Université de Lille

LUMEN
Lille University Management – ULR 4999

La Dynamique des prix des Cryptomonnaies et ses déterminants

Thèse de doctorat de l'Université de Lille
préparée au LUMEN

École doctorale n°74 Ecole doctorale des sciences juridiques, politiques et de gestion (SJPG)
Spécialité de doctorat: Sciences de Gestion

Thèse présentée et soutenue à Lille, le 3 Septembre 2025, par

Pierre FAY

Composition du Jury :

Deven Bathia Professeur des Universités, Queen Mary University of London	Rapporteur
Jane Binner Professeur des Universités, Université de Birmingham	Examinateur
Jean-Gabriel Cousin Professeur des Universités, Université de Lille	Examinateur
Catherine D'Hondt Professeur des Universités, Université de Louvain	Rapporteur
Richard Taffler Professeur des Universités, Warwick Business School	Examinateur
Iryna Veryzhenko MCF-HDR, CNAM Paris	Examinateur
Fredj Jawadi Professeur des Universités, Université de Lille	Directeur de thèse
David Bourghelle MCF-HDR, Université de Lille	Codirecteur de thèse

Thèse de doctorat

Acknowledgments

This research would not have been possible without the support of many people.

I would first like to express my deep gratitude to my thesis supervisors, Dr. David Bourghelle and Dr. Fredj Jawadi. Their patience, kindness, and wise advice have been invaluable throughout this journey. I thank them for believing in me and for giving me the honor of supervising this thesis. I will cherish many memories of this experience, which will stay with me throughout my life. This work would not have been possible without their insightful comments, their high standards, and their dedication.

I would also like to warmly thank Dr. Philippe Rozin for his constant support, kindness, and encouragement since the beginning of this adventure. His invaluable advice also contributed to the success of this project, and his words often comforted me and gave me the strength to continue in times of doubt.

I am very honored that Dr. Devan Bathia and Dr. Catherine D'Hont agreed to be the reviewers for the thesis. Their respective expertise will greatly enrich this work. I would also like to thank Dr. Jane Binner, Dr. Jean-Gabriel Cousin, Dr. Richard Taffler and Dr. Iryna Veryzhenko for their participation on the exam board. Thank you for giving me your time and knowledge to evaluate my work.

My thanks also go to all the members of LUMEN and the doctoral school. The warm welcome and kindness I received in the laboratory created a highly fulfilling environment. I am particularly grateful to all the professors whose classes I was able to follow.

My thanks also go to the IAE Lille for placing their trust in me and allowing me to gain insights into the teaching profession through an experience as a temporary teaching and research associate (ATER), which proved to be a rich learning experience.

I would also like to thank the French Association of Institutional Investors (AF2I) for allowing me to benefit from the “Young Researcher 2022” grant and for being able to present my research at their “Time to Change 2023” event.

Warm thanks also to my colleagues and clients at Decathlon, Fremaux, Capgemini, Hermès and Valiuz. Thank you for your trust in me; it has been a pleasure working with you during this period.

I would like to express my sincere gratitude to my friends and family who shared this intense period of my life. Their support allowed me to overcome moments of doubt and to persevere in my efforts.

I dedicate a special thought to my daughter, born during this return to studies; may this PhD thesis be proof that you should never give up and that hard work always pays off.

Finally, heartfelt thanks to my wife, the one who saw a one-year return to school aside from my job turn into two, and finally seven years. I am hugely grateful for her love and support. This work would not have been possible without her patience and understanding.

