in history, a distributed payment system like Bitcoin could achieve a state of consensus without the need for a central authority.

In addition to that, Bitcoin requires no central server as it is operated on a peer-to-peer basis where every participant from a global scale can leave or join at will. This means that the entire network can only be shut down if all of the network participants are prevented from contacting each other, which could only be attained under very extreme incidents, such as a collapse of the worldwide internet. Furthermore, no central storage is involved as all transactions are publicly announced and distributed to every participant of the network, meaning that failures of a single node would not affect the performance of the entire system. Altogether, these features, namely no central authority, no central server, no entry barrier, and no central data storage, are combined flawlessly in one single system and thus allowed this system to become the first purely decentralized digital currency where coins could be transferred directly from user to user and not through any intermediary.

2.2.2 More Than Just An Invention

So far in this section, we have seen as the digital currency solved a problem in computer science that was thought to be unsolvable for many years. However, no matter how brilliant an invention might be, if it could not tackle real-world issues then that invention is simply useless. A common skeptical view is that our society does not need a decentralized money system like Bitcoin. We have been developed fairly well without such a system in the past and might not need one to further advance in the future. This should be the case if the conventional banking system keeps performing safe and sound across the world. However, the next few paragraphs attempt to discuss this matter from a different perspective. By showing that there still exist several deficiencies in the traditional system, the study argues that Bitcoin could still be needed by many parts of the world, from the developing countries to the developed ones.

Before beginning with the developing countries, let's take an overall look at a global scale. In a report compiled by a group of World Bank researchers, it is estimated that worldwide, around 1.7 billion adults remained unbanked in 2017 (Demirguc-Kunt et al., 2018). The same report conducted a survey to ask adults without a bank account the reasons why they choose to not have one. The barriers most commonly cited by the respondents included not having enough money to open an account (65% of answers), high operating fees (26% of answers), far distance to the banks (22% of answers), lack of necessary documentation (20% of answers). Survey participants stating these reasons mostly came from developing countries such as Zimbabwe, Nigeria, Colombia, Peru, Kenya, Philippines, Indonesia, etc. All of the reasons pointed to a fact that the central parties of these countries' banking systems might have created, perhaps, too high entry barriers for their customers. High entry barriers result in many businesses and entrepreneurs staying unbanked and undocumented. De Soto (2000) has estimated that this situation caused over $10 trillion of the world's wealth, approximately 8.5% of the worldwide assets at the time of his analysis, to be locked up[4] inside these underbanked countries. On the one hand, in a traditional trust-based model, a good reason for high entry barriers could be that there have to be some mechanisms to limit new participants as allowing a new member into the club increases

[4] Kept with reduced value due to uncertainty of ownership and decreased ability of lending/borrowing (higher interest rates, higher transaction fees).

the risks for everyone else (T. B. Lee, 2013). Hence, while imposing high entry barriers could lead to expensive consequences, it is inevitable in a centralized system. On the other hand, because transaction authentication in a decentralized system like Bitcoin is not based on trust among its participants, there's no need to restrict access to the network, resulting in negligible entry barriers: no entry restriction, no operation fees, trivial transaction costs, etc. With Bitcoin or similar systems, the citizens of underbanked nations could perform financial transactions on an international scale with relatively small fees, compared with doing transactions through their traditional banks and get access to more financial services such as payment, lending, fundraising, etc. This has the potential to set free the locked-up $10 trillion so that these funds could be re-invested for further capital gains. Empirical evidence from various studies, such as Beck et al. (2006), Ellis et al. (2010), and Boldbaatar & Lee (2015), have already shown that more financial accessibility generally has a positive impact on economic development. Especially, Boldbaatar & Lee (2015) found that increases in financial access indicators lead to higher marginal increases in the economic growth of developing countries than those of developed countries. All things considered, one could argue that undeveloped countries with inefficient banking infrastructure could likely become the ideal markets for the general adoption of a decentralized digital currency like Bitcoin.

There also exist sociopolitical demands for money systems that are not controlled by any central authority in developed countries. While centralized systems worked most of the time, the world's history is full of events where they have failed. These events often happened due to wrong decisions made by central parties during crucial times. An excellent example could be the hyperinflations where central banks used their discretion to print as much money as they desired to respond for a crisis, resulting in excessive and out-of-control price increases in the economy. Most infamous hyperinflations through the history include Hungarian notorious inflation in 1945-1946 (every 15 hours for prices to double in the worst month), Zimbabwe in 2007-2008 (every 24.7 hours for prices to double in the worst month), Germany in 1922-1923 (every 3.7 days for prices to double in the worst month), China in 1947-1949 (every 5.3 days for prices to double in the worst month), Armenia in 1993-1994 (every 12.5 days for prices to double in the worst month), and the most recently, Venezuela from 2016-present (every 2.6 days for prices to double in the worst month). Data for all hyperinflations except for the one of Venezuela are collected from the study of Hanke & Krus (2012). Another good example is the so-called financial crises where systemically important institutions made serious losses due to their own mistakes and then had to cover the losses using support from the public money, for example, the global financial crisis of 2007-2008, the Greek sovereign debt crisis of 2009-2018, the Cypriot financial crisis of 2012–2013, etc. Such events (hyperinflations, financial crises, and events with similar themes) usually resulted in significant losses of public trust for the parties that are in charge of the economy (Chuen, 2015). Interestingly, perhaps in an unsurprising way, the previous World Bank survey also cites distrust in the financial system as one of the most common reasons why people stay unbanked with 16% of all respondents globally, however, the share is more than twice as high for respondents from a developed region like Europe. Because a portion of people in developed countries do not trust their banks, they might need a different system of transferring values that do not involve a central third party, such as Bitcoin.

2.2.3 Disruptive Cryptocurrency

In the last sub-section, we have seen that demands for Bitcoin genuinely existed and they arise from every corner of the world. According to Maranville (1992), when a newly-created product is associated with true demands, it will not be considered solely as an invention, but also innovation. Following this philosophy, it is reasonable to say that Bitcoin is an indispensable innovation. But, to examine more deeply about the decentralized currency's characteristics, a question that might be worth answered is which exact type of innovation Bitcoin could be categorized into. Before answering this, it is sensible to first consider a theoretical framework for classifying innovations that is very popular among the tech industry's practitioners. The framework, first proposed by Bower & Christensen (1995), classified innovations into two general categories: sustaining innovation and disruptive innovation. On the one hand, sustaining innovation is the improvement of an existing product based on the known needs of current customers. A great example of sustaining innovation is the current market of smartphones, computers, TVs, or other technological devices. Every year, companies offer new, better products with extra features and performance improvements, for instance, TVs with slimmer screens, smartphones, and laptops getting faster and lighter every few years, so on and so forth… On the other hand, disruptive innovation refers to a new product or service that has the potential to transform products that are historically very expensive and sophisticated. This type of innovation generally starts as a reduced and inferior version of costly and complicated products. While the disruptive product cannot satisfy the most demanding customers, it manages to meet the requirements of the very low-end customers, in a way that is much more affordable and accessible than its alternatives. Due to the low cost and simplicity, a larger audience could get involved with the product and new applications will constantly emerge. People gradually figure out how to use the product to perform tasks that previously required devices that cost multiple times as much. Up till a point, the disruptive product can even meet the requirements of the most demanding customers, which is when it finally creates a new market that replaces existing ones (often with bigger scales). Classic examples of *disruptive innovation* are personal computers (replaced mainframes and minicomputers), smartphones (replaced PCs and laptops), personal printers (replaced offset printing), video streaming services (replaced video rentals), Wikipedia (replaced traditional encyclopedias), etc.

Between the two categories of innovation: sustaining and disruptive, Bitcoin appears highly similar to the latter one. On the one hand, the digital currency looks inferior, in many ways, to the market that it has the potential to disrupt (i.e. the conventional banking/payment systems, as we have seen previously). For instance, as of August 2017, Visa, currently the largest global payment network, is capable of handling approximately 65,000 transactions per second (Visa, 2017). This number is significantly higher than the one allowed by Bitcoin which is between 3.3 and 7 transactions per second (Croman et al., 2016). Visa and Mastercard are also accepted at millions of locations around the world, while only less than 20,000 merchants accept Bitcoins worldwide as of August 2020 (Coinmap, 2020). Conventional banks also offer services that are not available with Bitcoin, such as automatic bill-paying, credit cards, deposit to earn interests, etc. On the other hand, interesting applications keep popping up around Bitcoin. There existed casinos, spot or derivatives exchanges, retailers, hedge funds, and much more that are based on Bitcoin (T. B. Lee, 2013). Also, there have been thousands of other digital currencies that were built upon the fundamental constituent of Bitcoin - the distributed ledger technology. As a result, these systems have several common elements with Bitcoin, such as the utilization of cryptographic proofs in order to achieve network consensus, or the use of native tokens as a way to incentivize participants to follow the network rules in the absence of a central authority. On a side note, because these currencies are secured by strong cryptography, there comes the term *cryptocurrency*. Each cryptocurrency is developed to serve a distinct purpose of its creator, but a majority of them have one thing in common - the decentralized nature. Following the same rationales as when we discuss how Bitcoin could potentially disrupt the banking and payment industry, these cryptocurrencies also have the same potential to disrupt many existing industries, including financial services, educational activities, transportation, cloud storage and computing, healthcare, gambling, etc. For example, the second-most-popular cryptocurrency after Bitcoin, Ethereum, is not only a digital money platform but also offers a novel feature called decentralized smart contracts. In essence, these smart contracts are essentially computer programs designed to automatically execute actions (without a third party) based on the events that are pre-defined in a contract between two certain parties (Savelyev, 2016; Tapscott & Tapscott, 2016). One of their fastest-growing use cases is decentralized finance (DeFi) in which financial instruments, such as bonds or loans, can be offered directly between the issuer and the buyer, without the need of any intermediary. This application has the potential to change dynamically how the finance industry works according to the argument of Schroeder (2020).

2.3 Hypothesis Development

In the last section, we discussed why Bitcoin - the first cryptocurrency and its descendants should be considered as potentially disruptive innovations, based on the popular classification framework by Bower & Christensen in 1995. One interesting characteristic of disruptive technologies is they are often exaggerated, or "*hyped up*" by the mass media as the *"the next big thing that could change the world"* during the very early stages of their adoption process. Consequently, the public will respond with excessive optimism and speculate on the future of these inventions. This phenomenon has been documented by a number of practitioners and researchers, with the most popular being the Gartner Inc[5]. The company even conceptualized the phenomenon into a graphical model that is used by itself and many other practitioners in the IT industry to draw insights on new technologies and predict their future. The following section examines this model in detail and develops a hypothesis in the context of the cryptocurrency market.

2.3.1 The Gartner Hype Cycle

In 1995, one year after joining the Gartner's Research Department, an analyst named Jackie Fenn published a report[6] in which proposed a diagram of patterns that she believed most disruptive technologies have to go through as they are adopted by the general public. The diagram was later introduced to the public as *the Gartner Hype Cycle Model* and got acquired by many practitioners and researchers due to its superior explanatory power (Jun, 2012). To understand the intuition behind the model, it is worth considering some previous theories that have influenced the author during the model development process.

---

[5] A global research and advisory firm working in multiple industries, predominantly technology and finance.. A global research and advisory firm providing information, advice, and tools for companies in various industries: finance, technology, communications, supply chain, etc. Gartner has been a member of the S&P 500 index (1 of 500 largest companies in terms of market value in the U.S. market) since 2017

[6] See Fenn (1995)

At first, we begin with a theory called *the linear model of innovation*. This is one of the conventional theories that people have been using to explain the innovation process. While the linear model has been mentioned by many scholars in the field of innovation management since at least World War II (Kline & Rosenberg, 1986), its original source has never been acknowledged or cited directly. According to some authors, such as Godin (2006), the model was largely inspired by a journal article named *Science: The Endless Frontier* by V. Bush (1945). Generally speaking, the model postulated that every innovation process contains four stages, starting with research, followed by development and production, and ending with diffusion. These events are visualized as a smooth, well-behaved linear process as given in Figure 2.1. The model has been highly influential since it was widely disseminated by many academic organizations and economists who are high-level advisors to policymakers and governments (Godin, 2006). We could even see the model's strong influence in the modern business world as nearly every corporation nowadays has a division named R&D responsible for creating new products (Kline, 1985). As a consequence, people working in the science and the technology industry circa 1980 often carried a deep conception that innovation is a purely linear process (Mowery, 1983).

Figure 2.1: The Conventional Linear Model. Source: Kline & Rosenberg (1986)

The major problem with this linear model of innovation, as pointed out by different scholars, for instance, Kline & Rosenberg (1986), is that it over-simplified the innovating process. According to Kline and Rosenberg, only in an ideal world of omnisciently technical people, the process of innovating could be as smooth and linear as described by the model since every innovator could get their creations workable and optimized the first time. In reality, innovation involves complicated feedback loops and countless trials within every stage (Kline, 1985), as if the process itself is a mechanical assembly with many interconnected gears and levels moving in conjunction in order to achieve a common purpose which is the final product. Because of this highly complex and variable nature, the process cannot be linear. On the basis of this school of thought, academicians started to develop another branch of models that describe innovation as a non-linear process.

Diffusion of Innovation, developed by Rogers (1962), is among the earliest theories that belong to the branch of non-linear innovation models. The theory seeks to explain how, why, and at what rate an idea or product can gain momentum and spread through a specific social system. Rogers suggested in his theory that people react and adopt differently to an innovation. He divided the population into five distinct groups of adopters, listed in the descending order of how fast they might adopt an innovation: the innovators (most willing to take risks with new ideas), the early adopters, the early majority, the late majority and the laggards (most conservative and very skeptical to change). An interesting point with the theory is that the groups are not distributed in equal proportions and the middle groups account for the majority of the population. Figure 2.2a shown a curve with its shape reminiscent of a bell (analogously to a Gaussian density curve) representing the five groups on a graph with a horizontal axis being the people's degree of difficulty in adopting new innovations. If we take a further step by graphing the cumulative adoption rate over time, we could see an S-shaped curve as in Figure 2.2b, implying that a typical innovation's adoption process is not a linear function of time.

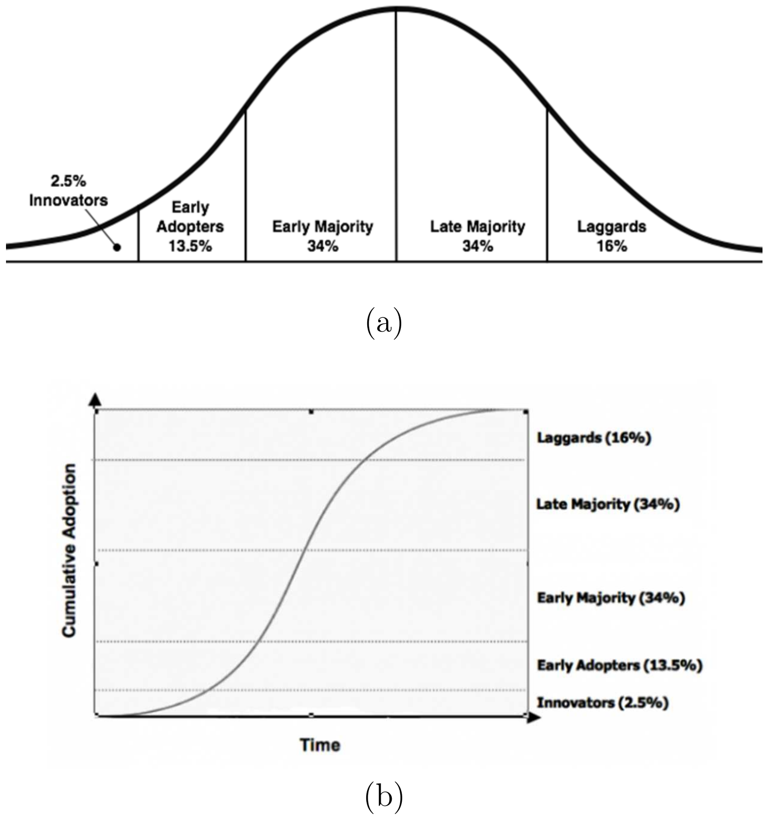


Figure 2.2: Diffusion of Innovation (panel a), Cumulative Adoption Curve (panel b)[7].

Source: Rogers (1962)

In 1991, Geoffrey Moore published a book introducing[8] an expanded version of Rogers' *Diffusion of Innovation* model. The new model, titled by its author as *Crossing the Chasm*, was designed specifically for disruptive technologies (Moore, 2014). In this model, Moore assumed that there exists a very large gap (or *chasm* in his words) to be

[7] The proportion of each group of adopters is given by the theory's author for representative purposes.

[8] See Moore (1991)

passed between the early adopters of an innovation and the early majority group (Figure 2.3). According to Moore, the visionaries (the early adopters) and the pragmatists (the early majority) possess very different expectations of a new product. For this reason, the originator of *Crossing the Chasm* considered the transition between these two groups as the most difficult stage of the entire adoption cycle. Moreover, he believed that disruptive innovations typically display most of their weaknesses at this transition step, and most of them will fail to surpass the chasm.

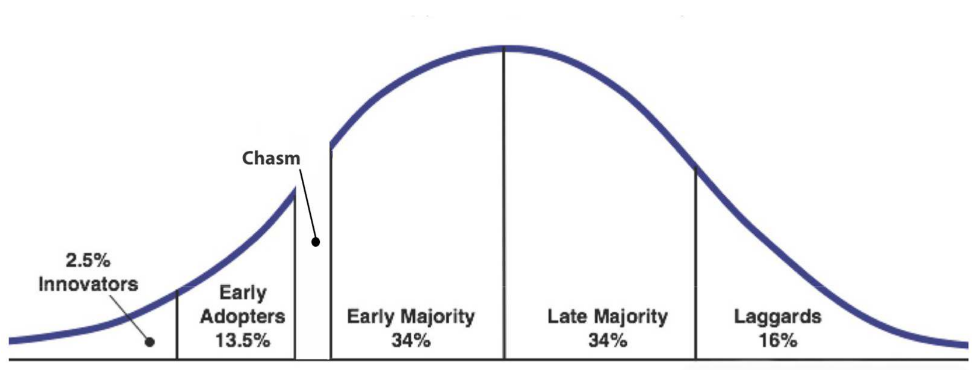


Figure 2.3: Crossing the Chasm. Source: Moore (1991)

Another non-linear innovation model that inspired Jackie Fenn to develop her *Hype Cycle* is the *Technology S-curve* introduced by Foster (1988). This model also involves an S-shaped curve as in Rogers' model (see Figure 2.4), but the curve of Foster measures an aspect that is completely different from the adoption rate, called the innovation's maturity. If adoption means the choice to acquire and utilize a new technology, the maturity of such technology can be interpreted as how developed it is, which could be measured based on characteristics such as performance, user-friendliness, reliability, costs saved, etc. The intuition behind Foster's S-curved is quite clear: a typical innovation goes through a beginning with a slow rate of improvement, then comes an acceleration phase (the steeper curve) where breakthroughs show up rapidly, and finally the rate of advance again decreases as the innovation becomes more matured and stabilized (the flattening curve).

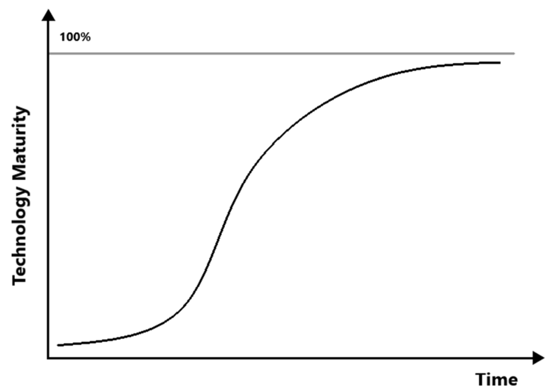


Figure 2.4: Technology S-curve. Source: Foster (1988)

Although adoption and maturity are two extremely different aspects of an innovation, the reader might notice how similar Roger's adoption curve and Foster's maturity curve look, making these very hard to be distinguished from each other. However, there actually exists a point where these curves break into different directions, which lies at the later stages of the innovation. This is often disregarded by people for one reason: Rogers' model stops explaining the adoption process once the percentage of people accepting to use the technology reaches nearly 100%. But in reality, is an innovation considered coming to an unchanged end if its adoption rate reaches 100%? (Veryzer, 1998) puts a *"No"* as the answer to this question as he believed what typically happens to a product after it has dominated the market for a long time is that some other products will eventually appear to replace it. Therefore, the adoption rate of a dominant product cannot stay high forever and will decrease at some points in time. On the other hand, the performance of a product will likely keep increasing, even at a very slow rate, in case some efforts are still devoted to the development process, and eventually stand still if people stop improving it.