Abstract

This thesis explores the behavioral foundations of Bitcoin price dynamics by focusing on the role of sentiment, emotions (fear, panic, joy), and investor attention. Using recent developments in nonlinear econometrics, advanced machine learning techniques, and high-frequency datasets, the study provides empirical evidence on the significant explanatory and predictive power of behavioral factors in cryptocurrency markets. Structured into four chapters, the thesis begins with a theoretical overview of blockchain technology and the main features of the crypto market, highlighting key stylized facts that challenge traditional assumptions of market efficiency and investor rationality. The remaining chapters are empirical and consist of three distinct essays. The first empirical essay examines the impact of investor sentiment and emotions on Bitcoin returns, the second investigates the role of sentiment and attention in shaping Bitcoin's volatility, and the third explores the potential of emotional indicators to detect speculative bubbles. Our research offers two main contributions. First, it develops a novel measure of investor sentiment and attention by applying natural language processing (e.g., FinBERT) to YouTube video content, helping us to differentiate between positive and negative emotions. Second, it uses various econometric models to demonstrate the key role of behavioral variables in forecasting returns, volatility, and the likelihood of bubbles. The findings reveal that blockchain literacy among investors significantly drives volatility, while emotional factors are instrumental in identifying speculative cycles. Moreover, the proposed models outperform traditional trading approaches

in terms of predictive accuracy.

Résumé

Cette thèse examine la dynamique du prix du Bitcoin en mettant l'accent sur l'influence des facteurs comportementaux, tels que le sentiment, les émotions (peur, joie), l'attention des investisseurs, etc. En mobilisant des outils récents de l'économétrie non linéaire, des techniques avancées de machine learning et des bases de données de haute fréquence, elle démontre l'importance de ces variables comportementales dans l'explication et la prévision de l'évolution du prix du Bitcoin. La thèse est structurée en quatre chapitres. Le premier chapitre, à vocation théorique, introduit les fondements technologiques de la blockchain et présente les principales caractéristiques du marché des cryptomonnaies. Il propose également un état des lieux des faits stylisés propres à ce marché, souvent difficilement conciliables avec les hypothèses d'efficience des marchés et de rationalité des investisseurs. Les trois chapitres suivants adoptent une approche empirique à travers trois essais distincts. Le deuxième chapitre analyse l'impact du sentiment et des émotions des investisseurs sur la rentabilité du Bitcoin. Le troisième chapitre explore le rôle du sentiment et de l'attention des investisseurs dans la dynamique de la volatilité. Le quatrième et dernier chapitre s'intéresse à la capacité des variables émotionnelles à détecter les phases spéculatives et les bulles sur le marché du Bitcoin. Cette recherche apporte deux contributions principales. D'une part, elle propose une nouvelle mesure du sentiment et de l'attention des investisseurs en exploitant des données issues de vidéos YouTube, à l'aide d'outils récents de traitement du langage naturel basés sur le machine learning, tels que FinBERT. Cette approche permet

notamment de distinguer les émotions positives et négatives exprimées dans les contenus audiovisuels. D'autre part, à travers différents modèles économétriques, la thèse met en évidence le rôle déterminant des facteurs comportementaux dans la prévision de la rentabilité, de la volatilité et de la probabilité d'émergence de bulles sur le marché du Bitcoin. Par exemple, nos résultats montrent que les compétences des investisseurs en matière de blockchain constituent un déterminant majeur de la volatilité du Bitcoin. De plus, les facteurs émotionnels permettent d'identifier les phases de surchauffe du marché, tels que les cycles haussiers excessifs ou les bulles spéculatives. Enfin, nos modèles surpassent les approches traditionnelles de trading en termes de performance prédictive.