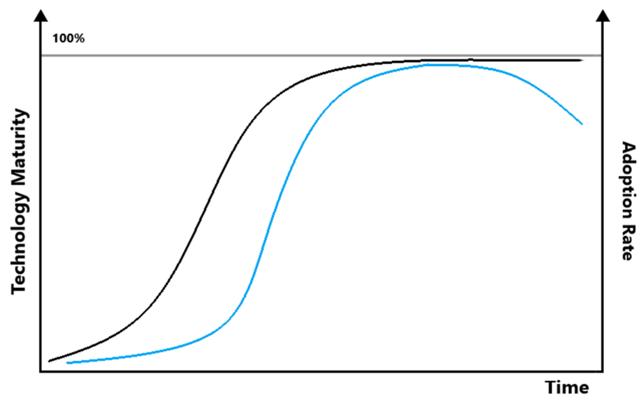
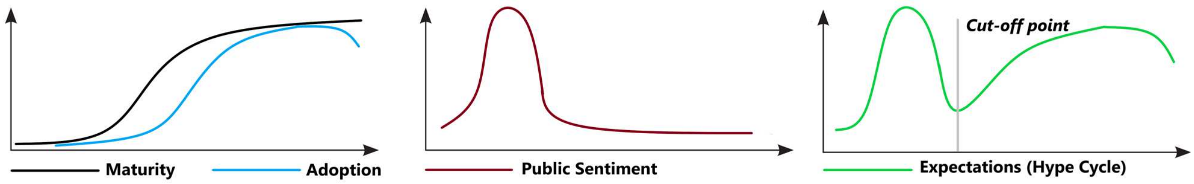


Figure 2.5: Maturity Curve (black) & Extended Adoption Curve (blue). Source: Fenn & Linden (2003)

If we interpret the difference between the maturity and the adoption of an innovation from a mathematical perspective, it could be said that the adoption rate is not an increasing function of time, while the maturity is likely to be an increasing one. To see this idea more clearly, let's take a look at the attempt of Fenn & Linden (2003) to plot the innovation's maturity curve and an extended adoption curve on the same x-axis of time (Figure 2.5). It could be seen that the adoption rate starts to drop after it reaches the highest level during the product lifecycle, portraying the belief of Veryzer that we have mentioned previously.

Now we are ready to return to the *Hype Cycle* model developed Jackie Fenn from

Gartner. In order to develop the theory, the first thing Fenn does is to inherit the main assumptions of the three models mentioned previously, namely Rogers' *Diffusion of Innovation*, Moore's *Crossing the Chasm*, and Foster's *Technology S-curve*, and then add a completely new element that allows her to describe the evolution of a disruptive technology more effectively (Fenn & Raskino, 2008). She called the patterns from those previous theories *the nature of innovation*, as visualized in Figure 2.6a. As one can notice, it is actually the graph Fenn drew together with Linden in 2003 that is shown previously in Figure 2.5. What is more important here is the new element introduced by Fenn - *the nature of human*. According to Fenn, this new element is essentially different from the conventional factor because the *nature of innovation* creates fundamental and genuine values while the *nature of human* causes exponential hype-driven expectations about such value creation. Fenn assumes that *the nature of human* follows a pattern as shown in Figure 2.6b, in which the public's irrational emotions emerge during the stage where a disruptive innovation tries to cross the *chasm* (the gap between the early adopters and the early majority). At the beginning of this stage, the new technology gets hyped up as soon as some of its potentials are noticed by the media and the public. After several success stories regarding the new product are shared, the media make unrealistic and over-optimistic predictions on its future. Enthusiasts frequently discuss the new innovation, and market observers begin to make irrational speculations. Soon enough, when the innovation enters the *chasm*: shortcomings and flaws of the product are discovered, people start to criticize and skepticize the new technology. Public sentiment is then driven to another direction: sharply downward. More importantly, Fenn believes that during this *chasm crossing* stage, not only the irrational sentiments arise powerfully, but they will also play the most important role in forming market expectations on the innovation.



(a) (b) (c)

Figure 2.6: Components of the Hype Cycle[9]: Nature of Innovation (panel a), Nature of Human (panel b), The Hype Cycle (panel c). Source: Fenn & Raskino (2008)

Despite that, Fenn further assumes that once the innovation has successfully crossed the *chasm* and begins to be accepted by the majority of the population, the effect of sentiment ceases to insignificant and the market expectations are now mainly based on the *nature of innovation* component. This idea is visualized in Figure 2.6c. As one can see, in the beginning, the expectation curve looks exactly the same as the start segments of the public sentiment curve in Figure 2.6b, while in the later stages, it is virtually identical to the maturity and adoption curves as in Figure 2.6a. The cut-off point, often known as a norm among the Gartner's analysts, lies around where the technology's adoption rate hits approximately 20% (Fenn & Raskino, 2008).

Since the origination of the *Hype Cycle* theory in 1995, the model has become extremely influential to practitioners in the technology industry due to its simplicity and superior explanatory power (Dedehayir & Steinert, 2016). More recently, many scholars have used the model to study the development paths of various kinds of innovations, such as O'Leary (2008), Järvenpää & Mäkinen (2008), Jun (2012), and many more. The key

[9] Note that the horizontal axes of all three graphs represent time.

takeaway from this theory is that expectations on disruptive innovations could be driven by hype and sentiment at the early stages of their life cycles. In some manners, this depicts the famous Amara's Law of Technology, stating that people usually overestimate the impact of a truly transformational technology in the short run and at the same time, underestimate it in the long run (Ratcliffe, 2017). As a quick recall, the reader might remember that in section 2.1, we have been argued that cryptocurrency possesses several characteristics of a disruptive innovation. Moreover, cryptocurrency is a disruptive innovation that is currently in the very first steps of its diffusion process[10]. Therefore, if one follows the assumptions of the *Hype Cycle* theory, he or she might also hypothesize that market expectations on cryptocurrencies might suffer from irrational hypes.

2.3.2 The Empirical Challenges

The important question is how exactly we could perform empirical tests on this assumption. One of the major issues is to find the quantitative measures for *market expectations* on cryptocurrencies and the *forces of hype* that are potentially behind them. This task is extremely challenging since it is not very straightforward to be measure or observe these *sort-of* abstract aspects[11].

Moreover, a limitation of the *Gartner Hype cycle*, admitted by its author Jackie Fenn

---

[10] Hays (2020) estimated that as of July 2020, approximately 40 million people are users of at least one cryptocurrency. If we compare this number with the 69% of adults globally (around 3.8 billion people) who registered an account at a bank or a payment service provider (Demirguc-Kunt et al., 2018), it could be seen that cryptocurrencies are very far from reaching the 20% adoption target set out by Gartner.

[11] Statisticians call these unobservable or latent variables.

in her book with Raskino in 2008, is that since the model tries to be applicable to all kinds of innovation - whether it is a management trend, a new business process, or a new technology, there is no standard unit of measure for the *expectations* of an innovation. Fortunately, the authors left a clue to solve this problem. According to Fenn & Raskino (2008), in case the new technology is related to some kinds of assets that are exchanged in financial markets, the *expectations* of that technology can be simply measured using the market prices of those related assets. It turns out that cryptocurrencies also fall into this category of innovations (innovation that possesses a market price). Although they are not traded in any traditional market, such as stock or commodity, cryptocurrencies have a market for themselves at least from 2010[12], and therefore, it is possible to utilize their prices for our empirical experiments later. At this stage, we only need the quantitative measures for the *forces of hype* in order to start our analysis.

Before looking for an answer to that problem, it might be interesting to take a closer look at the cryptocurrency market. According to Rauchs & Hileman (2017), this market is active 24/7 on a global scale, with thousands of cryptocurrencies being traded on multiple retail online exchanges. As of August 2020, the most dominant cryptocurrency remains to be Bitcoin, accounting for 57.91% of the total market capitalization[13] (CoinMarketCap, 2020). At the approximate price of $11,700, reached on 31st August 2020, the market value of all Bitcoins currently in circulation, is around $0.21 trillion. To put that number into perspective, there is approximately $1.98 trillion U.S. dollar

[12] On 17 March 2010, now-defunct BitcoinMarket.com (later Mt. Gox) started operating as the first cryptocurrency exchange (Sedgwick, 2018).

[13] Followed by Ethereum in the second place (at 12.81%), Ripple in the third (at 3.42%), and Tether in the fourth (at 2.69%).

and $1.38 trillion Euros in circulation today (European Central Bank, 2020; U.S. Federal Reserve Board, 2020). As some authors consider Bitcoin as the digital version of gold, such as Gkillas & Longin (2018) or Baur et al. (2018), we can also compare the market value of Bitcoin with the total value of all gold ever mined. If we price an ounce of gold at $1950[14] and use the estimate of World Gold Council (2020) on how much gold has been mined throughout history (197,576 tons), then the total value would reach over $10.72 trillion. Finally, in case we would like to put cryptocurrencies into comparison with the stock market, the world equity market capitalization has recently surpassed $95 trillion in valuation (World Federation of Exchanges, 2020). As one could see, the total value of Bitcoin and the cryptocurrency market is minor compared to the value of other mainstream markets.

Despite the relatively small valuation, the cryptocurrency market has captured significant attention from financial practitioners and academia in recent years due to one special characteristic, which is its immense volatility. Due to this characteristic, several laureates of the Nobel Memorial Prize in Economics, including Shiller (2014), Stiglitz (2017)[15], Thaler (2018)[16], Krugman (2018), and many other scholars have characterized cryptocurrencies as a speculative bubble. Although the market has displayed this behavior multiple times for the last ten years (Bouri et al., 2020), the bubble happened recently in late 2017 and early 2018 is possibly the one attracting the most attention from the media and academia. Various authors, such as Chen et al. (2019) or Kraaijeveld & De Smedt (2020), have even compared this bubble phase of

---

[14] Approximate price reached on 31[st] Agust 2020 (GoldPrice.org, 2020)

[15] See Costelloe (2017)

[16] See Wolff-Mann (2018)

cryptocurrencies specifically to the Dot-com bubble of the U.S. stock market occurred in the late 1990s. Recalling the Dot-com bubble, the event was caused by investors' excessive speculation in Internet-related companies. The late 1990s was a period of exponential growth in the use and the adoption of the Internet. Therefore, many investors misjudged the technology's potential and were hooked into a race of investing in stocks of any company that has an Internet-related prefix or a *.com* suffix. As a consequence, between the 1995 and the bubble's peak in March 2000, the Nasdaq Composite market index rose 400%, only to fall 78% from its peak by October 2002, giving up all its gains during the bubble (Lowrey, 2019). To put these numbers into perspective, let's perform a quick comparison with the cryptocurrency bubble in 2018. While the recent event in the market of digital money happened on a smaller scale[17], its decline in percentage terms (80% from the highest peak) is larger than the bursting of the Dot-com bubble in 2002 (Patterson, 2018).

2.3.3 A Shift In Perspective

It turns out that the Dot-com bubble is highly relevant to our discussion on the cryptocurrency market because it also originated from a technology that was considered extremely disruptive at the time the event happened (by the author of the term *disruptive innovation*[18]). If we think from the *Innovation Hype Cycle*'s theoretical

[17] Cryptocurrency market capitalization reached its highest point of around $830 billion on 7th January 2018 (CoinMarketCap, 2020). On the other hand, the market value of all Nasdaq companies peaked at more than $6.7 trillion in March 2000 (Gaither & Chmieleski, 2006).

[18] See (Christensen et al., 2000; Christensen & Raynor, 2003)

standpoint, the status of the Internet back then appears very similar to the state of the cryptocurrency ecosystem nowadays. What is more important is that during and after the Dot-com bubble, a large number of studies were written on the influence of irrational hypes on the stock market (the same as our current problem but targeted a different market). On the other hand, there are not many studies on the behavior of the crypto market at this moment since it is still a very young and small-sized market. Conditional on those facts, it is reasonable to first examine the studies implemented in the traditional market for stocks and try to look for some plausible ways to measure the unobservable *forces of hype.* Before moving on to discuss these in the first section of chapter 3, a quick note to the reader is that they were all conducted from financial perspectives and not from the perspective of innovation management as used in the last section. Therefore, our discussion will switch to the perspective of finance from this point. More importantly, those studies use the terminology *sentiment* in substitute for the *forces of hype* and, to ensure consistency, the study will also adopt this usage for the next chapters.

3 Methods Of Measuring Market Sentiment

3.1 Premise

The following chapter reviews a number of previous works studying the influences of *market hypes* (or *sentiment* according to their terminology) in the markets of stock and cryptocurrencies. The chapter begins by discussing why researchers in finance suspect the existence of *sentiment* and its potential effect on the market prices. The major problem with *sentiment* is that even if it truly exists and we can somehow find reliable measures for this unobservable variable, it is very challenging to empirically detect influences of sentiment on the financial markets. This is because the markets are well-known for their efficiency at eliminating mispricing opportunities. Despite the hardship, a large number of studies have emerged to solve the empirical challenges of the task. It turns out that some of those do possess feasible methods to quantify the unobservable *sentiment*. The chapter then discusses the strengths or weaknesses of such methods of measuring sentiment in detail. Besides, some results of the previous studies are also stated and commented on in this section. Subsequently, in section 3.3, we also examine a number of researches conducted recently studying the effect of sentiment on the cryptocurrency market. The chapter concludes by suggesting several directions for the later analysis.

3.2 In The Stock Market

3.2.1 Theoretical Effects Of Sentiment On Stocks

If one goes to Google Scholar[19] and performs searches for every study containing two exact words *market sentiment* and *stock*, year by year from 1980 to 2010, he could find the number of results as in Figure 3.1. It could be seen from the figure that there is an exponential growth in the number of studies related to sentiment and the stock market during and after the Dot-com bubble (meaning the late 1990s and the early 2000s). But the bubble is not the main reason that started the suspicions on an emotional factor that exists on a market scale. In fact, the doubts have arisen for quite a long time pre-Dot-com bubble, and the incident in the late 1990s itself should be considered merely a trigger for a wave of empirical studies testing if those doubts are correct. It turns out that the roots of the suspicions lie deeply in one of the most fundamental subjects in finance, the asset valuation problem.

For a better understanding, it is necessary to take a step back and discuss traditional asset pricing models. An assumption commonly made by those models is that all individual investors are fully rational and equipped with an infinite capacity of thinking. Such investors also have unlimited time and information so that they would consider every piece of relevant information until they could come up with the choice that maximizes their benefits. As a consequence, the market price will always stay at a rational level, which should be defined by the fundamental value of the asset. Moreover, any change in the market price only happens because new information arrives (which occurs in a completely random way by definition).

[19] (https://scholar.google.com) A freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines

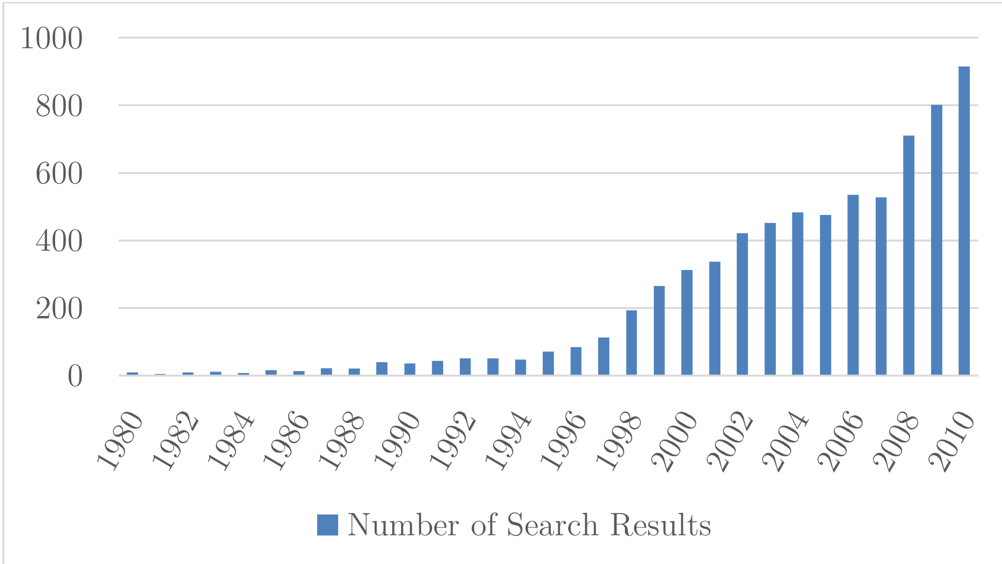


Figure 3.1: The Number of Studies Containing Two Exact Words *Market Sentiment* & *Stock* on Google Scholar (1980 - 2010). Source: Own elaboration based on data scraped from Google Scholar using the Python library BeautifulSoup

This assumption is simply unrealistic in the practical world since real investors are rarely praised for their good senses. In one of the masterpieces of his life, *The General Theory of Employment, Interest, and Money*, published in 1936, the English economist John Maynard Keynes coined the term *animal spirits* to describe how people arrive at bad financial decisions under stress and uncertainty (Lawson, 1993). Many psychologists, such as Lazarsfeld (1966) or Patch (1984), agreed that the behaviors of humans are very far from being perfect. In response, supporters of traditional models suggest that there is maybe another, more relaxed version of the assumption that could be used instead, without changing much the final results of those models. This variation assumed that although individual investors do sometimes make mistakes when they get emotional, these decision errors occur in a random way. Thus, the effects will be uncorrelated and cancel out each other so that there would be negligible impact on prices of the assets. In a sense, one could still count all investors in the market as a single rational decision-maker, which translates into no emotional factor exists on the scale of a market.

Academicians of behavioral economics, including[20] Amos Tversky, Daniel Kahneman, Richard Thaler, Robert Shiller, beg to disagree. These authors believed[21] that our human minds possess a number of flaws that make us subject to the same kinds of judgment errors. If this is true, the trader's decision mistakes should be somehow correlated with each other. As the mistakes are made, they drive the market prices farther and farther away from the assets' fundamental values. Moreover, even if imperfect psychology is not truly an issue, we have another problem regarding our social behaviors. People tend to perform activities collectively, as the legendary Greek philosopher Aristotle once stated in the magnum opus of his lifetime - the *Politiká*: *"Man, by nature, is a social animal. He who is unable to live in society, or who has no need because he is sufficient for himself, must be either a beast or a god."*. As one could imagine, people do many things together: they study, work, and play side by side. Investing is not an exception, as highlighted by Shiller et al. (1984): *"Investors spend a substantial part of their leisure time discussing investments, reading about investments, or gossiping about others' successes or failures in investing."*. When an

[20] 3 out of the 4 mentioned names are Nobel-prize winning authors. Tversky unfornately passed away six years before his collaborator Kahneman was nominated for the Nobel in 2002.

[21] See some of their highly influential works, e.g., Kahneman & Tversky (1974), Kahneman & Tversky (1979), Thaler (1985), Barberis & Thaler (2003), Akerlof & Shiller (2010), Shiller (2015).

investor sees the others' decisions before making up his own, there is a possibility that his final decisions might be actually influenced by those of others. This phenomenon, if occurring on a large scale, could lead to a multitude of correlated financial decisions. Behaviorists called the above phenomenon *herding* or *following the crowd* behaviors. *Herding* in the stock market has been heavily backed with empirical evidence by many studies, including De Bondt & Forbes (1999), Welch (2000), and Sias (2005). The latter study among the three shows that not only the individual traders and analysts suffer from *herding*, but the institutional investors are also affected by this crowd behavior. An argument to conclude this paragraph is that whether it is due to our psychology imperfections or the *herding* behaviors or a combination of both, it is possible that a large number of investors could make highly correlated decision errors while valuing assets simultaneously. The degree to which those investors misvalue the assets is regarded as *market sentiment*, or *sentiment* for short. A number of researches, including Grossman & Stiglitz (1980), Black (1986), De Long et al. (1990), Shleifer & Vishny (1997), Barberis et al. (1998), and Hong & Stein (1999), have augmented the traditional models of asset valuation with alternative models built on the formal assumption that *market sentiment* does exist.

Empirical tests of the pre-mentioned models faced two major challenges. The first is apparently the same as the problem proposed in the closing of section 2.3, which is to quantify the unobservable factor of sentiment. The second challenge, which is even more problematic, is to document real influences of sentiment on the financial markets. The reason why this challenge is so difficult to overcome is that the markets are well-known for their excellence at fixing investors' mistakes. This property of the market has been formalized into a very famous hypothesis in finance, which is the *efficient-*

market hypothesis. The hypothesis suggests that if the market price, for some reason[22], deviates from the true fair value by a degree (even by the slightest one), rational market observers will instantaneously arbitrage away the mispricing opportunities, which in turn reverts the price back to its fundamental level. This is only made possible because traditional finance assumes that arbitrage requires no capital and entails no to low risk. One more time, the behaviorists challenged this view of traditional finance. They noticed difficulties of the *efficient-market hypothesis* in fitting the patterns of prices during the events of speculative bubbles (Baker & Wurgler, 2007). These scholars suspected that arbitrages might not be unlimited as described in conventional theories, but rather are costly and risky operations in reality. Some authors, such as De Long et al. (1990) and Shleifer & Vishny (1997), turned these doubts into the formal assumptions for their behavioral asset-pricing models. According to these authors, limits of arbitrage prevent rational investors from forcing the price back to its fundamentals as aggressively as suggested by traditional finance. In other words, market corrections, instead of happening in a blink of an eye, should take place gradually over a period of time. To be more specific, behavioral pricing models, in general,[23] describe the market correction process as follows: In the short run, episodes of positive (negative) sentiment lead the price to the upper (lower) side of its fundamental level, possibly resulting in a *momentum*[24] price movement. The degree of mispricing get larger as time goes by. It also becomes clearer and clearer to a proportion of the investors that the asset might be overvalued (undervalued). In the long horizon,

[22] Due to the factor of sentiment or any other factor.

[23] See, e.g., Black (1986), De Long et al. (1990) and Shleifer & Vishny (1997)

[24] Returns are positively correlated with past returns (implying that price goes to the same direction for a period of time).

this proportion will increase to a point where those investors gather enough forces to push the price back to its fundamental level, in a movement called *reversal*[25]. The whole situation might be summarized using a common saying among investors: *"The trend is your friend... until the end when it bends."*. In case the pre-described circumstance is truly the case in reality, plus if one could define feasible measures for sentiment, then theoretically, there is a probability that the influence of sentiment on the stock market might be documented. As described above, the task is very troublesome and complicated, making this probability very small. Moreover, the task gets even more challenging at extremely short intervals because sentiment might not have enough time to build up a sufficient amount to be able to affect the prices significantly.

3.2.2 Empirical Studies On Stock Investors' Sentiment

Despite the hardship, a large number of studies have arisen to tackle the empirical challenges of the task. Many of them come after the trigger of the Dot-com bubble, as pointed out in the opening of this section. There are two aspects from these studies which should be referred to in accordance with our topic of sentiment and the crypto market.

The first is undoubtedly the empirical results. Despite the fact that the analysis in traditional markets cannot be generalized to the market of cryptocurrencies (Chen et al., 2019), it is still worth mentioning some of the results that might be relevant to our

[25] Returns are negatively correlated with past returns (implying that price is changing its direction after a period of time.)

discussion[26]. Two studies by Baker & Wurgler (2006, 2007) demonstrate that market sentiment imposes stronger effects on the category of stocks whose valuations are highly subjective. In the words of these authors, those stocks usually lack fundamental information for the purpose of valuation and are very hard to perform arbitrages. Examples are stocks of young or distressed firms, innovative stocks, unprofitable stocks, non-dividend-paying stocks, extreme growth stocks, and stocks with high volatility. Evidence from Glushkov (2006) also supports the view of Baker & Wurgler. Another interesting result from the study of Greenwood & Nagel (2009) shows that investors with less experience are more susceptible to buy assets with inflated prices during the bubble periods. On the researches concerning how sentiment influences the directions of stock prices in different investment horizons, experts generally agree with the theoretical predictions that sentiment waves generate temporary *momentum* in the short run (intraday to less than a week)[27]; and produce *reversal* movements in medium-to-long intervals (about 6-36 months)[28]. Needless to say, there still exist some controversial results, such as when Brown & Cliff (2004) claimed that their measure of sentiment has no predictive power for near-term returns, but rather the opposite, as returns can predict future sentiment. However, a little inconsistency in the results is considered understandable considering the empirical task's complexity (as anticipated from before).

Moreover, there is still a second aspect which can be learned from those studies of

[26] Shortly after, how and why these results are relevant to the crypto market will be discussed in the following part of this section. Here, to ensure a coherent presentation of ideas, they will only be mentioned for the sake of reference.

[27] See Tetlock (2007) and Tetlock et al. (2008) for periods of under-a-week. See Sun et al. (2015) and Renault (2017) for intraday intervals.

[28] See Brown & Cliff (2005) for 6-36 month intervals.

sentiment in the stock market, that is how to quantify the unobservable sentiment. In some sense, this aspect is even more valuable for our main discussion than the first one. The reason is that while it might be not sensible to apply the empirical results from the stock market on the market of cryptocurrencies, the methods of measuring sentiment are applicable in most cases. The most rigid restriction on the applicability of those sentiment indicators is perhaps their existence in the crypto market. But before getting into that, let's discuss some of the actual methods used by prior researchers to quantify sentiment in the stock market. Generally speaking, those methods could be divided into three major approaches.

The first approach is to **directly measure** the sentiment of market participants using one of the two techniques: through **surveys** or through **opinion mining**. Shiller has been one of the pioneers in collecting surveys on investors' behaviors since the late 1980s (Shiller, 1987; Shiller & Pound, 1989). Solt & Statman (1988) mentioned a popular survey-based sentiment index called the *Investors Intelligence*, although later the authors rejected its usefulness as a predictor of the direction of the stock market. A later study by Brown & Cliff (2004) also utilized the *Investors Intelligence* and another index called the *American Association of Individual Investors* as the proxy for institutional and individual investors, respectively. Qiu & Welch (2004) cited a sentiment measure based on consumer confidence surveys and further shows that this index correlates strongly with the returns of companies that are small or held disproportionately by retail investors. Overall, the survey-based sentiment measures were one of the classic picks for practitioners and researchers circa 2000. The major drawback of such measures is that they are only available at low frequencies such as weekly or monthly.

This leads to another technique of measuring sentiment directly called **opinion**

mining (also known as **sentiment analysis**). In essence, the technique refers to the computational process of identifying and categorizing opinions expressed in textual data extracted from online sources, namely social media platforms, web journals, message boards, etc. This method is becoming increasingly used in recent studies because it provides sentiment data in high frequencies such as intraday (if the input textual data are messages from social media and online message boards) or daily (if the input data are articles from online journals). It is possible that Antweiler & Frank (2004) are the initiators of this trend by applying opinion mining on more than 1.5 million messages posted on *Yahoo! Finance* and *Raging Bull* to measure the sentiment of stock investors. The two authors found that the effect of sentiment on market returns is statistically significant despite being economically small. They further discovered that sentiment is a strong predictor of market volatility. Two other studies by Tetlock (2007) and Tetlock et al. (2008) quantified financial news stories from the *Wall Street Journal* and *Dow Jones News Service* into sentiment information. Bollen et al. (2011) and Zhang et al. (2011) both measured the daily collective mood states of *Twitter* users using a large scale of *Twitter* feeds. Both studies then use these measurements to predict stock market returns. The sentiment measure of Bollen et al. (2011) reached an accuracy of 86.7% in predicting the daily direction changes in the closing value of the *DJIA (Dow Jones Industrial Average)*. And, Zhang et al. (2011) found that the percentages of *Tweets* showing hope, negativity, anxiety, and worry, are significantly positively correlated with the *VIX (Chicago Board Options Exchange's Volatility Index)*, and, at the same time, display significantly negative correlation with the market indices including the *DJIA*, the *NASDAQ*, and the *S&P 500*. A number of more recent studies, including Oh & Sheng (2011), Oliveira et al. (2013), and Renault (2017), switch to utilize messages from a microblogging service exclusively dedicated to financial discussions, called *StockTwits*. Two of the three studies mentioned above, Oh & Sheng (2011) and Renault (2017), verified that their sentiment indicators appear to have strong predictive power for intraday market directions, while the remaining study by Oliveira et al. (2013) found no evidence of return predictability using their indicator.

The second major approach used by researchers to **indirectly quantify** the unobservable sentiment is to find its proxy variables. Those indirect sentiment measures could be further divided into two groups of indicators. The first group, which we can call the **financial-based proxies**, is strictly related to the stock market's activities. According to Brown & Cliff (2004), technical analysts often considered several financial variables as the market's *weather vanes*[29]. One of the classic proxies for sentiment is the closed-end fund discount, which is defined as the difference between the net asset value of the actual security holdings and the market price of a closed-end fund[30]. Since the 1970s, plenty of studies, including Zweig (1973), Lee et al. (1991), and Neal & Wheatley (1998), have agreed that during the period when closed-end equity funds display high retail concentration, the average discount of those funds could be considered as an indicator of how pessimistic the retail investors currently are. However, in the opposite case where the closed-end equity funds are not held by a large number of retail investors, Qiu & Welch (2004) showed that this financial indicator might not be a proper measure for sentiment. Another sentiment proxy, that seems to stay more consistent over time, is the mutual fund flows. This indicator was used by Brown et al. (2005), Frazzini & Lamont (2008), Beaumont et al. (2008), and Ben-Rephael et al. (2012) to measure the sentiment of retail investors in certain markets.

---

[29] Weather vanes (or wind vanes) are the instruments used to show the direction of the wind.

[30] Investment funds which issue a fixed number of shares that are later tradable on stock exchanges.

Additionally, Kumar & Lee (2006) used the systematically correlated activities of retail traders as an indicator of market sentiment. The proxy later proved to be an excellent return predictor, especially for stocks that are costly to arbitrage or disproportionately held by retail investors. Scheinkman & Xiong (2003) proposed that both trading volume and market volatility could theoretically explain market bubbles. While trading volume, or in a broader sense, liquidity, was used as an investor sentiment index in Brown & Cliff (2004) and Baker & Wurgler (2007), the market volatility index *VIX* is often considered as the "*investor fear gauge*" by practitioners (Whaley, 2000) because the spikes of *VIX* are often accompanied by horrific market crashes.

Another branch of indirect sentiment proxies could be referred to as the **media-based indicators** because they originate from online platforms such as social media, message boards, search engines, web journals, etc. Not to confuse with pre-mentioned sentiment information that is directly extracted from detailed messages or articles, such indicators are merely the mass metrics of activities on those platforms, such as message volume or search volume. For instance, the daily message volume on *Twitter* (or *Tweet*[31] volume for short*)* is used by Rao & Srivastava (2012) and Ranco et al. (2015) as measures of sentiment. Both studies later prove that their sentiment indicators are strong predictors of future stock returns. Da et al., (2011, 2015) and Beer et al. (2013) proxy market sentiment using the *Google Search Volume,* reported by the website *Google Trend*[32], on stock tickers or a number of words with negative meanings in the context of finance such as bankruptcy, crisis, recession, poverty, liquidation, etc.

[31] A message posted on Twitter is called a *Tweet.*

[32] (https://trends.google.com/) A website by Google that analyzes the popularity of top search queries in Google Search across various regions and languages.

Furthermore, some studies consider the message volumes on *StockTwits*, the specialized-for-investor platform, as the measures of sentiment. Some of those studies include Oliveira et al. (2013) and Audrino et al. (2018).

Lastly, the third major approach used by prior researchers to measure sentiment is creating a **composite index**. Generally, the authors following this approach start by collecting a list of proxies for sentiment. They later implement signal extraction techniques, such as *Principal Component Analysis (PCA)* or *Kalman Filter*, to isolate all collected proxies into a common feature, which is then considered the final index of sentiment. The studies that pioneered this approach of measuring sentiment are possibly Brown & Cliff (2004) and Baker & Wurgler (2006), in which the former study employs a *Kalman Filter* method[33] and the latter applies the technique of *Principal Component Analysis*. Both studies agree that the fact sentiment individual proxies are often correlated with each other makes it very appealing to form an index as the common element of those proxies, in the hope that this index could measure sentiment in a somewhat clearer way. More importantly, Baker & Wurgler (2006) insists that it is nearly impossible for imperfect proxies to stay useful over time. In other words, while some sentiment proxies measure properly at a point in time, the others may only become valid at another time. Thus, for empirical experiments in the long horizons, it is sensible to combine a bunch of available proxies into a composite sentiment index that might have the potential to remain effective for a prolonged duration. Other studies following this approach include Chi et al. (2012) and Oliveira et al. (2017) (employing the *Kalman Filter* method); Chen et al. (2010), Finter et al. (2012), Khan

[33] Brown & Cliff (2004) actually also employed *PCA* to form two alternative indices, but for some reasons, they do not use those to predict market returns at the later analysis.

& Ahmad (2018), and Chong et al. (2017) (employing the *PCA* method).

Table 3.1: Methods of Measuring Sentiment in the Stock Market. Source: Own elaboration

Approach	**Family**	**Measures**
Direct	Survey	*Investors Intelligence* (II) Index; *American Association of Individual Investors* (AAII) Index; Consumer Confidence Surveys.
	Sentiment Analysis	Sentiment information extracted from social media platforms (*Twitter*, *StockTwits*); message boards (*Yahoo! Finance*, *Raging Bull*), web journals (*Wall Street Journal*, *Financial Times*).
Indirect	Financial Indicators	Closed-end Fund Discount; Mutual Fund Flows; Retail Traders Correlated Activities; Trading Volume; Liquidity; Market Volatility.
	Online Platforms' Metrics	*Twitter* or *StockTwits Message Volume*; *Google Search Volume.*
Composite		Common component extracted using *Kalman Filter*; Common component extracted using *Principal Component Analysis.*

Every method of measuring sentiment that is mentioned above could be summarized in Table 3.1. In the closing section of this chapter, we will finally examine some of the studies concerning sentiment and its possible effects in our target market - the market of cryptocurrencies.

3.3 In The Cryptocurrency Market

3.3.1 Theoretical Effects Of Sentiment On Cryptocurrency

Previously, there is a part where we have mentioned several studies empirically showing that sentiment tends to have a stronger effect on stocks whose valuations that are more subjective. If one recalls our discussion in section 2.3 on disruptive innovations, those results might ring a bell. In some sense, disruptive innovations are very similar to the group of stocks that lacks fundamental information and comparable counterparts for the purpose of valuation. At the moment when those innovations are newly born, they are often considered as unprecedented inventions, and there usually exists no reliable basis for the public to make assessments on their future. This could be one of the potential reasons why disruptive innovations tend to suffer from the early hypes, as discussed in the *Gartner Hype Cycle* model. Returning to our discussion on cryptocurrencies, if we consider those digital assets as disruptive according to the rationales presented in section 2.2, then theoretically there might be a possibility that they also possess the characteristic of subjective valuations. It turns out that many of the early studies investigating the pricing properties of this novel asset class agree with this opinion, including Cheah & Fry (2015), Hayes (2017), Berentsen & Schar (2018), Corbet et al. (2019), and Gurdgiev & O'Loughlin (2020). To be more specific, those studies argue that cryptocurrencies' prices are highly subjective because they generally offer no cash flow underlying their valuations, and there also exist no similar stocks or financial assets that could be used to perform comparable analysis. These statements are in line with our previous claim on disruptive innovations. Cheah & Fry (2015) even argue that if we only apply the traditional methods to value cryptocurrencies, their fundamental values should be equal to zero. Therefore, compared with other mainstream assets, it could be said that the market values of those digital assets are primarily speculative (Celeste et al., 2018; Corbet et al., 2019), and thus, should be considered prime targets to be affected by market sentiment. Furthermore, according to Bouri et al. (2018), many participants in the cryptocurrency markets are young and inexperienced individual investors, who often rely on social media and online chat forums for research. This might lead to even more herding behaviors and tendencies to buy assets with inflated prices among the investors, which is supported by the results of Menkhoff et al. (2006) and Greenwood & Nagel (2009) (mentioned in the previous section).

Overall, there seems to be a lot of studies suspecting the presence of sentiment inside the crypto market. However, an important question is that assuming sentiment truly exists, how should we expect this market emotional factor to affect the price formation of cryptocurrencies, at least from a theoretical standpoint. Should it be similar in the traditional stock market? Or should it be something more different? Recent studies on this matter seem to believe in the latter case. One of those studies, Chen et al. (2019), discusses in detail that sentiment might have a two-sided effect on the cryptocurrencies, meaning that this should be much more complex than the behaviors anticipated in the stock market. On the one hand, sentiment could also affect cryptocurrency prices in similar ways as with stocks, which is to generate temporary *momentum* in the short terms and create *reversal* in the medium-to-long terms. On the other hand, the authors argue that higher investor sentiment could also lead to an increase in the adoption rate of cryptocurrencies due to increasing awareness. According to a highly influential concept in technology, the so-called *network effect*, the more users connecting to a network system, the higher such system should be valued (Shapiro & Varian, 1999). In other words, besides the effects normally anticipated as in other financial markets, sentiment could also result in permanent *momentum* in the price of cryptocurrencies. All things considered, theoretically one should expect waves of sentiment in the crypto market to produce *momentum* in the short term (but the effect should be somewhat stronger, i.e. less harder to be documented compared with the situation with the stock market), and *uncertain* price movements in the medium-to-long (depending on which consequence might be stronger, the *network effect* or the *natural reversal*).

3.3.2 Empirical Studies On Cryptocurrency Investors' Sentiment

With the theoretical basis in mind, we now look into some of the empirical studies on sentiment and its effect in the cryptocurrency market. Again, based on the context of our discussion, there might be two aspects of those studies that one would like to pay attention to, namely the empirical results and the methods used to measure sentiment. Furthermore, we also prefer to compare such methods to those utilized by the studies conducted in the stock market. For this purpose, the following review will be organized in terms of the approach to measure sentiment adopted by the authors of those studies, meaning that whether they follow a **direct** or an **indirect** approach, or the combination of both, i.e. the **composite** approach.

First, we review some of the studies that attempt to measure sentiment in a **direct** way. As we discussed previously, there are two families of sentiment measures that belong to this approach: the survey-based indicators and the indicators derived from sentiment analysis.

Survey instruments seem to be not so popular with researchers in this field of study. Although there exist several survey-based indicators for the crypto market that are available on the Internet, such as the one introduced by Yuan & Wang (2018) or the one computed by a German company named *Sentix*[34], rarely we could find studies utilizing those on predicting cryptocurrency returns. Until now, there is perhaps only one study examining the relationship between Bitcoin's price and the investor sentiment index measured by *Sentix*, which is the one done by Rakovská (2018). This study performed a monthly time-series analysis in the timeframe from December 2013 to April 2018. The final results were quite promising since the author found that rising (falling) sentiment of investors towards Bitcoin increases (decreases) the price of this popular cryptocurrency in subsequent months, although the *momentum* becomes weaker over time.