Contents

List of Figures

List of Tables

List of Acronyms

ACPR: Autorité de Contrôle Prudentiel et de Résolution

AML: Anti-Money Laundering

AMLD: Anti-Money Laundering Directive

AMMs: Automated Market Makers

AMF: Autorité des Marchés Financiers

AUM: Assets Under Management

BC: blockchain-competent

BPPI: Bitcoin Price Pattern Index

BIC: Bénéfices Industriels et Commerciaux

CBDC: Central Bank Digital Currency

CEX: Centralized EXchange

CFT: Countering the Financing of Terrorism

DAO: Decentralized Autonomous Organization

DeFi: Decentralized Finance

DEX: Decentralized EXchange

e-CNY: Digital yuan

EMH: Efficient Market Hypothesis

ETFs: Exchange-Traded Funds

ETH: Ethereum

ESG: Environmental, Social, and Governance

GDP: Gross Domestic Product

HTML: HyperText Markup Language

IBIT: iShares Bitcoin Trust

ICO: Initial Coin Offering

IMF: International Monetary Fund

IPO: Initial Public Offering

KYC: Know Your Customer

LPs: Liquidity Providers

NASDAQ: National Association of Securities Dealers Automated Quotations

NBC: non-blockchain-competent

NFT: Non-Fungible Token

NLP: Natural Language Processing

MiCA: Markets in Crypto-Assets

PACTE: Plan d'Action pour la Croissance et la Transformation des Entreprises

PBoC: People's Bank of China

PoS: Proof of Stake

PoW: Proof of Work

PACTE: Plan d'Action pour la Croissance et la Transformation des Entreprises

PSAN: Prestataire de Services sur Actifs Numériques

RGPD: Règlement Général sur la Protection des Données

S&P500: Standard and Poorś 500

SEC: Securities and Exchange Commission

SWIFT: Society for Worldwide Interbank Financial Telecommunications

TFR: Transfer of Funds Regulation

TWh: Terawatt-hours

UST: TerraUSD

VAR: Vector Autoregression

WHO: World Health Organization

General introduction

Bitcoin is by far the most popular and liquid crypto-currency. Its capitalization is currently estimated at 2.181 trillion dollars. Blockchain technology and crypto-currencies have attracted a great deal of attention in recent years. For different reasons, they have become a source of controversy among investors, speculators, economists, and regulators. First, this burgeoning new technology is gaining ground in the global economy by facilitating cross-border trade and financing activities. Its ability to serve as a competitive alternative to international fiat money systems is also crucial. Second, the fact that regulators are concerned about the involvement of crypto-currencies to finance illegal activities underscores the importance of electronic currencies in all financing cycles. However, many economists argue that pricing of these digital currencies is too volatile to allow them to be used as a store of value, a unit of account, or a medium of exchange, and that this volatility cannot reasonably be explained by economic fundamentals. The first cryptocurrency, Bitcoin, was built around two powerful narratives. Sometimes presented as digital gold for its inflation-resistant capacity, Bitcoin is also seen by some as a financial revolution, allowing investors to move away from the traditional banking system. Academic research has examined the nature of Bitcoin, and while it shares similarities with gold (Dyhrberg 2016), notably through resistance to inflation thanks to its supply limited to 21 million units, the price of Bitcoin is not significantly correlated with that of gold (Yermack 2015; Corbet et al. 2018; Hu, Parlour, and Rajan 2019; Liu and Tsyvinski 2021). The asset is nonetheless of interest to portfolio managers, enhancing diversification in otherwise traditional asset portfolios (Bouri, Molnár, et al. 2017; Corbet et al. 2018).

While the Bitcoin narrative is generally about a financial revolution that allows us to break away from the traditional system (De Filippi 2022; Giacomini and Rossi 2025), the reality is quite different. First, it is important to recall that Bitcoin is an early version of cryptocurrency, and despite the significant potential of blockchain, its underlying technology, there are still few real-world applications that use Bitcoin. Second, most transactions currently take place in centralized exchanges that are subject to strict regulatory frameworks such as Know Your User (KYC), Anti-Money Laundering (AML), and Combating the Financing of Terrorism [1]. The rules allow regulators to monitor transactions and to detect suspicious movements of money in cryptocurrencies. Third, its growing adoption by institutional investors, particularly since the arrival of Bitcoin ETFs and the inclusion of MicroStrategy in the Nasdaq 100, makes it an asset that is increasingly linked to the traditional economy. Indeed, Bitcoin's movements are significantly correlated with stocks like the NASDAQ, the SP500, and the US Dollar (Wu 2025; Aliyev and Eylasov 2025).

The recurring up-and-down cycles, characteristic of the cryptocurrency market, raise many questions about the determinants of Bitcoin's value and its investors' rationality. Cryptocurrencies are a rapidly expanding field of research. Existing work attempts to understand the factors that influences the extreme volatility and returns of this asset (Koutmos 2020; Naeem et al. 2020; Pagnotta 2021), to evaluate its informational efficiency (Urquhart 2016; Tran and Leirvik 2019), and