As mentioned in section 3.2, one drawback of the survey-based indicators is that they are only available in low frequencies. This is admitted by the previous study's author Rakovská herself. In particular, Rakovská wished that she could employ the analysis at a higher frequency but unfortunately, the *Sentix* index is just published every month. This drawback is probably one of the reasons why those survey-based indicators are scarcely employed in the studies of sentiment in the crypto market. Moreover, the investors in cryptocurrencies also actively discuss their investment decisions on social media platforms and online message boards. Thus, applying *sentiment analysis* to

[34] https://www.crypto-sentiment.com/bitcoin-sentiment

derive measures appears to be a better choice than surveys, in terms of both data frequencies and feasibility. A variety of studies, for instance, Kaminski (2014), Perry-Carrera (2018), Karalevicius et al. (2018), Chen et al. (2019), Salac (2019), Valencia et al. (2019), Gurdgiev & O'Loughlin (2020), and Kraaijeveld & De Smedt (2020), has followed this methodology of measuring sentiment. Except for Salac (2019), all of the pre-mentioned studies found important results on the effect of sentiment on cryptocurrency returns. The earliest among those studies, Kaminski (2014), collected *Twitter* posts within the timeframe of 104 days and assessed their sentiment using a very small lexicon consisting of only 15 words. The study shows that *Twitter* sentiments actually do correlate with Bitcoin prices, but a later Granger-causality analysis fails to confirm any causal effect, thus the author concluded that *Twitter* sentiment only mirrors and not predicts the market. Karalevicius et al. (2018), Valencia et al. (2019), Chen et al. (2019), Gurdgiev & O'Loughlin (2020), and Kraaijeveld & De Smedt (2020) went further by using much more sophisticated lexicons than the one used by Kaminsky. More specifically, Perry-Carrera (2018), Valencia et al. (2019), and Kraaijeveld & De Smedt (2020) employed a lexicon usually used in the domains of social media, called *VADER (Valence Aware Dictionary and sEntiment Reasoner)* to assess sentiments of *Twitter* public messages. The measures of all three studies show significant predictive power[35] for daily market movements of cryptocurrencies such as Bitcoin, Ethereum, Ripple, etc. On the other hand, a financial lexicon, created by Loughran & McDonald (2011), is utilized by Karalevicius et al. (2018) and Gurdgiev & O'Loughlin (2020). Karalevicius et al. (2018) again applied this lexicon on *Twitter*

[35] Salac (2019) also utilized the *VADER* lexicon but fails to find any significant relationship between sentiment and cryptocurrency prices.

public messages. This study later predicts a tendency for investors to overreact on news in a short period of time. The other paper conducted by Gurdgiev & O'Loughlin (2020) did it differently by measuring sentiment using messages, not from a social media platform for general purposes like *Twitter*, but rather from a crypto-specific discussion website called *BitcoinTalk*. A disadvantage of *Twitter* is that this platform has an astonishing amount of messages daily, making it infeasible to collect messages for a long analysis timeframe (such as 1 year or over). Also, since *Twitter* is a platform for general users, a lot of irrelevant messages could be collected wrongly due to misclassification, resulting in a very inefficient data collection processes. Those weaknesses are admitted by Perry-Carrera (2018)[36], one of the previously mentioned studies that used *Twitter* messages as the main source for sentiment analysis. Compared with *Twitter*, the number of messages on platforms that are specialized for investors, such as *BitcoinTalk*, *StockTwits*, or crypto-relevant boards on the discussion website *Reddit*, seems to be more feasible to be collected in a prolonged analysis timeframe. Also, due to their specialization, the chance of collecting irrelevant messages is smaller as well. Thus, Chen et al. (2019) took this process of measuring sentiment to another level. Not only did they attempted to extract sentiment information from messages posted on two crypto-specific platforms (*StockTwits*, and crypto-relevant message boards on *Reddit*) over a daily span of nearly 5 years, but the authors also followed the automatic procedures of Oliveira et al. (2016) and Renault (2017) to create a highly sophisticated lexicon with over 10,000 terms specially related to cryptocurrencies. They further showed that their crypto-specific lexicon greatly

[36] In this study, it took the author 4 days to collect the Twitter messages that are relevant to cryptocurrencies in an interval of 1 day.

outperforms the one created by Loughran & McDonald (for financial terms) and the one created by Renault (2017) (for general social media) in classifying sentiment in messages concerning cryptocurrencies. Moreover, the results given by Chen et al. (2019) are quite interesting as they are conditional on the bubble status of the cryptocurrency market. Overall, the authors found significant evidences of increasing (decreasing) prices 1 day after a surge (decline) in sentiment during the bubble-growing and post-bubble periods. On the other hand, an increase in sentiment causes prices to fall 3 days afterwards, during the post-bubble periods. There are also some studies practicing *sentiment analysis* on articles from online newspapers (instead of social media messages), such as Cao & Rhue (2019) which apply the *AFINN* lexicon to assign sentiment scores for headlines of articles on the *Wall Street Journal* and the *Financial Times* containing the keyword *"bitcoin"* inside. The problem with this approach is that because cryptocurrencies are not mainstream assets, the frequency of them appearing on common financial news sources is very low. Sometimes, there are weeks or months between the publication of new articles on cryptocurrencies. Therefore, the studies following this method have to accept either a smaller sample size or a lower time frequency (since they have to obliterate all days without articles). For example, Cao & Rhue (2019) faced this trouble in their analysis. Initially, the original timeframe of their analysis is 454 days, but since there was no article on the *Wall Street Journal* and the *Financial Times* for respectively 397 and 324 days in the analysis period, the authors have to decrease the number of observations substantially. One could say that this drawback of online articles is, in some sense, very similar to the challenges faced by survey instruments. Therefore, at least until cryptocurrencies become somewhat mainstream assets, it is not recommended to exploit online articles as the textual sources used for computing crypto investors' sentiment.

Next, we look into some of the studies that follow an **indirect** approach to quantify sentiment, i.e., using proxies. We start by reviewing the family of financial-based indicators. Some of the indicated belonging to this family that is often used in the stock market, such as closed-end fund discounts, mutual fund flows, and retail trader activities, do not have their counterparts in the cryptocurrency market. Thus we would dismiss such indicators from this point of our discussion. Regarding trading volume, although it is quite straightforward to collect the data of this indicator, it is uncertain why researchers rarely use it as a proxy for sentiment in the cryptocurrency market. Concerning the market volatility index, it appears that only since 2018, the crypto market has possessed an index similar to the *VIX* index of the stock market. Due to its novelty, there seems to not exist any study that implemented this index as a sentiment measure of the cryptocurrency market. This crypto market volatility index, called the *VCRIX (Volatility CRypto IndeX)*, is created by Kolesnikova (2018) based on another cryptocurrency market index called the *CRIX (CRypto IndeX)*, proposed by Trimborn & Härdle (2018). Intuitively, one could think that the *VCRIX* and the *CRIX* perform similar functions in the cryptocurrency market as the *VIX* and the *S&P500* index do in the stock market, respectively. In any case, we will discuss those indices more in detail in the following chapters as they will play a major role in our analysis later.

Additionally, we also examine several studies that employed metrics of online platforms with a large presence of crypto investors as indirect proxies for sentiment. As one might guess, this family of indicators is also popular among the studies on cryptocurrency sentiment, based on the same rationales that we explained previously about the measures derived by using sentiment analysis on online textual data. Kristoufek (2013) and Abraham et al. (2018) found a high correlation between *Google Search Volume* and Bitcoin returns. They also considered this index as a strong measure for market sentiment. Abraham et al. (2018) even argued that *Tweet Volume* and *Google Search Volume* are better predictors of price directions than *Tweets* sentiment, which is invariably overall positive regardless of the price direction[37]. For this reason, they recommended using proxies for general interest such as *Google Search Volume* or *Tweet Volume* over the direct measures of sentiment.

Finally, when one attempts to look for studies that implement the third approach, i.e. creating a **composite** sentiment index, it turns out that at least until this moment, there seems to exist no attempt of measuring sentiment that follows this method. While it is not strange to see studies that utilized a lot of sentiment measures to predict cryptocurrency returns[38], such as Abraham et al. (2018), Chen et al. (2019), or Gurdgiev & O'Loughlin (2020), but none of these authors have used any signal extraction method for the purpose of deriving a common component out of their measures. As discussed previously in section 3.3, this approach could bring several advantages, such as dimension reduction that simplifying the later analyses, or the potentials to create a sentiment measure that could stay reliable over a long period of time. Despite having strengths with no obvious weaknesses, it remains unclear why no one has ever employed such an approach to measure sentiment in the cryptocurrency.

To conclude the section, generally speaking, to measure sentiment in the crypto market, it is recommended to use the indicators that originate from sources such as social media or discussion website due to the active engagement of crypto investors on these online

---

[37] The phenomenon that investors are often optimistic when they share opinions online has been documented empirically documented by Avery et al. (2011) and Kim & Kim (2014).

[38] This type of studies is even perhaps the most commonly seen among studies on cryptocurrency investors' sentiment.

platforms, plus the fact that those indicators are often available in high frequency. It is an important note that such indicators include both the direct measures derived using sentiment analysis and the indirect proxies that are simply online platforms' metrics. Concerning the sources to apply sentiment analysis, the platforms which are specialized for cryptocurrency discussions, such as *StockTwits*, *BitcoinTalk*, *Reddit* (crypto-relevant boards only), are suggested over the ones for general purposes like Twitter. Additionally, we could also utilize some financial-based proxies that are popular among the studies in the stock market but uncommon among the ones conducted in the market of cryptocurrencies, conditional on their existence in the latter. Two potential candidates are trading volume and market volatility. Lastly, it appears that for some unknown reasons, no one has ever employed a **composite** approach in measuring sentiment in the crypto market. This interesting detail could help us decide the later directions of this study. Because the **composite** approach has some benefits, along with uncertain drawbacks (except the certain need to collect a lot of data), it is perhaps not a bad idea to become the first follower of such an approach, especially if we could document significant correlations between individual indicators of sentiment. The following chapters will describe and execute this idea in more detail.

4 Data Collection

4.1 Premise

To create a **composite** sentiment index and carry out relevant analyses around it, the first step is to collect the data required by those processes. This task will be performed in the following chapter. Since our main purpose is to study the influences of sentiment in the cryptocurrency market, we begin by obtaining a benchmark indicator representing the performance of this asset class in section 4.2. The study chooses to analyze the market from an aggregated viewpoint, instead of focusing on a subset of several cryptocurrencies. Therefore, a capitalization-weighted market index called *CRIX (CRyptocurrency IndeX)* will be selected for the pre-mentioned purpose. Next, a list of sentiment measures will be collected according to the considerations presented in the conclusion of section 3.3. This list includes both direct and indirect measures of sentiment. On the one hand, the direct indicators, presented in section 4.3, will not contain any survey-based measure, but only the ones derived through sentiment analysis. The online platforms selected for our sentiment analysis are the ones designed exclusively for crypto-related conversations, including the social media *StockTwits* and the discussion website *Reddit* (only crypto-relevant message boards). General-purpose discussion platforms such as *Twitter* will be excluded from our study based on the rationales given in section 3.3.2 (see page 48). On the other hand, the indirect proxies, discussed in section 4.4, involves both financial-based indicators and online platforms' metrics. In the realm of financial-based indicators, we have two candidates, namely the trading volume and a novel market volatility index called the *VCRIX (Volatility CRypto IndeX)*. It is noteworthy that the first day that *VCRIX* has been measured is

28 November 2014, thus restricting our analysis timeframe to the period between such date and 25 July 2020[39]. Regarding the online platforms' metrics, in addition to the usage of message volumes on both of our target platforms, *StockTwits* and *Reddit,* this study also utilizes *Google Search Volume* as one of our sentiment proxies.

4.2 CRIX - A Cryptocurrency Market Index

The *CRIX (CRyptocurrency IndeX)* was constructed by Trimborn & Härdle (2018) in order to track the entire cryptocurrency market performance as close as possible. The study adopts the usage of this index for broad representations of the cryptocurrency market as in a number of pioneering papers on cryptocurrencies, such as Hafner (2018), Chen et al. (2019), and Alexander & Dakos (2020). The index formula is given as:

$$CRIX_t = \frac{\sum_{i=1}^{n} MV_{it}}{Divisor_t} \tag{4.1}$$

in which MV_{it} is the market capitalization of the cryptocurrency i at time t, and n is the number of constituents of the index. $Divisor_t$ is the coefficient ensuring that the changes in the number of crypto tokens in circulation do not affect the value of the *CRIX*. At the starting point of the *CRIX, Divisor* is defined as:

$$Divisor_0 = \frac{\sum_{i=1}^{n} MV_{i0}}{1000} \tag{4.2}$$

where 1000 is the chosen starting value of the *CRIX.* Whenever the number of tokens of a cryptocurrency changes, the *Divisor* will be adjusted accordingly, thus assuring

[39] 1 day before when the analysis was performed.

that the changes in *CRIX* are only caused by the price changes of the constituents.

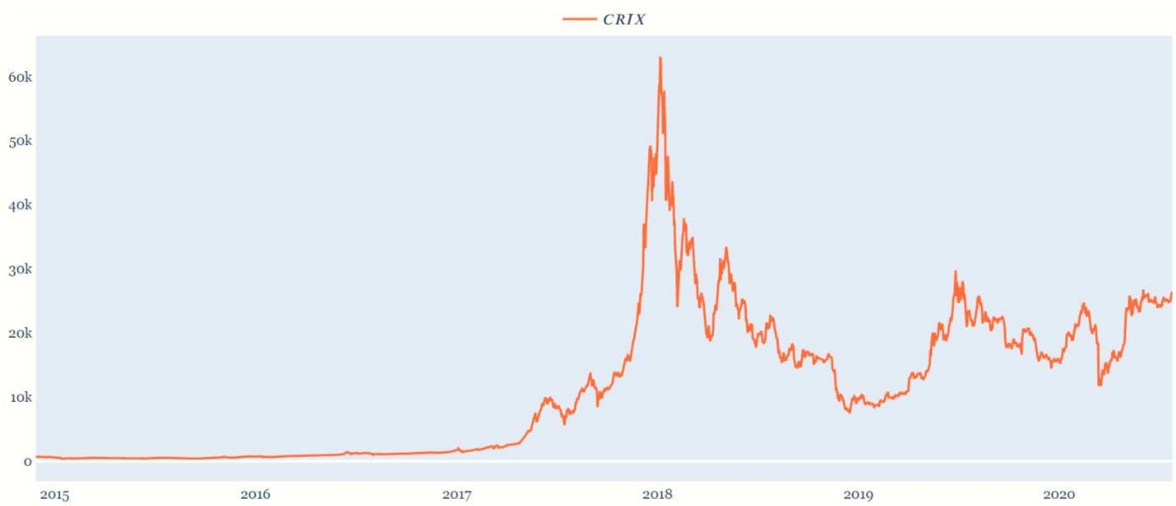


Figure 4.1: The *CRIX* series (28/11/2014 - 25/07/2020). Source: Own elaboration based on collected data.

In order to become eligible to be a constituent of the index, a cryptocurrency has to fulfill at least one of the two liquidity rules set out by its creators. The rules will not be mentioned explicitly in this study[40], but in principle, the components of *CRIX* have to be traded very frequently so that they can be easily converted to traditional fiat money or other cryptocurrencies. Also, the number of constituents n is not fixed as with the indices of relatively stable markets (such as the stock market with the *S&P500*) but actually allowed to variate to match a fast and dynamic market as cryptocurrencies. This number will be checked every 3 months to see if it still fits the market well, based

[40] For more details, see http://data.thecrix.de/#about

on certain rules set out by the index's authors. As of 30 August 2020[41], the number of *CRIX*'s components is currently set at 30. In addition, the index is also reallocated monthly or more frequently in case the need arises. Lastly, the full data series is provided by the authors at the link http://data.thecrix.de/data/crix.json. In Figure 4.1, the reader could find a plot of the index for our analysis timeframe (from 28 November 2014 to 25 July 2020). It could be seen crucial moments of the cryptocurrency market, especially the unprecedented boom (and subsequent crash) occurred in late 2017 and early 2018, have been captured quite effectively by the *CRIX* series.

4.3 Direct Sentiment Measures

4.3.1 StockTwits Sentiment

This section begins with a detailed re-introduction of *StockTwits*. This platform is a social media similar to *Twitter*, but dedicated specifically to financial discussions where investors can express opinions on any financial assets supported by the platform by posting messages with a maximum length of 140 characters. *StockTwits* currently supports over 2,500 financial assets, including 532 different cryptocurrencies[42]. There are three features that make *StockTwits* a great candidate for performing sentiment analysis. Firstly, the site has a public API that allows automatic retrieval of users' historical messages. Secondly, conversations on *StockTwits* are ordered using a feature

[41] See https://thecrix.de

[42] According to the site's API documentation, could be found at: https://api.stocktwits.com/symbol-sync/symbols.csv

called "*cashtags*". For example, we have $BTC.X referring to Bitcoin on this platform. This feature enables one to download messages that are exclusively related to his or her desired topics. Finally and most importantly, StockTwits has an optional feature for the users to tag their posts as "*Bullish*" or *"Bearish*". Conceptually, this feature makes applying supervised machine learning or similar algorithms very appealing to do because we have a very large dataset that is naturally labeled. Chen et al. (2019) exploited this to build a sophisticated lexicon that consists of nearly 10,000 terms frequently used by the investors of cryptocurrencies. According to the same study, this lexicon outperforms the financial lexicon by Loughran & McDonald by a substantial 32% of accuracy in classifying out-of-sample observations. Due to its superior performance, this lexicon will be employed to assess the sentiment of our retrieved messages at later stages.

The *StockTwits Public API*[43] is utilized for retrieving all messages posted between 28 November 2014 and 25 July 2020. Every retrieved message must contain at least one *cashtag* ending with ".X" (i.e, cashtag referring to a cryptocurrency on *StockTwits*). To avoid the domination of fake messages, possibly spammed by automatic bots, according to Pang et al. (2002), a maximum proportion of 1% of the dataset has to be imposed per user. The final dataset consists 2,045,322 messages (~287MB) from 48,648 distinct users and related to 519 cryptocurrencies. Among those, 999,720 messages are labeled by their posters as *"Bullish"* (818,452 messages ~ 40.02%) or *"Bearish"* (181,268 messages ~ 8.86%). The rest are unclassified messages. This imbalance is considered expected based on the previous claim on how investors are often optimistic when they

[43] https://api.stocktwits.com/developers

share opinions on social media (see footnote 37). Table 4.1 displays 10 cryptocurrencies with the highest number of messages posted on the platform during our analysis timeframe. It could be seen that messages about Bitcoin dominate nearly half of the total messages on *StockTwits*.

Table 4.1: 10 cryptocurrencies with the highest number of messages on *StockTwits* (28/11/2014 - 25/07/2020)

Cryptocurrency	**Number of Messages**
Bitcoin ($BTC.X)	964,167
Ethereum ($ETH.X)	169,481
Litecoin ($LTC.X)	147,808
TRON ($TRX.X)	140,884
Ripple ($XRP.X)	128,281
Dogecoin ($DOGE.X)	41,293
Bitcoin Cash ($BCH.X)	32,779
Ethereum Classic ($ETC.X)	24,342
Stellar ($XLM.X)	23,962
Bitcoin SV ($BSV.X)	23,472

Next, a *text processing* procedure is applied to the collected messages so that only relevant information remains. The main idea is to remove all words that are unnecessary for classifying the sentiment of sentences. The first step is to lowercase all messages. For some reason, messages collected using *StockTwits API* have all punctuations represented by their HTML codes instead of the Unicode (human-readable) characters. For example, "!" was expressed as "!". All of these HTML codes are converted back to Unicode format. All words with more than 3 repeated letters, such as *"hellooooooooooo*", which appear frequently as a form of emphasized sentiment on online discussion platforms[44], are shrunk to a maximum length of 3 (i.e. *"hellooo"*). Currencies hashtags (such as *"$BTC.X"*, *"$ETH.X"*), money values (such as *"$10k"*, *"€200"*), hyperlinks (such as *"http://stocktwits.com"*), numbers (such as 10,000 or 5k), username (such as *"@username"*) are replaced respectively by the word *"cashtag"*, *"moneytag"*, *"linktag"*, *"numbertag"*, and *"usertag"*. Next, all stop words such as *"I"*, *"he"*, *"she"*, *"it"*, *"herself"*, *"himself"*, *"the"*, *"an"*, *"a"*, etc.[45] are also removed. All punctuations are also removed except *"!"* and *"?"*. Lastly, the prefix *"negtag_"* is added to any word consecutive to negative words such as *"no"*, *"nor"*, *"isn't"*, etc.[46] For example, *"can't fly"* is converted into *"negtag_fly".* The *text processing* methodology adopted by this study is inspired largely by Renault (2017) and Chen et al. (2019). However, this study modifies their procedures by adding some

[44] See Brody & Diakopoulos (2011)

[45] The complete list of 136 stop words to be removed could be found in the file *stopwords.csv* in the GitHub repository of this project, which will be linked in the conclusion chapter.

[46] The complete list of 43 negative words could be found in the file *negative.csv* in the GitHub repository of this project.

extra steps, namely escaping HTML symbols, or providing a more detailed list of stop words and negative words. Examples of messages before and after processing are given in Table 4.2.

Table 4.2: Text processing examples

Before Processing	**After Processing**
$BTC.X wow this is way ahead of binance!	cashtag wow way ahead binance!
$BTC.X needs to hold above 10,200.. I really don't want to be bored by a 2 day consolidation between 10,000-10,200	cashtag needs hold numbertag really don't want bored numbertag day consolidation numbertag numbertag
$1ST.X Got fouled, this is a POS. need to hodl cos I'm fucked at the moment with this one. Will take 5 years to break even	cashtag got fouled pos need hodl cos fucked moment one take numbertag years break even
$SPY wallstreet is piling into bitcoin and crypto. Smart money leaving before the crash $BTC.X	cashtag wallstreet piling bitcoin crypto smart money leaving crash cashtag
$BTC.X to all the bitcoin bears out there.	cashtag bitcoin bears there

At this stage, we utilize the crypto-specific lexicon created by Chen et al. (2019)[47] to measure sentiment in each processed message. To make this happen, the messages must be transformed into vectors of words that could be unigrams (i.e. single word) or bigrams (i.e. two words side by side) in a procedure called *text vectorization*. For example, suppose there is a sentence such as *"I have a lovely dog"*, then the vectorizer will split that sentence into a vector of unigrams:

$$[\text{“}I\text{”}\ ,\text{“}have\text{”},\text{“}a\text{”}\ ,\text{“}lovely\text{”},\text{“}dog\text{”}]$$

and a vector of bigrams:

$$[\text{“}I\ have\text{”}\ ,\text{“}have\ a\text{”},\text{“}a\ lovely\text{”}\ ,\text{“}lovely\ dog\text{”}]$$

The reason why we have to convert the messages into vectors of unigrams and bigrams is that the lexicon of Chen et al. (2019) consists of both types of words. Moreover, bigrams even account for the vast majority of the words inside their lexicon with 7,329 out of 9,613 words are bigrams. At this moment, we could proceed to compute the sentiment score of each individual message in the exact order as follows. First, sentiment of the vector of bigrams is measured, then any bigram that has been matched with a word in the lexicon is removed. Next, sentiment of the vector of unigrams is measured. The sentiment scores of both unigrams and bigrams are summed up and divided by the total number of terms that have been matched with the lexicon. For instance, suppose we have a processed message such as *"moneytag pulls back pump"*. The bigram *"pulls back"* will be first taken into account with a sentiment weight of 0.88. The unigram *"pulls"* with a sentiment score of 0.54 will be skipped according to

[47] The crypto-specific lexicon is published by the main author (Professor Cathy Y. Chen) on her personal website: https://sites.google.com/site/professorcathychen/resume

the previous rules. However, the unigram *"pump"* will be considered with a weight of -0.30. Since there are 2 terms that have been matched with the lexicon, the sentiment score of the message is evaluated at $(0.88 - 0.30)/2 = 0.29$. After every individual message has been computed for its sentiment, the sentiment score for each day is derived by averaging the sentiment scores of all messages published on that day. The series of the daily sentiment of StockTwits users will be denoted as $SENT_{StockTwits}$ from this point. Figure 4.2 plots this data series versus the $CRIX$ index. For some reason, the sentiment score on *StockTwits* fluctuates strongly during the early days of our sampling period (pre-2017), then becomes more stable after that. Interestingly, the sentiment score seems to be almost always positive after June 2017. We will discuss this phenomenon in more detail when we examine the descriptive characteristics of the dataset in chapter 0.