[1] source: https://finance.ec.europa.eu/financial-crime/anti-money-laundering-and-countering-financing-terrorism-eu-level_en

to understand the impact of regulations (Auer and Claessens 2018; Shanaev et al. 2020; Liu and Tsyvinski 2021), and its links with traditional financial markets such as gold (Dyhrberg 2016; Corbet et al. 2018; Panagiotidis, Stengos, and Vravosinos 2018; Hu, Parlour, and Rajan 2019; Liu and Tsyvinski 2021), the stock market (Ciaian, Rajcaniova, and Kancs 2016; Bouri, Molnár, et al. 2017; Corbet et al. 2018), economic uncertainty (Demir et al. 2018; Bouri and Gupta 2019; Shaikh 2020), and other cryptocurrencies (Wei 2018; Hu, Parlour, and Rajan 2019; Lahiani, jeribi, and Jlassi 2021).

Despite a wide-ranging literature on the topic and the European Union's regulatory advances, the study of Bitcoin price dynamics remains a complex field of research given the rapidly evolving regulations, technologies, and investor behavior. There is still no unified economic theory to explain the determinants of Bitcoin's value despite investors' massive adoption of the asset and its high volatility. At a time when political initiatives are part of a broader trend toward financial deregulation, questions may arise about the future of the international monetary system[2]. Indeed, the lack of oversight and regulations could encourage speculative behavior and excessive risk-taking, make it more difficult to combat money laundering and terrorist financing, and increase the influence of private actors in the creation and circulation of money. These issues represent a major challenge to the integrity of the international financial system.

The present thesis aims to provide answers to the issues of bitcoin returns and related risk, with a focus on several questions: What are the factors that influence Bitcoin returns? How can we model and predict its excessive volatility? And can

[2]https://www.mediapart.fr/journal/international/210125/la-cryptomonnaie-trump-ou-la-tentation-de-la-privatisation-totale-du-pouvoir-monetaire

we identify and characterize the speculative bubbles that form in this market? To answer these questions, we develop specific econometric models involving behavioral factors that we compare with historical data. Interestingly, these behavioral factors are captured using machine learning tests. The results not only improve our understanding of cryptocurrency price dynamics but also contribute to the development of new investment and risk management tools.

The thesis is organized as follows: In Chapter 1, we explain the key concepts to give the reader an understanding of the basics of cryptocurrencies. We explore how Bitcoin works, its history, and its regulatory framework, and we analyze its market structure. The chapter provides a concise overview of cryptocurrency markets and its rules. In Chapter 2, we focus on factors that influence Bitcoin returns, exploring an innovative new avenue: the impact of YouTube videos on Bitcoin returns. Indeed, despite YouTube being the second most used and visited social media platform in the world, no research has studied its impact on the price of Bitcoin. Yet, its video format makes it the ideal platform for technical analysis and information exchange. Indeed, YouTube is full of market analyses and price predictions that can encourage viewers to buy or sell. We also look at the impact of investor sentiment and attention to YouTube videos and propose a direct measure of investor attention. By analyzing a large database, covering the period 2018–2023, we seek to determine whether attention and sentiment indicators extracted from YouTube can improve the prediction of Bitcoin returns.

This study reveals multiple dynamics with respect to investor attention and sentiment. Regarding investor attention (daily number of videos viewed), the results indicate an overall improvement in the predictability of Bitcoin returns through analysis of YouTube activity. However, this influence is not uniform and varies significantly, depending on the content of the videos. Our findings show that increased aggregated attention on the platform is associated with positive pressure on Bitcoin prices. Disaggregating topics, it appears that interest in 'Tutorial' videos dedicated to learning and understanding how Bitcoin works, 'Robot Trading' strategies involving automated systems, and 'Network Activities' discussions, related to blockchain infrastructure, exert a positive influence on Bitcoin returns. However, this impact appears to be transient, generally fading after one or two days. Conversely, increased attention to videos focused on influential 'Personalities' in the crypto ecosystem or on videos discussing 'NFT and Metaverse' is correlated with a negative reaction in terms of Bitcoin returns. Building on these observations regarding the topical influence of YouTube attention, we next explore how the sentiment expressed in these videos—whether positive or negative—further impacts Bitcoin's price dynamics. Analyzing investor sentiment (the relationship between the number of positive and negative videos and Bitcoin returns) reveals that the impact of opinions expressed varies by topic and sentiment. Indeed, only optimistic videos specifically addressing influential 'Personalities' or 'Institutional and Central Banks' (videos with titles mentioning the central bank or a major company like "MicroStrategy", "PayPal") appear to have a significant effect on Bitcoin prices. The study observes an interesting phenomenon of investor overreaction following a positive shock in the number of positive videos regarding "institutional investors and central banks". An initial surge in optimism can lead to an initial price increase, followed by a market correction in the following days, suggesting possible profit-taking after the initial enthusiasm. Notably, the number of negative videos on YouTube, whether considered in aggregate or by individual topic, appears to be driven primarily by Bitcoin's past performance rather than a predictive indicator of its future movements.

Finally, the study explores the interaction between investor attention and sentiment, revealing a bidirectional causal relationship. When a shock occurs simultaneously to investor attention and the number of negative videos posted, this can potentially self-reinforce, leading to sharp price movements. The significance of these findings is twofold. For cryptocurrency investors and traders, this research offers valuable insights into the influence of social media dynamics on the Bitcoin market, particularly YouTube. Distinguishing the impact of different types of video content can help refine investment and trading strategies. Identifying specific attention and sentiment signals that precede price movements can help to better anticipate short-term trends. For regulators, the study highlights the important role of social media platforms in shaping prices in the digital asset market. Understanding how aggregated attention and sentiment on platforms as popular as YouTube can impact market dynamics is essential for effective oversight. The discovery of potential feedback loops exacerbating positive or negative returns may require careful consideration in the development of regulatory frameworks aimed at ensuring the integrity and stability of the cryptocurrency market.

In Chapter 3, we study the role of behavioral factors, especially the impact of market sentiment on Bitcoin volatility. We also extend the existing literature by exploring a new question: does investors' blockchain knowledge impact Bitcoin volatility? To this end, we use natural language processing techniques to extract user sentiment on Reddit, as well as users' knowledge level (whether or not they use blockchain vocabulary in their messages) in order to examine how the sentiment of blockchain-competent and non-blockchain-competent investors influences Bitcoin price fluctuations. Our study presents significant results regarding the influence of investors' blockchain knowledge level. First, the model reproduces the dynamics of Bitcoin volatility, validating its suitability to analyze this complex issue. Second, identification of distinct low- and high-volatility regimes provides further evidence of asymmetry and nonlinearity in the relationship between investor sentiment and Bitcoin volatility. Third, the research demonstrates the crucial value of distinguishing between positive and negative sentiment and comments of blockchain competent (BC) and non-blockchain-competent investors (NBC) to better comprehend the dynamics of these volatility regimes. Fourth, positive comments by BC users are observed to have a contrasting impact on volatility: they increase it during periods of low volatility, potentially boosting activity and optimism, but reduce it during periods of high volatility, suggesting a stabilizing or stepping-back effect of skilled investors in the face of uncertainty. Fifth, in low volatility regimes, an increase in the number of negative comments, whether by BC or NBC users, tends to increase volatility, possibly by sowing concern and uncertainty. However, in periods of high volatility, only negative comments by NBC users continue to amplify volatility, indicating that less informed investors are more likely to be affected by negative sentiment in times of crisis. Finally, the study shows that news articles do not influence BC and NBC investor sentiment in the same way, highlighting the importance of blockchain expertise on how information is perceived and absorbed by investors. These findings are important for cryptocurrency investors, traders, and regulators alike. For investors and traders, our research offers a more detailed understanding of the factors that influence Bitcoin volatility, a crucial element in risk management and decision-making. Distinguishing between the impact of informed and uninformed investor sentiment, as well as how this impact varies depending on the volatility regime, can help us identify more accurate signals and anticipate market movements. For regulators, our study highlights the complexity of cryptocurrency market dynamics and the need to account for the heterogeneity of actors. Understanding how different investor segments react to information and contribute to volatility is essential to the development of effective monitoring and regulatory policies. The difference in reaction to media sentiment between informed and uninformed investors highlights the potential role of investor education in reducing excessive cryptocurrency volatility.