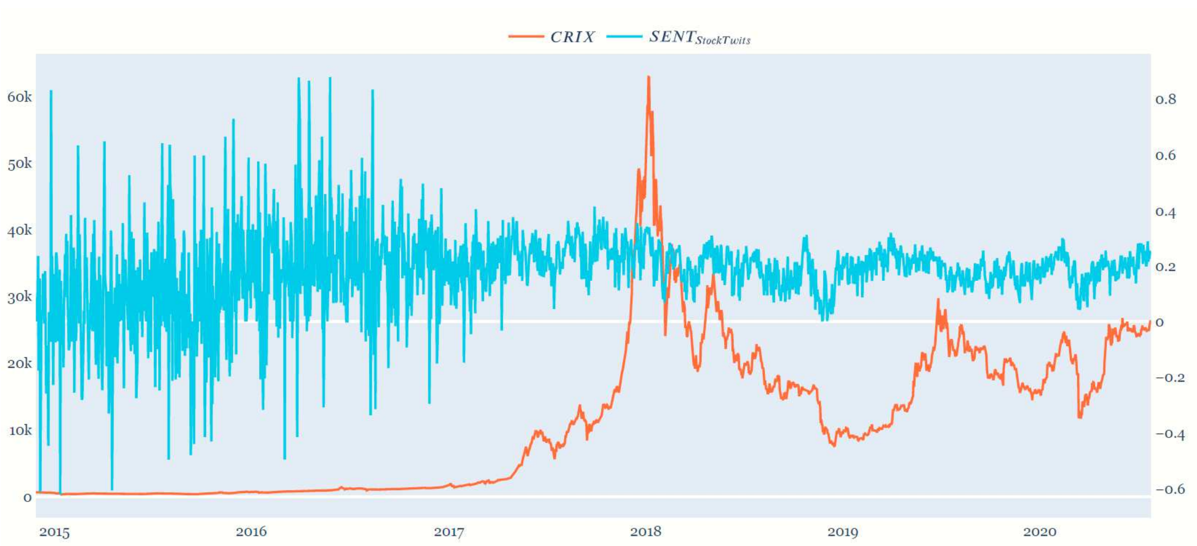


Figure 4.2: The $SENT_{StockTwits}$ & The $CRIX$ series (28/11/2014 - 25/07/2020).

Source: Own elaboration based on collected data.

4.3.2 Reddit Sentiment

Previously in section 3.3, we learned that *Reddit* is also an excellent source for topic-specialized textual data. Speaking more about *Reddit*, this is an American online discussion website that has been ranked as the 18th-most visited website globally (as of 30 August 2020) by the popular web traffic analysis company *Alexa Internet*[48]. Registered members (also known as *"Redditors"*) can submit content to the site such as links, text posts, images, videos, etc. (called *"submissions"*). These submissions are then voted up or down and discussed by other members. Posts are organized by subject into user-created boards (called *"subreddits"*), which cover a variety of topics, from politics, science, sports, fitness to video games, cooking, pets, investing, etc. Each subreddit is started with the prefix *"/r/"*.

This study employs the *Pushshift Reddit API* designed by the *"/r/datasets"* moderators to extract all submissions and comments posted in two subreddits strictly related to cryptocurrencies having more than 1 million subscribers, namely *"r/Bitcoin"* (~1.6m members), and *"r/CryptoCurrency"* (~1.1m members). Those submissions and comments are collected, once again, from 28 November 2014 to 25 July 2020. It is noteworthy that although *Reddit* also has its own official public API, the *Pushshift Reddit API* provides access to *Reddit* data in a much faster and more efficient way. Thus, it is chosen for the analysis in this study. After imposing a similar maximum proportion of 1% of the dataset on each poster, the final dataset contains 1,290,318 submissions (~327MB) from 291,009 distinct users, and an astonishing 13,015,447 comments (~2.61GB) from 498,460 distinct users. Since Reddit is also an informal

[48] https://www.alexa.com/siteinfo/reddit.com

source of information where people also use emojis, slangs, and short sentences, the same procedures of text processing, vectorization, and sentiment analysis are applied to extract sentiment information from Reddit submissions and comments as with StockTwits messages. Also, because comments are usually shorter and more informal than submissions, the analysis is done separately for both sources of data, resulting in two unique sentiment indices, denoted as $SENT_{\mathrm{RedSub}}$ and $SENT_{\mathrm{RedCom}}$. The reader could find below those series plotted versus the *CRIX* index in Figure 4.3 and Figure 4.4. Both of the series appears to vary in ranges relatively narrower compared to the one of the *StockTwits* messages' sentiment score. Over the sampling period, the $SENT_{\mathrm{RedSub}}$ series reaches its highest point during the market downturn in late 2018, while the $SENT_{\mathrm{RedCom}}$ peaks during the famous boom of late 2017.

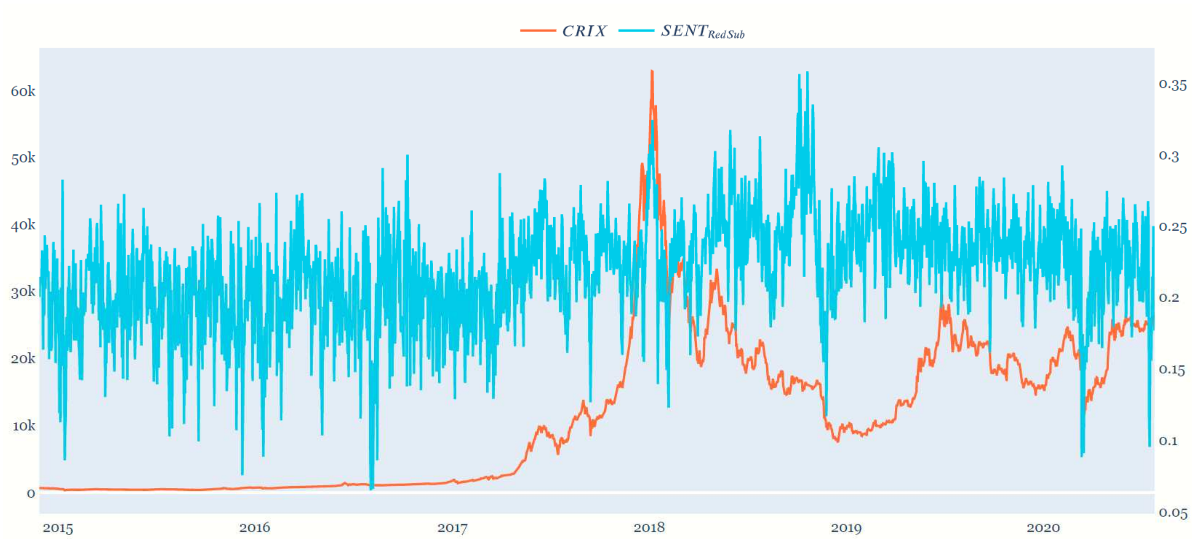


Figure 4.3: The $SENT_{RedSub}$ & the *CRIX* series (28/11/2014 - 25/07/2020). Source: Own elaboration based on collected data.

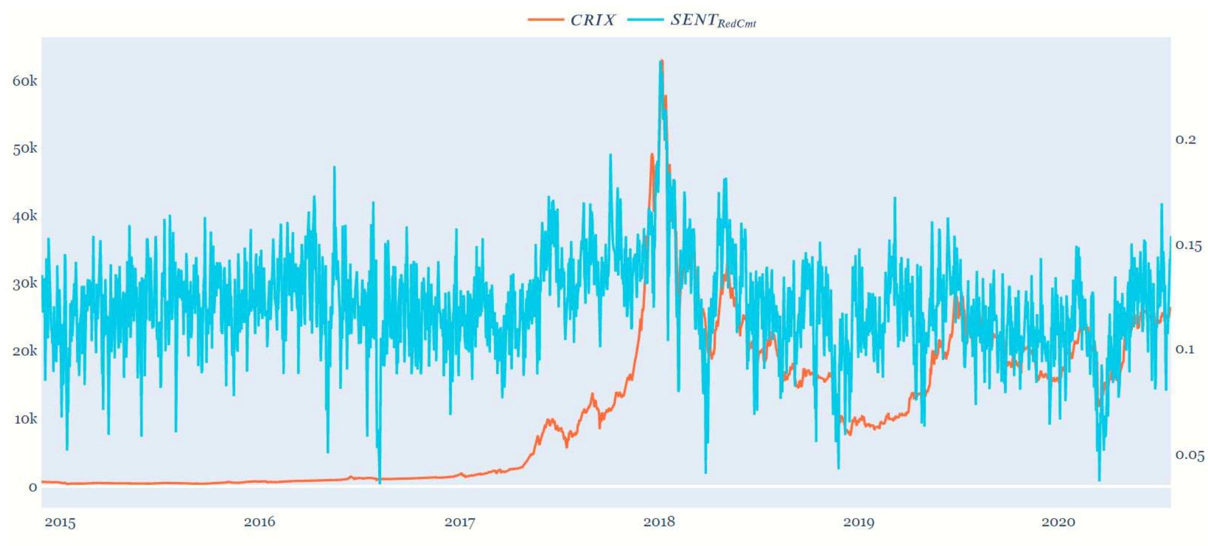


Figure 4.4: The $SENT_{RedCmt}$ & The *CRIX* series (28/11/2014 - 25/07/2020). Source: Own elaboration based on collected data.

4.4 Indirect Sentiment Measures

4.4.1 Financial Indicators

Following the rationales given in the closing part of section 3.3, we attempt to collect data of two potential sentiment proxies that are measures of market performance.

VCRIX - A Cryptocurrency Volatility Market Index

The first candidate is the *VCRIX (Volatility CRypto IndeX)*, created by Kolesnikova (2018) to grasp the risk induced by the cryptocurrency market. This index plays similar roles in the market of crypto as the *VIX* in the US stock market, or the German implied volatility index *VDAX* in the German stock market. Normally, computing the implied volatility of a financial market requires the presence of derivative instruments inside that market. However, despite the lack of a developed cryptocurrency derivatives market, Kolesnikova has successfully simulated the *VCRIX* on the basis of the *CRIX* market index. The data series could be collected at the author's website[49]. As usual, in Figure 4.5, one could find a plot of the *VCRIX* versus the *CRIX* during our analysis timeframe. It could be seen over our sampling period, the *VCRIX* index reaches its peak of 2324 points very recently, on 14 March 2020, and hits its bottom being 113 points on 3 September 2016.

Figure 4.5: The *VCRIX* series & The *CRIX* series (28/11/2014 - 25/07/2020).

Source: Own elaboration based on collected data.

Market Trading Volume

Another possible candidate in this family of financial indicators is the market's aggregated trading volume. The data series has been downloaded using *Nomics Public*

[49] http://data.thecrix.de/data/crix11.json

API[50]. *Nomics* is a price-tracking website for crypto assets that is comparable to websites such as *CoinMarketCap* or *CoinGecko*. In terms of popularity among crypto investors, *Nomics* is behind *CoinMarketCap* and *CoinGecko*, however, this site is employed since only its API provides the information required by our analysis. This data series, denoted as VOL_{Trade}, is quoted in USD and available in a daily frequency (Figure 4.6). Initially, the trading volume series appears to co-move very well with the market index series, but stops to do so until the early of this year (2020) when it bursts significantly with a peaked trading volume that equals to approximately 4 times of the highest point reached during the bubble of late 2017.

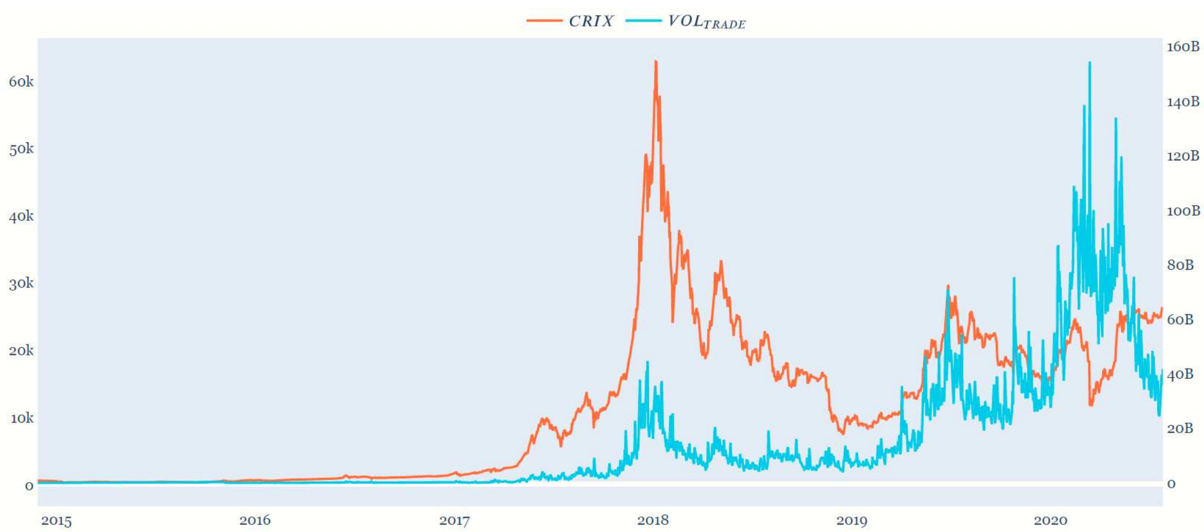


Figure 4.6: The VOL_{Trade} series & the CRIX series (28/11/2014 - 25/07/2020).

Source: Own elaboration based on collected data.

[50] https://docs.nomics.com

4.4.2 Online Platforms' Metrics

Google Search Volume

Based on the important recommendations of Abraham et al. (2018) that have been discussed in section 3.3, *Google Search Volume*, a search index provided by *Google Trends*, is employed as an indirect sentiment measure. At this stage, the *Python* package *Pytrends*[51] is implemented to collect the *Search Volume* of the keyword *"bitcoin"* from 28 November 2014 to 25 July 2020. The reason why this keyword is selected for our analysis is simply that according to *Google Trends*, it is the crypto-related keyword that is most searched for on *Google*. The search volume is scaled automatically across the time period by the *Pytrends* package. Once again, this sentiment measure, denoted in our analysis as VOL_{Google}, is plotted agaisnt the *CRIX* in Figure 4.7. It is not surprising that the *Google Search Volume* peaks during December 2017 and January 2018. Moreover, there seem to be more observers paying attention to the crypto market after the market boom of 2018 than before the incident.

[51] The Pseudo (Unofficial) API for *Google Trends* designed for *Python* users: https://pypi.org/project/pytrends

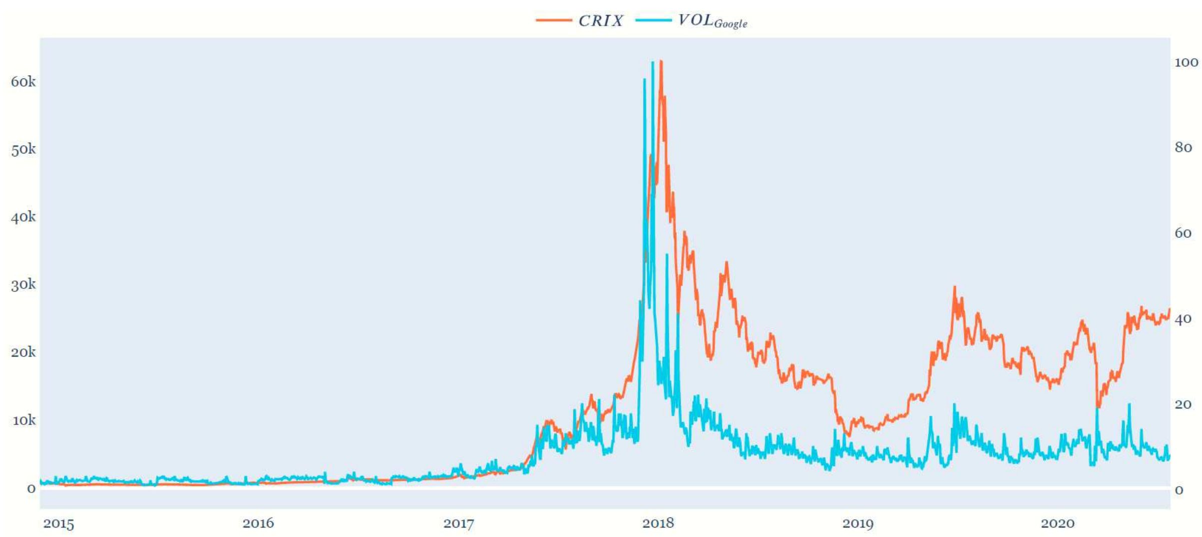


Figure 4.7: The VOL_{Google} & The *CRIX* series (28/11/2014 - 25/07/2020). Source: Own elaboration based on collected data.

Metrics of StockTwits and Reddit

Also, since a large number of messages and discussions on *StockTwits* and *Reddit* have been collected in previous sections, it is quite straightforward to also compute the metrics of these online platforms and utilize those as proxies for sentiment. The data series of *StockTwits* message volume, the number of *Reddit* submissions, and the number of *Reddit* comments are denoted respectively as $VOL_{StockTwits}$, VOL_{RedSub}, and VOL_{RedCmt}. The reader might also find their plots versus the *CRIX* in the figures below. The submission and comment volumes on *Reddit* appear to co-move extremely closely with each other and the market index. There exist some point where their increases (decreases) even precede the movement of the market index. In contrast, the message volume on *StockTwits* seems to only spike during crucial events of the market. A possible explanation might be that discussions on *Reddit* take place regularly every day since there are not only investors but also tech enthusiasts getting involve on this platform. In contrast, a vast majority of *StockTwits* users are probably investors and they will mostly talk about the market when there are some large movements ongoing.

To conclude this chapter, we have collected the daily time series of 9 different individual measures for sentiment, namely $SENT_{StockTwits}$, $SENT_{RedSub}$, $SENT_{RedCmt}$, *VCRIX*, VOL_{Trade}, VOL_{Google}, $VOL_{StockTwits}$, VOL_{RedSub}, and VOL_{RedCmt}. The sample period is from 28 November 2014 to 25 July 2020. Besides from that, the index chosen to become the aggregated measure of the crypto market is the *CRIX* time series, also obtained for the pre-mentioned period. At this point, the data collection process has been completed. In the next chapter, we will move forward to the actual analysis where our true measure of sentiment in the crypto market is finally computed, according to the plan proposed at the end of chapter 3. This sentiment measure will also be utilized to test the suspicion set out in section 2.3 that public expectations on cryptocurrencies (evaluated by the market prices of those digital assets) are currently suffering from irrational hypes.

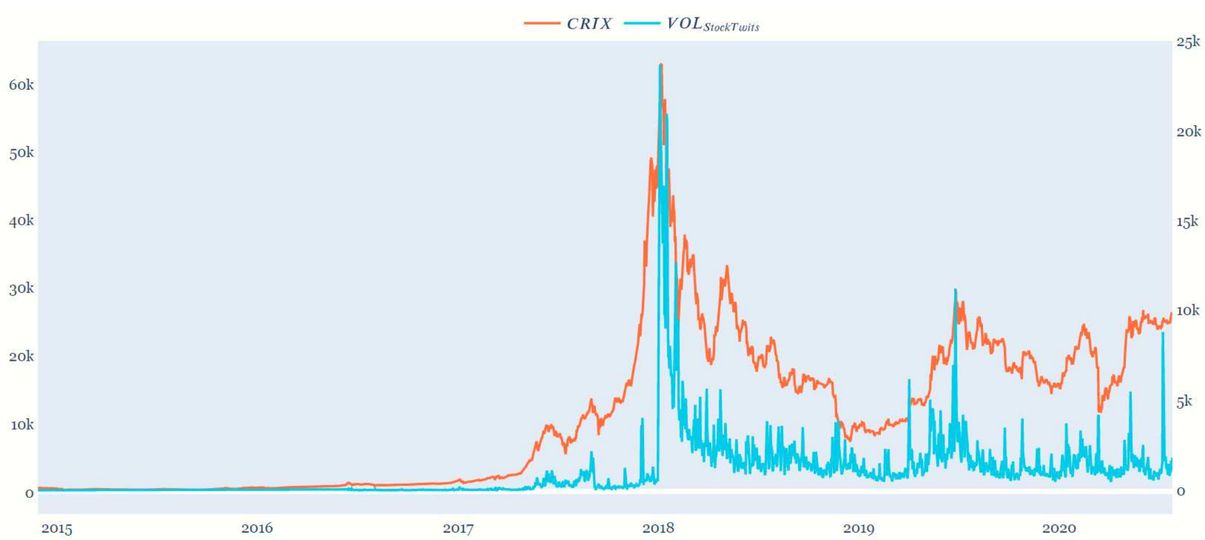


Figure 4.8: The $VOL_{StockTwits}$ series & The *CRIX* series (28/11/2014 - 25/07/2020). Source: Own elaboration based on collected data.

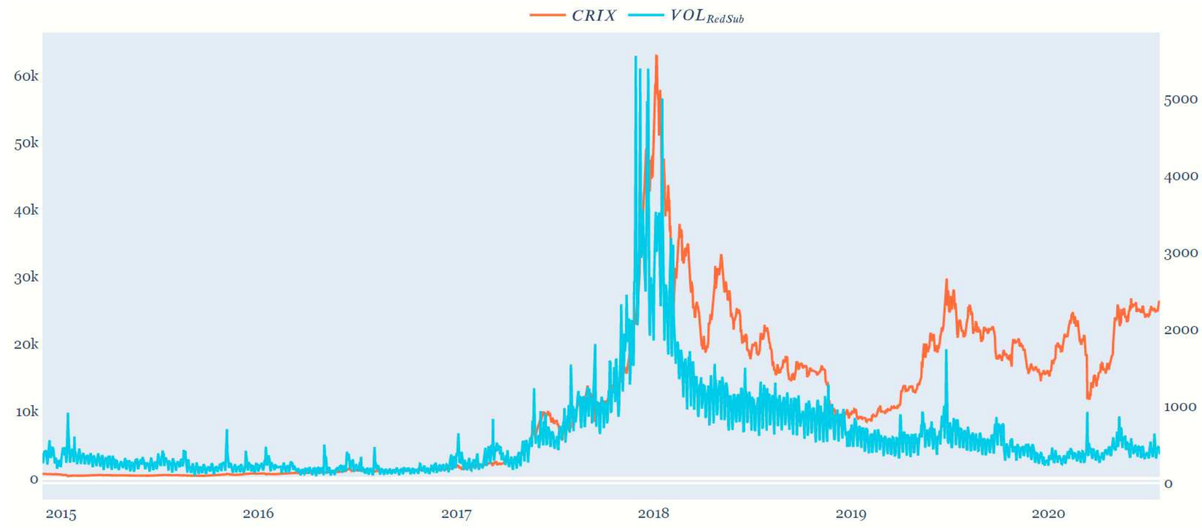


Figure 4.9: The VOL_{RedSub} series & The $CRIX$ series (28/11/2014 - 25/07/2020).

Source: Own elaboration based on collected data.

Figure 4.10: The VOL_{RedCmt} series & the $CRIX$ series (28/11/2014 - 25/07/2020).

Source: Own elaboration based on collected data.

5 Sentiment and Cryptocurrency Return

5.1 Premise

In this chapter, we turn into the step of creating our composite sentiment index using the data available at hand. For this purpose, the signal extraction technique *Principal Component Analysis* will be employed. Section 5.2 discusses this exercise in detail. After the sentiment index has been created, in section 5.3, we perform a *Vector Autoregressive Analysis* and other analytical methods to see how this new index interacts with the crypto market index *CRIX*. The result indicates that the sentiment index is a strong predictor of crypto market returns in the short term. In the closing section of this chapter, a simple trading strategy based on the sentiment index is simulated to see if the index is capable of generating any *alpha*[52] for the investors who utilize it as a return predictor.

5.2 A Composite Sentiment Index

5.2.1 Descriptive Statistics

[52] Technically speaking, *alpha* refers to excess returns of a trading strategy or an investment, in comparison with a market index. The term *generating alpha* is also be used in a somewhat fancy way by some quantitative hedge funds, Wall Street investors, or algorithmic trading firms to describe how they could predict the market superiorly than the average investor.

We begin the analysis by taking a look at some of the important characteristics of the collected dataset. In Table 5.1, one could find the summary statistics reported for the ten variables acquired in the last chapter. In fact, there are not too many points that could be put into comparison between those variables since they are mostly measured on different scales, except the ones that are derived using sentiment analysis, namely $SENT_{StockTwits}$, $SENT_{RedSub}$, and $SENT_{RedCmt}$. While these measures are generally positive (which is in line with the statement above about the optimism of online investors), the sentiment score of *StockTwits* seems to display greater volatility than the ones coming from *Reddit*. This phenomenon is exactly similar to in the study of Chen et al. (2019) where the crypto-specific lexicon is also applied on *StockTwits* and *Reddit* textual data.

Table 5.1: Summary Statistics (Before Standardization). Source: Own elaboration.

Variable	**Mean**	**Std. Dev**
CRIX	11391.7	11269.3
$\boldsymbol{SENT_{StockTwits}}$	0.17860	0.14028
$\boldsymbol{SENT_{RedSub}}$	0.21820	0.03836
$\boldsymbol{SENT_{RedCmt}}$	0.12006	0.02369
VCRIX	820.852	386.930
$\boldsymbol{VOL_{Trade}}$	1.37E+10	2.19E+10
$\boldsymbol{VOL_{Google}}$	7.75068	7.87399
$\boldsymbol{VOL_{StockTwits}}$	1002.12	1847.41
$\boldsymbol{VOL_{RedSub}}$	609.919	602.834
$\boldsymbol{VOL_{RedCmt}}$	6095.96	6565.66
Number of Observations	2041	

Next, all sentiment measures (except *CRIX*) are standardized for zero-mean and unit-variance. This is because later we want to apply *PCA* on those variables. In *PCA*, we are interested in the components that maximize the joint-variation between variables. If the variables are left unscaled, some (e.g. $SENT_{RedCmt}$) will vary less than others (e.g. VOL_{Trade}) because of their respective scales and *PCA* might determine that the direction of maximum variance more closely corresponds with the variable with high relative variation. The summary statistics after standardization are presented in Table 5.2 below.

Table 5.2: Summary Statistics (After Standardization). Source: Own elaboration

Variable	Kurtosis	Skewness	Minimum	Maximum
$\boldsymbol{SENT_{StockTwits}}$	5.28	-0.68	-5.70	5.00
$\boldsymbol{SENT_{RedSub}}$	0.77	-0.33	-4.00	3.65
$\boldsymbol{SENT_{RedCmt}}$	2.06	0.29	-3.55	4.93
VCRIX	1.98	1.17	-1.83	3.89
$\boldsymbol{VOL_{Trade}}$	5.45	2.25	-0.63	6.43
$\boldsymbol{VOL_{Google}}$	32.02	4.30	-0.88	11.72
$\boldsymbol{VOL_{StockTwits}}$	47.27	5.67	-0.54	12.29
$\boldsymbol{VOL_{RedSub}}$	14.85	3.21	-0.87	8.20
$\boldsymbol{VOL_{RedCmt}}$	17.47	3.65	-0.79	9.23
Number of Observations	2041			

One might also examine the correlations between standardized sentiment measures in Table 5.3. The sentiment indicators appear to be highly correlated with one another with 34 out of 36 pairs showing statistically significant relationships (with p-values less than 0.01). Two pairs that do not show significant correlations (colored in red) both involve the volatility index *VCRIX*. The results further reveal that most relationships are positive (which is in line with conventional expectations), except the ones that entail *VCRIX* and the market trading volume. This result is still understandable, to a certain degree, as both of these variables are not very straightforward measures of sentiment. Usually, market volatility indices could only become meaningful during the time when investors feel extremely bearish about the market. In addition, while it is true that investors tend to trade more as they feel more confident about future prospects, the trading volumes could also rise in the episodes of market crashes where people panic sell their assets. In other words, while financial-based indicators could sometimes turn into valid proxies of sentiment during special moments of the market, their performance (on measuring sentiment) will not stay consistent over time. Therefore, they are recommended to be used together with other indicators for the purpose of measuring sentiment. Back to the discussion on the relationship between sentiment individual indicators, another noteworthy detail is that the highest correlations (around 0.8 ~ 0.9) are present among the metrics of online platforms, namely VOL_{Google}, $VOL_{StockTwits}$, VOL_{RedSub}, and VOL_{RedCmt}, meaning that they seem to be measuring the same phenomenon (possibly sentiment). Put this differently, one might guess that those variables could be among the most valid individual measures for the sentiment of crypto investors. It should also be noted that this speculation is in line with the results of Abraham et al. (2018) (mentioned previously in section 3.3).

According to the arguments of Brown & Cliff (2004) and Baker & Wurgler (2007) given in section 3.2, significant correlations between sentiment indicators make it conceptually appealing to construct a composite index that can represent sentiment in a clearer way and hope that it could stay valid for a prolonged time. In order to isolate as much information as possible into the final index, the *PCA* method used by Baker & Wurgler (2007) will be adopted to combine the sentiment indicators. The next section will describe this process in more detail.

Table 5.3: Contemporanous Correlations. Source: Own elaboration

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
(1)	1.000								
(2)	0.217	1.000							
(3)	0.220	0.374	1.000						
(4)	-0.044	-0.046	0.069	1.000					
(5)	0.058	0.131	-0.130	0.160	1.000				
(6)	0.158	0.184	0.243	0.276	0.327	1.000			
(7)	0.089	0.217	0.244	0.259	0.339	0.497	1.000		
(8)	0.137	0.199	0.340	0.290	0.107	0.878	0.613	1.000	
(9)	0.106	0.124	0.348	0.342	0.118	0.839	0.693	0.955	1.000

(1): $SENT_{StockTwits}$ (4): $VCRIX$ (7): $VOL_{StockTwits}$
(2): $SENT_{RedSub}$ (5): VOL_{Trade} (8): VOL_{RedSub}
(3): $SENT_{RedCmt}$ (6): VOL_{Google} (9): VOL_{RedCmt}

5.2.2 Principal Component Analysis

Before applying *PCA* on the standardized sentiment indicators, Bartlett's test of sphericity is employed to check if their correlation matrix diverges significantly from the identity matrix. Since the p-value is less than 1% (Table 5.4), the null hypothesis of orthogonal variables can be rejected, implying that *PCA* can be used to derive the compression of all available sentiment indicators.

Table 5.4: Bartlett's Test of Sphericity. Source: Own elaboration

Bartlett's Test of Sphericity	Approx. Chi-Square	12125.57
	df	36
	Sig.	.00

Similar to Baker & Wurgler (2007) and Khan & Ahmad (2018), the composite sentiment index is the first principal component of the nine individual indicators, which is simply a linear combination of those variables with the coefficients chosen to capture the most joint-variation across all nine series. All indicators' coefficients, also called *component loadings*, are presented in the following equation, portraying that the indicators have the signs as expected.

$$\begin{aligned} SENT_9 = \ & 0.116 SENT_{StockTwits} + 0.166 SENT_{RedSub} + 0.226 SENT_{RedCmt} \\ & + 0.207 VCRIX + \ 0.162 VOL_{Trade} + 0.460 VOL_{Google} \\ & + 0.394 VOL_{StockTwits} + 0.484 VOL_{RedSub} + 0.487 VOL_{RedCmt} \end{aligned} \tag{5.1}$$

It could be seen that five indicators have *component loadings* less than 0.3. Based on a rule of thumb set by Hair et al. (2009), we can attempt to eliminate every variable with a loading below 0.3 and construct another composite index from the four remaining variables. The new index (SENT_4) has the *loadings* as follows:

$$SENT_4 = 0.499VOL_{Google} + 0.419VOL_{StockTwits} + 0.534VOL_{RedSub} + 0.539VOL_{RedCmt} \tag{5.2}$$

The equation of $SENT_4$ shows that all indicators have the same signs and almost the same (but larger) coefficients as in the original index ($SENT_9$). The correlation between $SENT_9$ and $SENT_4$ is extremely high (0.976), indicating that eliminating four sentiment indicators did not cause any significant change. This can also be seen in Figure 5.1 where we plot the two series against each other. The only crucial difference between the series perhaps occurs in the early years (from 2015 to 2017) where the original index $SENT_9$ seems to vary much stronger than the compact version $SENT_4$. This is easy to understand since the five indicators removed to form $SENT_4$ fluctuate very widely during those years, as one could see in the previous Figure 4.2, Figure 4.3, and Figure 4.4. To avoid losing early variation, the study proceeds with the sentiment index computed using all nine indicators for the final analysis. From this point, the index will be denoted solely as $SENT$ instead of $SENT_9$.

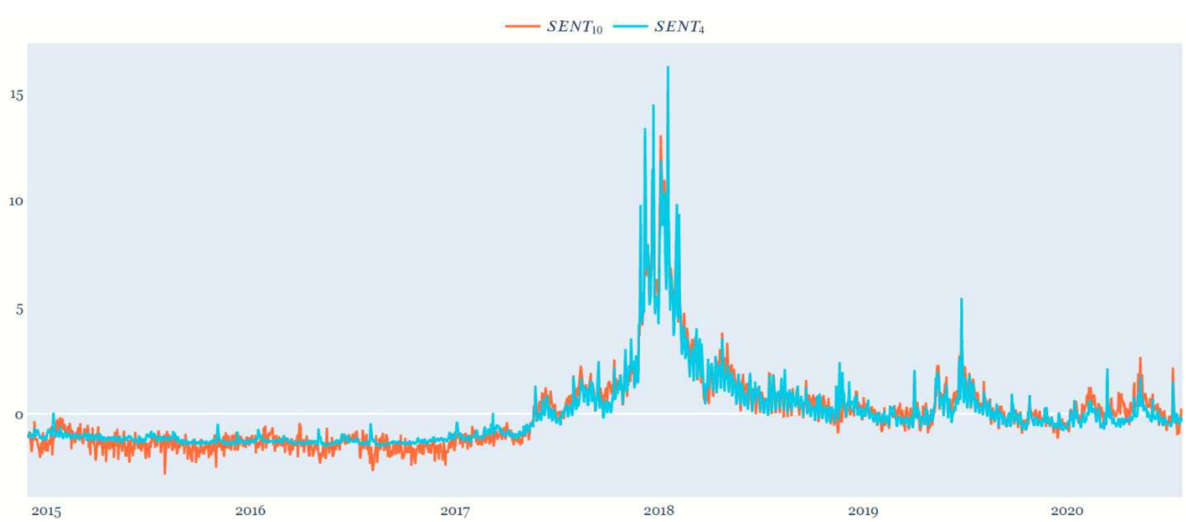


Figure 5.1: $SENT_9$ & $SENT_4$. Source: Own elaboration.

The correlation vector of nine individual sentiment indicators and the composite sentiment index, given in Table 5.5, shows that all sentiment proxies are directly related to the index. To summarize, it appears that we have successfully isolated the unobserved factor of sentiment from various individual indicators.

Table 5.5[53]: Correlation Vector of Individual Sentiment Measures with the Composite Index. Source: Own elaboration.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
SENT	0.222	0.319	0.443	0.401	0.302	0.883	0.761	0.934	0.940
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

(1): $SENT_{StockTwits}$ (4): *VCRIX* (7): $VOL_{StockTwits}$
(2): $SENT_{RedSub}$ (5): VOL_{Trade} (8): VOL_{RedSub}
(3): $SENT_{RedCmt}$ (6): VOL_{Google} (9): VOL_{RedCmt}

5.3 Vector Autoregressive Analysis

With the composite index in hand, we now turn to test the key hypotheses about how sentiment affects the market of cryptocurrency. Some of the earlier discussion in section 3.3 suggests that an increase (decrease) in sentiment may lead to a temporary price change in the same direction in the short term. In an initial plot (Figure 5.2), the new sentiment index *SENT* appears to be highly promising as it seems to precede most of the trends happening in the market index *CRIX*. A *Vector Autoregressive (VAR)*

[53] Values in parenthesis represent significance levels of the corresponding correlation coefficients.

model will be employed to see the interaction between the two series in a formal way. The analysis will also attempt to identify any possible causality between the sentiment index and the crypto market index.

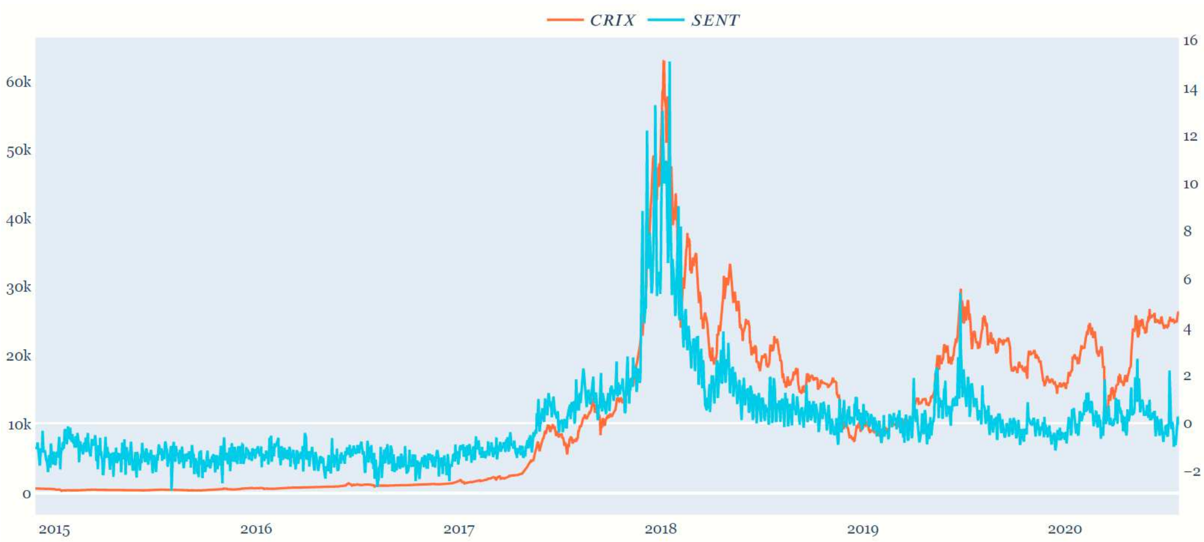


Figure 5.2: The *SENT* index & the *CRIX series*. Source: Own elaboration.

5.3.1 Stationarity Test

To cover various combinations of relationships, the market return series (denoted as RM) and the first difference of the sentiment index (denoted as $\Delta SENT$) will also be studied. The formulas of these series (at time t) are given as follows:

$$RM_t = \frac{CRIX_t \; - \; CRIX_{t-1}}{CRIX_{t-1}} \tag{5.3}$$

&

$$\Delta SENT_t = SENT_t - SENT_{t-1} \tag{5.4}$$

Before applying *VAR*, it is necessary to test the series' stationarity using the KPSS test (Kwiatkowski et al., 1992) and the ADF test (Dickey & Fuller, 1979). It is noted that the two tests have opposite null and alternative hypotheses where the KPSS test has the null of stationarity. The test results are summarized in Table 5.6. The $\Delta SENT$ and the *RM* series are found to be stationary and not contain any unit root, as could also be seen from their graphs in Figure 5.3.

Table 5.6: Stationarity Tests. Source: Own elaboration

	KPSS	**p-value**	**ADF**	**p-value**
SENT	4.3250	<0.01	-2.3810	0.1472
CRIX	1.9274	<0.01	-2.1351	0.2306
$\boldsymbol{\Delta SENT}$	0.1561	>0.1	-11.0189	0.0000
RM	0.0342	>0.1	-16.3843	0.0000

This also implies that the *SENT* and the *CRIX* series are non-stationary time series with an order of integration of 1[54]. The later *VAR* analysis will be applied to the two stationary series, namely $\Delta SENT$ and *RM*. A plot of these two series could be found in Figure 5.3. A side note is that it is also possible to apply *VAR* on the pre-transformed variables (*SENT* and *CRIX*) if we could verify whether they are co-integrated. If this is the case, our *VAR* model will become a *Vector Error Correction Model (VECM)*, which could be thought of as a restricted version of *VAR*. However, since the

[54] Non-stationary series that takes 1 time of differencing to become stationary.

relationship between the two stationary variables ($\Delta SENT$ and RM) is already straightforward to be interpreted in from a financial perspective (and also commonly investigated in studies concerning market sentiment), the author decides to proceed with these two variables and does not check for the co-integration between the $SENT$ and $CRIX$ series.