Finally, in Chapter 4, we look at ways to detect Bitcoin bull run regimes. We explore the application of concepts from behavioral finance to the cryptocurrency market, focusing specifically on the detection of bull runs. We introduce new indicators to detect price escalation, and use investors' attention and sentiment to test whether they can discriminate between bull run and non-bull run states, putting forward a logistic regression model to identify the rising phase of a crypto bubble.

The main results highlight the interest of our new indicator, the Bitcoin Pattern Price Indicator (BPPI), in distinguishing speculative bubble periods from other periods. In similar vein, the sentiment (number of positive and negative videos views) and emotions (fear and joy) features are shown to be relevant signals for identifying bull run phases, suggesting that both positive and negative attention of investors, fear and joy, play a crucial role in their development. Application of a logistic regression model significantly improved the performance of Bitcoin bull run detection compared to the traditional PSY (Phillips, Shi, and Yu 2015) methodology, highlighting the power of behavioral data-based approaches for identifying these periods. The findings are important for both cryptocurrency investors and traders. The ability to identify speculative bubble phases more accurately and earlier can help provide them with valuable risk management tools. The indicators used could serve as early warning signals, allowing stakeholders to take advantage of the bubble's upward phase and reduce their exposure before it potentially bursts, thus protecting their capital. The increased effectiveness of behavioral finance and machine learning-based models provides regulators with more powerful tools to monitor the market and implement preventive measures to ensure cryptocurrency market stability.

This research offers multiple contributions across conceptual, empirical, managerial, and regulatory domains. First, by exploring the impact of YouTube videos and investor knowledge on Bitcoin price dynamics, the study contributes to the field of behavioral finance within the cryptocurrency context; it underscores the complex nature of its valuation and highlights the importance of acknowledging the effects of sentiment and emotions on valuation.

Empirically, the thesis provides substantial evidence for the influence of previously unexplored factors on Bitcoin returns and volatility. The detailed analysis of YouTube videos disaggregated by topic and sentiment demonstrates that not all attention or sentiment has the same effect. The distinction between blockchain competent and non-competent investor sentiment on Bitcoin volatility offers a clearer understanding of Bitcoin investors. Finally, leveraging sentiment and emotional factors, the successful application of a logistic regression model for bull run detection provides a new tool to identify speculative periods in cryptocurrency markets.

From a managerial perspective, the findings offer practical insights for cryptocurrency investors and traders. The ability to distinguish the impact of diverse YouTube content types and to understand how blockchain-competent investors versus non-blockchain-competent investors' sentiment affects volatility allows for more sophisticated investment and risk management strategies. The early warning signals provided by the logistic regression model for bull run detection can help to capitalize on upward trends and mitigate losses during high speculative periods. Finally, for regulators, the research underscores the critical role of social media in price formation and the importance of blockchain literacy among crypto investors. It highlights the need to include investor education and influencer accountability initiatives in regulatory frameworks to ensure market integrity and stability.

Despite its contributions, this work has certain limitations inherent to the evolving and complex nature of the cryptocurrency market. First, the study period for analysis of YouTube's impact begins in 2017 and ends in 2025. Although significant, it may not capture all phases of the cryptocurrency market evolution, which is constantly changing. The dynamics observed could well change, for example, with the recent increase in adoption by institutions or with future regulations (as an example, the adoption of the MiCA European regulation and the fact that USDT has been banned from Binance for European users).

Secondly, in terms of correlation, while Granger causality tests are used, establishing true causality in financial markets, especially with complex psychological factors, is inherently difficult. The thesis demonstrates significant influence and lead-lag effects of sentiment and attention, but there are certainly other unobserved variables that also come into play. Thirdly, from the investor's point of view, many of them are classified as "blockchain-competent/non blockchain-competent", based on a dictionary-based NLP method for identifying technical discussions on Reddit. While innovative, this method may be limited in its ability to fully capture the nuanced levels of blockchain expertise, or may misclassify users who are knowledgeable but do not use technical jargon in their posts. Fourth, the type of investment is of some importance, and it would be interesting to analyze the impact of institutional vs. retail investor sentiment. Different investor profiles are only known by exchanges and are not shared. This highlights a significant data accessibility limitation, preventing a more granular understanding.