Figure 5.3: The $\Delta SENT$ index & the $CRIX$ Return (RM). Source: Own elaboration.

5.3.2 VAR Results

The general *VAR(p)* model specification is given as:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \tag{5.5}$$

where $Y_t = [RM_t, \Delta SENT_t]^T$, α is the vector of intercepts, β_i is the time-invariant (2x2)-matrix of coefficients, and ε_t is the vector of error terms. Goodness-of-fit criteria BIC suggests $p = 5$.

Table 5.7 reports the result from estimating the VAR model using daily market return (RM) and daily changes in the sentiment index ($\Delta SENT$). The blocks of rows indicate the contribution of each independent variable at lags 1 - 5. As we want to find out the influences of sentiment on market return, the first block of rows and the first column should be our primary interest.

The first block of rows shows that $\Delta SENT$ is a powerful predictor of itself. All lagged values of the series from 1 to 5 day(s) are negative and significant at 1%. This outcome is very similar to what is observed by Brown & Cliff (2004) in the stock market where their composite index also displays strongly negative autocorrelations.

However, at the part showing from the relationship between the market return variable RM and lagged values of $\Delta SENT$, we acquire different results compared to Brown & Cliff (2004). Our market return variable RM is significantly positively correlated with $\Delta SENT$ at lag 1, 3, 4 (at 1%), and lag 5 (at 5%), while their sentiment index shows no correlation at all with their market variables. One possible explanation for this phenomenon has been discussed in section 3.3, where we anticipate that sentiment influences should be stronger (and consequently, easier to be captured) in a young and innovation market like crypto than a well-established market like stocks. Moreover, the positive correlation of RM and $\Delta SENT$ at all lags is also consistent with the expectation that rising (falling) waves of sentiment lead to temporary increase (decrease) in cryptocurrency prices in the short horizon.

On a side note, changes in the sentiment index are also significantly correlated with the lagged market return (at 1% level with lag 1, 4; at 5% level with lag 2). Thus, one might also suspect that returns could drive the changes in sentiment in the opposite direction.

Table 5.7: VAR(5) Results (RM & $\Delta SENT$). Source: Own elaboration

Independent Variable	Lag	Dependent Variable: RM	Dependent Variable: $\Delta SENT$
$\boldsymbol{\Delta SENT}$	1	0.0088***	-0.2349***
	2	0.0004	-0.3481***
	3	0.0045***	-0.2748***
	4	0.0044***	-0.2132***
	5	0.0038**	-0.2154***
$\boldsymbol{RM}$	1	-0.0427*	0.9456***
	2	0.0198	0.6894**
	3	0.0152	0.2291
	4	0.0317	1.3735***
	5	0.0034	0.2589

* Indicate significance at the 10% level
** Indicate significance at the 5% level
*** Indicate significance at the 1% level

There is a popular mantra in the world of statisticians that goes *"Correlation is not causation"*, referring to the inability to formally deduce a cause-and-effect relationship between two variables solely on the basis of an observed correlation between them. Thus, it is not quite straightforward to interpret the significant estimates of the previous VAR model into a statement of predictability between the sentiment index and the market return. To determine whether the composite sentiment index is useful in forecasting the market returns series, one has to rely on a different statistical test called *Granger-causality*, which we will turn our attention to in the next section.

5.3.3 Granger-causality Analysis

The *Granger-causality* test, first proposed by the Nobel laureate Clive W. J. Granger (1969), is a statistical test for assessing whether adding a time series leads to significantly better *forecasts*[55] of another one. A time series X is said to *Granger-cause* (i.e., have a predictive power of) another series Y if it can be shown that the past values of X provide statistically significant information about the values of Y. By applying the *Granger-causality* test on the changes in sentiment index and the market return, we retrieve the results as presented in Table 5.8. Since the null hypotheses of no *Granger-causality* are rejected in both test directions with very small p-values, it could be seen that the two variables, namely $\Delta SENT$ and RM, are both strong predictors of the remaining one. Interestingly, our results coincide with the results of similar papers (but done in different markets) such as the ones by Alrabadi & Waleed (2015) (who studied the stock market of Jordan), Khan & Ahmad (2018) (who did a nearly identical analysis in the stock market of Pakistan). It could be seen that the markets where sentiment index is documented to have a predictive power of returns (using the above approach) are often young, emerging, and volatile ones. This further confirms our previous anticipation of how investors in novel markets often get swayed by waves of sentiment. In contrast, studies conducted in a well-established market,

[55] *Forecast* is the term used by Hamilton (1994) rather than the original author of the test.

such as Brown & Cliff (2004), show no short-term return predictability in their sentiment index.

Table 5.8: Granger-causality Test Results. Source: Own elaboration.

Null Hypothesis	Test Statistics	Critical Value	p-value
$\Delta\boldsymbol{SENT}$ does not Granger-cause RM	7.597	2.216	0.000
RM does not Granger-cause $\Delta\boldsymbol{SENT}$	7.305	2.216	0.000

5.4 A Simple Trading Strategy

Since the short-run return predictability of the composite sentiment index has been proved to exist, intuitively, one could think about the possibility of profitable trading strategies that are implemented based on this index. Thus, the following section will attempt to simulate and evaluate the economic value of a simple trading strategy with the sentiment index as the main indicator. The methodology of this section is inspired largely by Hilpisch (2014) and Chen et al. (2019).

Based on the sentiment index, one could try to predict (point estimate) the market return using the regression equation of the VAR model from section 5.3.2, given formally as:

$$\widehat{RM}_t = \widehat{\propto} + \sum_{i=1}^{5} \hat{\beta}_i \Delta SENT_{t-i} + \sum_{i=1}^{5} \hat{\delta}_i RM_{t-i} \tag{5.6}$$

where the constant $\widehat{\propto} = 0.002516$ and other lagged coefficients of $\Delta SENT$ and RM

($\hat{\beta}_i$ and $\hat{\delta}_i$, respectively) could be found in

Table 5.7. The rule to generate trading signals is simple, in which we go long ($BUY =$ 1) when the forecasted return is greater than 0, go short (BUY = -1) when the forecasted return is less than 0, and wait (do nothing) otherwise. In mathematical terms, this could be expressed as:

$$BUY_t = \begin{cases} 1, & if\ \widehat{RM}_t > 0 \\ -1, & if\ \widehat{RM}_t < 0 \\ 0, & otherwise \end{cases} \tag{5.7}$$

Thus, the cumulative return of the sentiment strategy at day t is given by:

$$R_t^{SENT} = \prod_{i=1}^{t}(BUY_i * RM_i + 1) - 1 \tag{5.8}$$

By fitting the sample data of RM and $\Delta SENT$ from 29/11/2014 to 20/07/2020 (a total of 2035 days) using equation 5.6, we receive the point estimates of RM (or $\widehat{RM}$) from 05/12/2014 to 25/07/2020. Next, trading signals and cumulative returns of the strategy will be derived accordingly, based on equation 5.7 and 5.8. The cumulative returns of sentiment-based strategy are then plotted against the ones of the buy-and-hold strategy in Figure 5.4. It could be seen that although the sentiment strategy could not capture the whole upside during the crypto market boom in late 2017 and early 2018, the strategy outperforms the market quite significantly over the whole sampling period. It appears that our strategy performs relatively well (in comparison with the market index) during the downturn after January 2018. Over the entire examining period, the sentiment-based strategy achieves a daily return of ~31bps[56], around 1.73

[56] 1 basis point (bp) is equal to 0.01%.

times of the buy-and-hold strategy return (~18 bps).

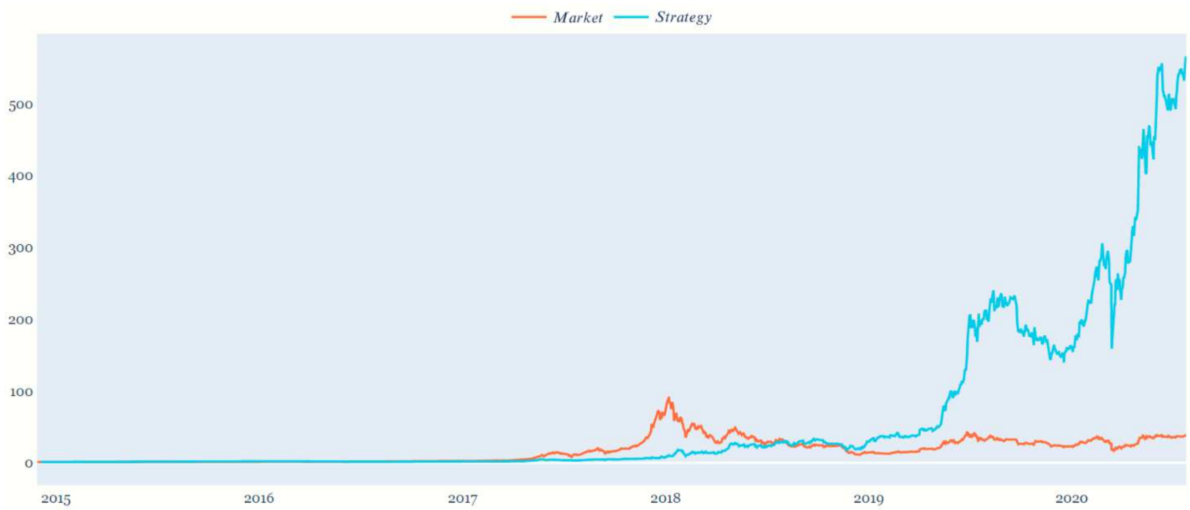


Figure 5.4: Cumulative Returns of the CRIX buy-and-hold Strategy & Sentiment-based Strategy (05/12/2014 to 25/07/2020). Source: Own elaboration.

The results indicate that the composite sentiment index can potentially gain value to the investors who use them as a trading indicator. In the meantime, it further confirms our argument that sentiment plays an important role in predicting future cryptocurrency market returns. However, an important note to be mentioned here is that there is a long way to go from a simulated strategy to a trading system applicable in the real-world. Speaking in an ambitious manner, the strategy in practice could either perform better or worse than the one simulated here. In a way, it could be worse, since in order to backtest the simulated strategy, operational issues (e.g. trade execution) and relevant market microstructure elements (e.g., transaction costs) are completely neglected. Generally speaking, these real-world factors often increase risks and diminish returns in our trading. In another way, the strategy could also be better since the one simulated here is very simple and there are a lot of rooms for improvement. For example, an investor may attempt to impose several more sophisticated fund management techniques than the one employed in our simulated strategy (i.e, long and short the same amount every time). An interesting question that might arise is what would happen if our trading volume varies according to the confidence interval of the predicted returns. The answer is difficult to say and we save this question for further researches or practical implementations in the future.

6 Conclusion

The thesis follows a *composite* approach in order to measure the unobservable factor of sentiment in the cryptocurrency market. While the chosen approach has been popularized by a number of studies investigating the sentimental effects in the traditional market of stocks, it is very likely that this is the first time such an approach has been utilized to quantify public sentiment in the innovative market of cryptocurrencies. The ambitious early motive is not only to discover an effective measure of sentiment in the crypto market, but also to capture empirical evidence on how this factor could influence the decision-making process of the crypto investors. In case this intention leads to significant results, the thesis can contribute to the existing literature a novel method to study sentiment in the crypto market.

The analytical process later proves the above conjecture to be correct. It begins by collecting a comprehensive set of sentiment indicators, which are divided into two major categories. The first group, referred to as **direct** measures, are computed by applying sentiment analysis on a fairly large textual dataset. This dataset consists of over 2 million messages on the social media platform *StockTwits*, 1.3 million submissions, and an astonishing number of 13 million comments on the discussion platform *Reddit*, posted over a 6-year period from November 2014 to July 2020. The lexicon employed for sentiment analysis is adopted from the study of Chen et al. (2019), in which the authors developed a sophisticated lexicon comprised of nearly 10,000 words specifically related to the crypto market. The second group, referred to as **indirect** sentiment proxies, consists of financial-based indicators (namely the volatility index *VCRIX* and the market trading volume) and the metrics of online platforms (including the volume of *Google Search*, *StockTwits* messages, *Reddit* submissions, and *Reddit* comments).

After individual indicators are collected, one crucial condition that must be fulfilled for the composite approach to appear valuable is the high correlations between individual indicators. Through examining their pairwise correlations, it turns out that this is indeed the case with our collected data, thus making the composite approach very attractive. For the purpose of isolating the individual indicators into the final composite index, the thesis employs the technique of *Principal Component Analysis*, which is inspired by two highly influential studies on sentiment in the stock market: Brown & Cliff (2004) and Baker & Wurgler (2007). Our final sentiment index is simply the differences of the first principal component of nine individual proxies.

The time series analysis in chapter 5 shows that the newly-created sentiment index has strong predictive power of cryptocurrency returns in the short term. To be more specific, positive (negative) changes in sentiment usually result in price movements in the same direction for nearly a week. The findings agree with the theoretical expectations mentioned in section 3.2. Interestingly, a simulated trading strategy utilizing the sentiment index as the base predictor of returns outperforms significantly the market index over our sampling period.

The empirical results might carry some profound implications for a variety of parties who are involved with cryptocurrencies. For the crypto traders or investors, a quite straightforward suggestion is that effective sentiment monitoring could possibly be one of the crucial keys able to help their trading strategies stay profitable during the downturns of the market. For the policymakers, given that the trends of sentiment appear to precede most of the large market movements, these folks could look deeper into how the index might interact with the crypto market in longer horizons. In case that kind of long-term relationship truly exists and could be documented, theoretically, they could use the index to predict the incidents of investors' irrational exuberance, or at least to some extent, assess the degree of irrationality in the market. From there, the policymakers could exploit the time to examine carefully which approach should be taken appropriately based on the current status of the market, and then issue timely and accurate regulations afterward. For the developers or other enthusiasts in the crypto industry, these individuals could employ the index as a basis to form their decisions on when and how to release a new application related to their cryptocurrencies. With the right use and the correct timing, one can exploit the market momentum to make his products recognized by more customers. While it is absolutely true that the real value of technological products depends prominently on such products' usefulness and functionality (which could only be improved through time by the efforts of the developers), some free marketing will not harm anyone. As suggested by the author of *Crossing the Chasm* Geoffrey A. Moore, a little precise timing is sometimes what it takes for a product to reach its desired destination.

Finally, the data used for this thesis is publicly available through the corresponding APIs mentioned explicitly in chapter 4. The reader might find the scripts (written in *Python*) to collect the required data and produce the analysis in the public GitHub repository at https://github.com/dang-trung/CryptoSentiment.

Bibliography

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

Akerlof, G. A., & Shiller, R. J. (2010). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. In *Princeton University Press*. https://doi.org/10.1111/j.1475-4932.2009.00576.x

Alexander, C., & Dakos, M. (2020). A Critical Investigation Of Cryptocurrency Data And Analysis. *Quantitative Finance*, *20*(2), 173–188. https://doi.org/10.1080/14697688.2019.1641347

Alrabadi, D., & Waleed, H. (2015). A new proxy for investor sentiment: Evidence from an emerging market. *Afro-Asian Journal of Finance and Accounting*, *5*(4), 334–343. https://doi.org/10.1504/AAJFA.2015.073471

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *Journal of Finance*, *59*(3), 1259–1295. https://doi.org/10.2139/ssrn.282320

Audrino, F., Sigrist, F., & Ballinari, D. (2018). The Impact of Sentiment and Attention Measures on Stock Market Volatility. *International Journal of Forecasting*, *36*(2), 334–357. https://doi.org/10.2139/ssrn.3188941

Avery, C., Chevalier, J. A., & Zeckhauser, R. J. (2011). The "CAPS" Prediction System and Stock Market Returns. *SSRN Electronic Journal.*

https://doi.org/10.2139/ssrn.1404750

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-section of Stock Returns. *Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. In *NBER Working Paper* (Issue No. 13189). http://www.nber.org/papers/w13189

Barberis, N., Shleifer, A., & Vishny, R. W. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, *49*(3), 307–343.

Barberis, N., & Thaler, R. H. (2003). A Survey of Behavioral Finance. In *Handbook of the Economics of Finance* (Vol. 1, pp. 1053–1128). Elsevier.

Baur, D. G., Dimpfl, T., & Kuck, K. (2018). Bitcoin, Gold and the Us Dollar – a Replication and Extension. *Finance Research Letters*, *25*, 103–110. https://doi.org/10.1016/j.frl.2017.10.012

Beaumont, R., van Daele, M., Frijns, B., Lehnert, T., & Muller, A. (2008). Investor Sentiment, Mutual Fund Flows And Its Impact On Returns And Volatility. *Managerial Finance*, *34*(11), 772–785. https://doi.org/10.1108/03074350810900505

Beck, T., Demirguc-kunt, A., Soledad, M., & Peria, M. (2006). Banking Services for Everyone? Barriers to Bank Access and Use around the World. *World Bank Policy Research Working Paper*, *No. 4079*.

Beer, F., Herve, F., & Zouaoui, M. (2013). Is Big Brother Watching Us? Google, Investor Sentiment And The Stock Market. *Economics Bulletin*, *33*(1), 454–466. https://doi.org/10.2139/ssrn.2185979

Ben-Rephael, A., Kandel, S., & Wohl, A. (2012). Measuring Investor Sentiment With Mutual Fund Flows. *Journal of Financial Economics*, *104*(2), 363–382. https://doi.org/10.1016/j.jfineco.2010.08.018

Berentsen, A., & Schar, F. (2018). A Short Introduction to the World of Cryptocurrencies. *Federal Reserve Bank of St. Louis Review*, *100*(1), 1–16. https://doi.org/10.20955/r/2018.1-16

Black, F. (1986). Noise. *Journal of Finance*, *41*(3), 529–543.

Boldbaatar, M., & Lee, C. L. (2015). Financial Accessibility and Economic Growth. *East Asian Economic Review*, *19*(2), 143–166. https://doi.org/10.11644/kiep.jeai.2015.19.2.294

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts The Stock Market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bouri, E., Gupta, R., & Roubaud, D. (2018). Herding Behaviour in the Cryptocurrency Market. *Montpellier Business School Working Paper 2018-34*, Article 2018–34.

Bouri, E., Lucey, B., & Roubaud, D. (2020). The Volatility Surprise of Leading Cryptocurrencies: Transitory and Permanent Linkages. *Finance Research Letters*, *33*. https://doi.org/10.1016/j.frl.2019.05.006

Bower, J. L., & Christensen, C. M. (1995). Disruptive Innovation: Catching The Wave. *Harvard Business Review*.

Brody, S., & Diakopoulos, N. (2011). Cooooooooooooooollllllllllllll!!!!!!!!!!!!!! using word lengthening to detect sentiment in microblogs. *Proceedings of the Conference on Empirical Methods in Natural Language Processing 2011*, 562–570.

Brown, G. W., & Cliff, M. T. (2004). Investor Sentiment and the Near-term Stock Market. *Journal of Empirical Finance*, *11*(1), 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001

Brown, G. W., & Cliff, M. T. (2005). Investor Sentiment and Asset Valuation. *Journal of Business*, *78*(2), 405–440. https://doi.org/10.1086/427633

Brown, S. J., Goetzmann, W. N., Hiraki, T., Shiraishi, N., & Watanabe, M. (2005). Investor Sentiment in Japanese and U.S. Daily Mutual Fund Flows. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.302829

Cao, M., & Rhue, L. (2019). Disposed of Bitcoin? Using the Disposition Effect to Understand Financial News Sentiment and Bitcoin Returns. *25th Americas Conference on Information Systems, AMCIS 2019*, 1–5.

Celeste, V., Corbet, S., & Gurdgiev, C. (2018). Fractal Dynamics and Wavelet Analysis: Deep Volatility Properties of Bitcoin, Ethereum and Ripple. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3232913

Cheah, E. T., & Fry, J. (2015). Speculative Bubbles in Bitcoin Markets? An Empirical Investigation Into the Fundamental Value of Bitcoin. *Economics Letters*, *130*, 32–36. https://doi.org/10.1016/j.econlet.2015.02.029

Chen, C., Despres, R., Guo, L., & Renault, T. (2019). What Makes Cryptocurrencies Special? Investor Sentiment and Return Predictability During the Bubble. *SSRN Electronic Journal*, 1–36. https://doi.org/10.2139/ssrn.3398423

Chen, H., Chong, T. T. L., & Duan, X. (2010). A Principal-component Approach To Measuring Investor Sentiment. *Quantitative Finance*, *10*(4), 339–347. https://doi.org/10.1080/14697680903193389

Chi, L., Zhang, Guang-sheng; Zhuang, X., & Song, D. (2012). A Study on Investor Sentiment Index and Chinese Stock Market Based on the Extended Kalman Filter Method. *Journal of Industrial Engineering and Engineering Management*.

Chong, T. T.-L., Cao, B., & Wong, W. K. (2017). A Principal Component Approach to Measuring Investor Sentiment in Hong Kong. *Journal of Management Sciences*, *4*(2), 237–247. https://doi.org/10.20547/jms.2014.1704206

Christensen, C., Johnston, C., & Barragree, A. (2000). After the Gold Rush: Patterns of Success and Failure on the Internet. *Innosight, LLC*.

Christensen, C., & Raynor, M. (2003). The Innovator's Solution: Creating and Sustaining Successful Growth. In *Harvard Business Review Press*. https://doi.org/10.1111/j.0737-6782.2005.116_1.x

Chuen, D. L. K. (Ed. . (2015). *Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data*. Academic Press. https://doi.org/10.1016/B978-0-12-802117-0.09989-6

Coinmap. (2020). *Crypto ATMs & merchants of the world | Coinmap.org*. https://coinmap.org/

CoinMarketCap. (2020). *Cryptocurrency Market Capitalizations | CoinMarketCap*. CoinMarketCap. https://coinmarketcap.com/

Corbet, S., Lucey, B. M., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a Financial Asset: A Systematic Analysis. *International Review of Financial Analysis*, *62*, 182–199. https://doi.org/10.2139/ssrn.3143122

Costelloe, K. (2017). *Bitcoin 'Ought to Be Outlawed,' Nobel Prize Winner Stiglitz Says*. Bloomberg.

https://web.archive.org/web/20180612224313/https://www.bloomberg.com/news/articles/2017-11-29/bitcoin-ought-to-be-outlawed-nobel-prize-winner-stiglitz-says-jal10hxd

Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., Gün Sirer, E., Song, D., & Wattenhofer, R. (2016). On Scaling Decentralized Blockchains. *Proc. 3rd Workshop on Bitcoin and Blockchain Research*, 106–125. https://doi.org/10.1007/978-3-662-53357-4_8

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS: Investor Sentiment and Asset Prices. *The Review of Financial Studies*, *28*(1), 1–32. https://doi.org/10.2139/ssrn.1509162

De Bondt, W. F. M., & Forbes, W. P. (1999). Herding in Analyst Earnings Forecasts: Evidence From the United Kingdom. *European Financial Management*, *5*(2), 143–163. https://doi.org/10.1111/1468-036X.00087

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1093/0198292279.003.0002

De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs In The West And Fails Everywhere Else*.

Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, *108*, 28–41. https://doi.org/10.1016/j.techfore.2016.04.005

Demirguc-Kunt, A. ., Klapper, L. ., Singer, D. ., Ansar, S. ., & Hess, J. . (2018). The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution. In *The World Bank*.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431.

Ellis, K., Lemma, A., & Rud, J.-P. (2010). Investigating the Impact of Access to Financial Services on Household Investment. *Overseas Development Institute*, *August*, 1–66. http://www.tzdpg.or.tz/fileadmin/_migrated/content_uploads/Impact_financial_service_on_investment_2010_02.pdf

European Central Bank. (2020). *Banknotes and Coins Circulation*. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/banknotes+coins/circulation/html/index.en.html%0A

Fenn, J. (1995). When To Leap On The Hype Cycle. *Gartner Research Report*, *June*, 1–5. http://www.cata.ca/_pvw522C275E/files/PDF/Resource_Centres/hightech/reports/indepstudies/Whentoleaponthehypecycle.pdf

Fenn, J., & Linden, A. (2003). Understanding Gartner's hype cycles. In *Strategic Analysis Report No R-20-1971. Gartner Research* (Issue May). https://www.gartner.com/en/documents/3887767

Fenn, J., & Raskino, M. (2008). Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time. In *Harvard Business Press*. https://lib.ugent.be/en/catalog/rug01:001311261

Finter, P., Niessen-Ruenzi, A., & Ruenzi, S. (2012). The Impact Of Investor Sentiment On The German Stock Market. *Zeitschrift Für Betriebswirtschaft*, *82*(2), 133–163. https://doi.org/10.1007/s11573-011-0536-x

Foster, R. N. (1988). Innovation: The Attacker's Advantage. In *Summit Booksooks*. https://doi.org/10.1016/0024-6301(87)90084-7

Franco, P. (2014). Understanding Bitcoin : Cryptography, Engineering and Economics. In *Wiley*. https://ebookcentral.proquest.com/lib/nmt-ebooks/reader.action?docID=1823060

Frazzini, A., & Lamont, O. A. (2008). Dumb Money: Mutual Fund Flows and the Cross-section of Stock Returns. *Journal of Financial Economics*, *88*(2), 299–322. https://doi.org/10.1016/j.jfineco.2007.07.001

Gaither, C., & Chmieleski, D. (2006). *Fears of Dot-Com Crash, Version 2.0*. Los Angeles Times. http://articles.latimes.com/2006/jul/16/business/fi-overheat16

Gkillas, K., & Longin, F. (2018). Is Bitcoin the New Digital Gold? Evidence From Extreme Price Movements in Financial Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3245571

Glushkov, D. (2006). Sentiment Beta. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.862444

Godin, B. (2006). The Linear Model of Innovation. *Science, Technology, & Human Values*, *31*(6), 639–667. https://doi.org/10.1177/0162243906291865

GoldPrice.org. (2020). *Gold Price on 31 August 2020*. GoldPrice.Org. https://goldprice.org/gold-price-today/2020-08-31

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and

Cross-Spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/doi:10.2307/1912791

Greenwood, R., & Nagel, S. (2009). Inexperienced Investors and Bubbles. *Journal of Financial Economics*, *93*(2), 239–258. https://doi.org/10.1016/j.jfineco.2008.08.004

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

Gurdgiev, C., & O'Loughlin, D. (2020). Herding and Anchoring in Cryptocurrency Markets: Investor Reaction to Fear and Uncertainty. *Journal of Behavioral and Experimental Finance*, *25*. https://doi.org/10.1016/j.jbef.2020.100271

Hafner, C. (2018). Testing for Bubbles in Cryptocurrencies with Time-Varying Volatility. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3105251

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham., R. L. (2009). *Multivariate Data Analysis*. Prentice Hall. https://doi.org/10.1002/9781119720492.ch8

Hamilton, J. D. (1994). Time Series Analysis. In *Princeton University Press*. https://doi.org/10.1016/B978-008044910-4.00546-0

Hanke, S. H., & Krus, N. (2012). World Hyperinflations. In *Cato Institute Working Paper*. https://ssrn.com/abstract=2226980

Hayes, A. S. (2017). Cryptocurrency Value Formation: An Empirical Study Leading to a Cost of Production Model for Valuing Bitcoin. *Telematics and Informatics*, *34*(7), 1308–1321. https://doi.org/10.1016/j.tele.2016.05.005

Hilpisch, Y. (2014). Python for Finance. In *O'Reilly Media*.

https://doi.org/10.1111/febs.12952

Hong, H., & Stein, J. C. (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *Journal of Finance*, *54*(6), 2143–2184.

Järvenpää, H., & Mäkinen, S. J. (2008). Empirically Detecting the Hype Cycle With the Life Cycle Indicators: An Exploratory Analysis of Three Technologies. *2008 IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2008*, 12–16. https://doi.org/10.1109/IEEM.2008.4737823

Jun, S. P. (2012). An Empirical Study Of Users' Hype Cycle Based On Search Traffic: The Case Study On Hybrid Cars. *Scientometrics*, *91*(1), 81–99. https://doi.org/10.1007/s11192-011-0550-3

Kahneman, D., & Tversky, A. (1974). Judgment Under Uncertainty. *Science*, *185*(4157), 1124-1131. https://doi.org/10.7551/mitpress/3080.003.0038

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kaminski, J. (2014). Nowcasting the Bitcoin Market with Twitter Signals. *ArXiv Preprint ArXiv:1406.7577*. http://arxiv.org/abs/1406.7577

Karalevicius, V., Degrande, N., & Weerdt, J. De. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *Journal of Risk Finance*, *19*(1), 56–75. https://doi.org/10.1108/JRF-06-2017-0092

Khan, M. A., & Ahmad, E. (2018). Measurement of Investor Sentiment and Its Bi-directional Contemporaneous and Lead-lag Relationship With Returns: Evidence From Pakistan. *Sustainability (Switzerland)*, *11*(1). https://doi.org/10.3390/su11010094

Kim, S. H., & Kim, D. (2014). Investor Sentiment From Internet Message Postings and the Predictability of Stock Returns. *Journal of Economic Behavior and Organization*, *107*(PB), 708–729. https://doi.org/10.1016/j.jebo.2014.04.015

Kline, S. J. (1985). Innovation Is Not A Linear Process. *Research Management*, *28*(4), 36–45. https://doi.org/10.1080/00345334.1985.11756910

Kline, S. J., & Rosenberg, N. (1986). An Overview of Innovation. *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, 275. https://doi.org/10.1142/9789814273596_0009

Kolesnikova, A. (2018). *VCRIX - Volatility Index for Crypto-currencies on the Basis of CRIX*. https://edoc.hu-berlin.de/bitstream/handle/18452/20056/master_kolesnikova_alisa.pdf?sequence=3&isAllowed=y

Kraaijeveld, O., & De Smedt, J. (2020). The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices. *Journal of International Financial Markets, Institutions and Money*, *65*. https://doi.org/10.1016/j.intfin.2020.101188

Kristoufek, L. (2013). Bitcoin Meets Google Trends and Wikipedia: Quantifying the Relationship Between Phenomena of the Internet Era. *Scientific Reports*, *3*(1), 1–7. https://doi.org/10.1038/srep03415

Krugman, P. (2018). *Bubble, Bubble, Fraud And Trouble*. The New York Times. http://ezproxy.unal.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsgin&AN=edsgcl.525529677&lang=es&site=eds-live

Kumar, A., & Lee, C. M. C. (2006). Retail Investor Sentiment and Return Comovements. *Journal of Finance*, *61*(5), 2451–2486.

https://doi.org/10.1111/j.1540-6261.2006.01063.x

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lawson, T. (1993). Keynes And Conventions. *Review of Social Economy*, *51*(2), 174–200. https://doi.org/10.1080/00346769308616163

Lazarsfeld, S. (1966). The Courage for Imperfection. *Journal of Individual Psychology*, *22*(2), 163–165.

Lee, C. M. C. ., Shleifer, A., & Thaler, R. H. (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance*, *46*(1), 75. https://doi.org/10.2307/2328690

Lee, T. B. (2013). Bitcoin Is A Disruptive Technology. *Forbes*. https://www.forbes.com/sites/timothylee/2013/04/09/bitcoin-is-a-disruptive-technology/#1fa027029562

Loughran, T., & McDonald, B. (2011). When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Maranville, S. (1992). Entrepreneurship in the Business Curriculum. *Journal of Education for Business*, *68*(1), 27–31. https://doi.org/10.1080/08832323.1992.10117582

Menkhoff, L., Schmidt, U., & Brozynski, T. (2006). The Impact of Experience on Risk Taking, Overconfidence, and Herding of Fund Managers: Complementary Survey Evidence. *European Economic Review*, *50*(7), 1753–1766.

Moore, G. A. (1991). Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers. In *New York HarperBusiness: Vol. Rev.* https://doi.org/10.1017/CBO9781107415324.004

Moore, G. A. (2014). *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers* (3rd ed.).

Mowery, D. C. (1983). Economic Theory And Government Technology Policy. *Policy Sciences*, *16*(1), 27–43. https://doi.org/10.1007/BF00138466

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Bitcoin.Org. https://bitcoin.org/bitcoin.pdf

Neal, R., & Wheatley, S. M. (1998). Do Measures of Investor Sentiment Predict Returns? *The Journal of Financial and Quantitative Analysis*, *33*(4), 523. https://doi.org/10.2307/2331130

O'Leary, D. E. (2008). Gartner's Hype Cycle and Information System Research Issues. *International Journal of Accounting Information Systems*, *9*(4), 240–252. https://doi.org/10.1016/j.accinf.2008.09.001

Oh, C., & Sheng, O. R. L. (2011). Investigating Predictive Power Of Stock Micro Blog Sentiment In Forecasting Future Stock Price Directional Movement. *International Conference on Information Systems 2011, ICIS 2011*, *4*, 2860–2877.

Oliveira, N., Cortez, P., & Areal, N. (2013). On The Predictability Of Stock Market Behavior Using Stocktwits Sentiment And Posting Volume. In *Portuguese conference on artificial intelligence* (pp. 355–365). https://doi.org/10.1007/978-3-642-40669-0_31

Oliveira, N., Cortez, P., & Areal, N. (2016). Stock Market Sentiment Lexicon

Acquisition Using Microblogging Data and Statistical Measures. *Decision Support Systems*, *85*, 62–73. https://doi.org/10.1016/j.dss.2016.02.013

Oliveira, N., Cortez, P., & Areal, N. (2017). The Impact Of Microblogging Data For Stock Market Prediction: Using Twitter To Predict Returns, Volatility, Trading Volume And Survey Sentiment Indices. *Expert Systems with Applications*, *73*, 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *ArXiv Preprint Cs/0205070*. https://doi.org/10.3115/1118693.1118704

Patch, A. R. (1984). Reflections on Perfection. *American Psychologist*, *39*(4), 386–390.

Patterson, M. (2018). *Crypto's 80% Plunge Is Now Worse Than Stocks' Dot-Com Crash - Bloomberg*. Bloombergoomberg. https://www.bloomberg.com/news/articles/2018-09-12/crypto-s-crash-just-surpassed-dot-com-levels-as-losses-reach-80

Perry-Carrera, B. (2018). Effect of Sentiment on Bitcoin Price Formation. *Economics in Trinity College of Duke University*, 49. https://sites.duke.edu/brianperrycarrera/files/2018/04/Perry-Carrera2018.pdf

Pitta, J. (1999). *Requiem for a Bright Idea*. Forbes. http://www.forbes.com/forbes/1999/1101/6411390a.html

Qiu, L., & Welch, I. (2004). Investor Sentiment Measures. *NBER Working Paper*, *No. 10794*. http://www.nber.org/papers/w10794

Rakovská, Z. (2018). Beliefs in Cryptocurrency: Exploring the Relationship between Sentiment and Bitcoin. *European Financial Systems 2018*, 567.

https://is.muni.cz/do/econ/sborniky/2018/Proceedings_finalni_verze_Septembe r_3.pdf#page=567

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects Of Twitter Sentiment On Stock Price Returns. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138441

Rao, T., & Srivastava, S. (2012). Analyzing Stock Market Movements Using Twitter Sentiment Analysis. *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, 119–123. https://doi.org/10.1109/ASONAM.2012.30

Ratcliffe, S. (2017). Oxford Essential Quotations. In S. Ratcliffe (Ed.), *Oxford University Press*. https://doi.org/10.1093/acref/9780191794230.001.0001

Rauchs, M., & Hileman, G. (2017). Global Cryptocurrency Benchmarking Study. In *Cambridge Centre for Alternative Finance*.

Renault, T. (2017). Intraday Online Investor Sentiment and Return Patterns in the U.s. Stock Market. *Journal of Banking and Finance*, *84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Rogers, E. M. (1962). Diffusion Of Innovations. In *New York: Free Press of Glencoe* (1st ed.). https://doi.org/10.1097/01.GIM.0000061743.67794.C4

Salac, A. (2019). *Forecasting of the Cryptocurrency Market Through Social Media Sentiment Analysis*. 1–12. http://essay.utwente.nl/78607/

Scheinkman, J. A., & Xiong, W. (2003). Overconfidence and Speculative Bubbles. *Journal of Political Economy*, *111*(6), 1183–1219.

Schroeder, S. (2020). *Defi Could Become the Next Big Thing in Finance*. Mashable.

https://mashable.com/article/defi-growing-fast/?europe=true

Sedgwick, K. (2018). *Bitcoin History Part 6: The First Bitcoin Exchange*. Featured Bitcoin News. https://news.bitcoin.com/bitcoin-history-part-6-the-first-bitcoin-exchange/

Shapiro, C., & Varian, H. R. (1999). Information Rules. *Harvard Business School Press*, 352. http://portal.acm.org/citation.cfm?doid=776985.776997%0Ahttp://ntst-aegean.puas.gr/sites/default/files/forum/information-rules.pdf

Shiller, R. J. (1987). Investor Behavior in the October 1987 Stock Market Crash: Survey Evidence. *NBER Working Paper Series*, 1–41. https://doi.org/10.3386/w2446

Shiller, R. J. (2014). *In Search of a Stable Electronic Currency*. The New York Times. http://www.nytimes.com/2014/03/02/business/in-search-of-a-stable-electronic-currency.html

Shiller, R. J. (2015). Irrational Exuberance. In *Princeton University Press.*

Shiller, R. J., Fischer, S., & Friedman, B. M. (1984). Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity*, *1984*(2), 457. https://doi.org/10.2307/2534436

Shiller, R. J., & Pound, J. (1989). Survey Evidence On Diffusion Of Interest And Information Among Investors. *Journal of Economic Behavior and Organization*, *12*(1), 47–66. https://doi.org/10.1016/0167-2681(89)90076-0

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. *Journal of Finance*, *52*(1), 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Sias, R. W. (2005). Institutional Herding. *SSRN Electronic Journal.*

https://doi.org/10.2139/ssrn.307440

Solt, M. E., & Statman, M. (1988). How Useful is the Sentiment Index? *Financial Analysts Journal*, *44*(5), 45–55. https://doi.org/10.2469/faj.v44.n5.45

Sun, L., Najand, M., & Shen, J. (2015). Stock Return Predictability and Investor Sentiment: A High-Frequency Perspective. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2692328

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Journal of Finance*, *63*(3), 1437–1467. https://doi.org/10.2469/dig.v38.n4.20

Thaler, R. H. (1985). Mental Accounting And Consumer Choice. *Marketing Science*, *4*(3), 199–214. https://doi.org/10.1287/mksc.1070.0330

Trimborn, S., & Härdle, W. K. (2018). CRIX an Index for cryptocurrencies. *Journal of Empirical Finance*, *49*(October), 107–122. https://doi.org/10.1016/j.jempfin.2018.08.004

U.S. Federal Reserve Board. (2020). *How Much U.S. Currency is in Circulation?* https://www.federalreserve.gov/faqs/currency_12773.htm

V. Bush. (1945). Science: The Endless Frontier. *Chemical and Engineering News*, *23*(16), 1420–1430. https://doi.org/10.1021/cen-v023n016.p1420

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning.

Entropy, *21*(6). https://doi.org/10.3390/e21060589

Veryzer, R. W. (1998). Discontinuous Innovation and the New Product Development Process. *Journal of Product Innovation Management*, *15*(4), 304–321. https://doi.org/10.1016/s0737-6782(97)00105-7

Visa. (2017). *Visa Fact Sheet*. 1. https://www.visa.gr/dam/VCOM/download/corporate/media/visanet-technology/aboutvisafactsheet.pdf

Welch, I. (2000). Herding among Security Analysts. *Journal of Financial Economics*, *58*(3), 369–396. https://doi.org/10.2469/dig.v31.n2.857

Whaley, R. E. (2000). The Investor Fear Gauge. *The Journal of Portfolio Management*, *Vix*.

Wolff-Mann, E. (2018). *"Only Good for Drug Dealers": More Nobel Prize Winners Snub Bitcoin*. Yahoo Finance. https://finance.yahoo.com/news/good-drug-dealers-nobel-prize-winners-snub-bitcoin-184903784.html

World Federation of Exchanges. (2020). *World Federation of Exchanges Database*. https://www.world-exchanges.org/%0A

World Gold Council. (2020). *How Much Gold Has Been Mined? | World Gold Council*. https://www.gold.org/about-gold/gold-supply/gold-mining/how-much-gold

Yuan, H., & Wang, S. (2018). Huobi Cryptocurrency Investor Sentiment Index Report-March. *Huobi Research Weekly*, *3*. https://huobiblog.zendesk.com/hc/en-us/article_attachments/360002880172/Huobi_Cryptocurrency_Investor_Sentiment_Index_Report-March.pdf

Zetter, K. (2009). *Bullion and Bandits: The Improbable Rise and Fall of E-Gold |*

WIRED. Wired. https://www.wired.com/2009/06/e-gold/

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter "I Hope It is Not as Bad as I Fear." *Procedia - Social and Behavioral Sciences*, *26*, 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562

Zweig, M. E. (1973). An Investor Expectations Stock Price Predictive Model Using Closed-End Fund Premiums. *The Journal of Finance*, *28*(1), 67. https://doi.org/10.2307/2978169