Finally, with increased institutional involvement, Bitcoin's role is expanding to include traditional portfolio diversification and even, for some, a store of value. The models and conclusions drawn might need recalibrating as adoption continues to grow.

Chapter 1: Theoretical framework

1 History and concepts

1.1 The history of cryptocurrencies

In the 1990s, with the arrival of the Internet, access providers became the nodes through which all users' private information passed. This new phenomenon posed a risk to users' privacy and freedom of expression. Aware of this danger, a group of computer scientists with an interest in cryptography began working on solutions to improve privacy and protect freedom of expression. Among these so-called cypherpunks were small groups, some developing encrypted means of communication to enable users to communicate more securely and to preserve their privacy, while others, like the "crypto-anarchists", sought to develop means of independence from big business and the state. To be able to live freely on the Internet without leaving a trace, you need a decentralized, anonymous payment system. For the crypto-anarchists, disintermediation of financial transactions is an important issue. Several attempts were made to create an anonymous virtual currency whose transactions could not be traced. In 1997, Adam Back created the Hashcash system. The system obliged any individual wishing to use an online service to perform a certain amount of "work" on his or her computer in order to gain access to the service. This "proof of work" discouraged certain behaviors such as spam, denial-of-service attacks, and so on. Although the system was not initially intended to create a new digital currency, it was this “proof of work” system that would be adopted by Bitcoin in the future. In 1998, Wei Dai, a computer engineer, took up the proof-of-work principle to create a virtual currency, and introduced B-Money. However, B-Money suffered from a major limitation: with no means of verification, pirates could use the virtual currency and generate it several times over (known as the “double spending” problem). The project was abandoned and was never developed to its conclusion. That same year, Nick Szabo, a cypherpunk cryptographer, described a similar model called BitGold. He took up what had been developed on B-Money but added an asynchronous system to verify transactions. During proof-of-work, the solution to each equation became part of the next equation to be solved, thus resolving the “double spending” problem. This system would also be adopted by Bitcoin. A few years later, in 2008, an individual (or group of individuals), hiding behind the pseudonym Satoshi Nakamoto, published a white paper on a mailing list dedicated to cryptography called “Bitcoin: A peer to peer electronic cash system” Nakamoto (2008). The white paper describes a cryptocurrency using blockchain technology. A few months later, on January 3, 2009, the first 50 bitcoins attributed to Satoshi Nakamoto were generated... the beginning of blockchain technology and cryptocurrencies. To this day, nobody knows who Satoshi Nakamoto is, but he or she is the largest holder of this cryptocurrency, with an estimated fortune of 1.1 million bitcoins. Inspired by Bitcoin, many cryptocurrencies were born, and the technology has been evolving ever since. There are many cryptocurrencies, each of which may be based on different blockchain technologies or uses. Blockchain technology is still in its infancy, and its evolution is at the heart of what is known as “web 3.0”. Its application in the field of finance is called decentralized finance or DeFi[1].

1.2 Towards web 3.0

Blockchain technology was born with bitcoin, but it is the evolution of this technology that led to the emergence of a decentralized Internet known as Web 3.0, mainly thanks to its notion of the "smart contract" (popularized in particular thanks to Ethereum). We will come back to this notion later. In the first version of the Internet, Web 1.0, information was made available to users via an Internet platform. Websites were static, and the web was principally made up of HTML pages and hypertext links. These pages allowed users to consult information from the company that owned the website. As technology evolved, web pages became dynamic thanks to new programming languages. This is what we call Web 2.0. Users can now fill in forms, and data is sent and processed by servers, then stored in databases belonging to the companies that own the website. Since this revolution, a great deal of content has been created by users. We have seen the emergence of social networks (Facebook, Twitter, etc.), Wikipedia, personal blogs, YouTube, and even sites offering services such as TripAdvisor and Blablacar. Users generate a lot of data via these free services (e.g. Facebook), which is then used by companies to optimize their processes in what is known as "big data". In Web 2.0, users can now exchange information with each other via a website belonging to a trusted third party (the company owning the website). Data passes through company-owned servers, and information is mainly centralized. This is the web we know today. However, the web continues to evolve and we are now at the begin-

[1]For more information about the history of cryptocurrencies, readers may refer to De Filippi (2022) and Voshmgir (2020), which inspired this chapter.

ning of Web 3.0, a new vision of the future of the internet. In Web3, users have their own identity on the Internet, enabling them to identify themselves, authenticate themselves, and give/receive access to data. Data is stored decentralized in a blockchain, which is far more secure than a traditional server, limiting the risk of hacking and loss of information in the event of failure. Only the owner of the data decides whether or not to offer access to his or her data to a third party. This identity is also linked to a means of payment, enabling the user to access services or make purchases without going through a trusted third party, since it is the technology that ensures transaction security, rather than the trusted third party. It is also to this identity that information issued by institutions, such as a driving license or diploma, could be linked. This information could then be verified by third parties, without necessarily giving them all the information about a user. Indeed, if a rental agency scans a driver's license and receives validation of verification of the driver's license and identity directly from the prefecture, it does not need to have the driver's exact address or age. This change in operation is also in line with regulations introduced by various countries on privacy protection (e.g. in Europe with the RGPD), which gives users greater control over their personal data. In Web3, an identity can also be attached to an object such as a car. It could, for instance, be assigned an owner, and this information could be validated by the prefecture. We could then check that a car is not stolen before it accesses any service such as buying petrol or using a parking lot (thus rendering a stolen car unusable). Smart contracts are an important feature of Web3. These are algorithms which, once published, run autonomously on a blockchain when they are called up. The algorithms can be verified by anyone with the computer skills to read them. To fully understand how a smart contract works, let's imagine a simple use case: disintermediating the sale of a vehicle. To change ownership and complete the sale, all the current owner (the seller) and the future owner (the buyer) would have to do is call up a smart contract with their 2 respective digital signatures. After verifying the 2 signatures and the presence of the funds required for the purchase, the smart contract would transfer the money from the buyer's account to the seller's account and notify the prefecture of the change of ownership. In addition to the money transfer, the "owner" information attached to the car's identity would be changed immediately and automatically by the prefecture upon receipt of the money by the seller. As the smart contract code is freely readable by everyone on the blockchain, everyone can check what the script is going to do and thus have confidence in its execution. We are currently only at the beginning of Web 3.0. The technology, its use, and the associated regulatory framework are currently undergoing rapid change, and many potential uses remain to be discovered[2].

1.3 The blockchain

Bitcoin and most cryptocurrencies are based on blockchain technology[3].The goal of this technology is to be able to store information permanently, decentralized, and securely. Permanent, because once the information is entered into the blockchain, it can no longer be modified or deleted. Decentralized, it relies on a peer-to-peer network. The database is shared by all network participants. Each transaction must be validated by the entire network to be included in the database. Secure, because it is governed by a consensus algorithm and cryptographic algorithms

[2]For more information about Web 3 and its prospects, readers may refer to Voshmgir (2020).

[3]To know more about the Bitcoin and how does blockchain works, the reader can refer to Antonopoulos (2017).

that make it very difficult to hack. The blockchain is therefore like a highly secure public ledger. It is a so-called “trustless” system, allowing network participants to work together without needing to trust each other. Indeed, trust is placed in the technology (the code is public, readable by everyone).

Blockchains can be public or private. On a public blockchain (like Bitcoin or Ethereum), anyone can view the information in the shared database, enabling transactions to be validated using an algorithm. Anyone can participate in validations and therefore attempt to cheat; a high level of security is required to ensure it works to prevent cheating. Computing power is required to operate and validate transactions. The blockchain compensates the stakeholders who allocate their computing power to its network by issuing its own cryptocurrency (also known as a protocol token).

On a private blockchain, the reading of database information and transaction verification is controlled by a limited number of trusted nodes, still operating via a distributed consensus mechanism, but which is generally lighter and less resource-intensive. On the other hand, the blockchain is vulnerable to collusion between validators. In short, the blockchain is a secure, decentralized ledger. It can be managed publicly or privately. You can store whatever you want in each block. The ledger is shared by all participants in the blockchain network. Once information is recorded in the blockchain, it is permanently stored. Miners perform the calculations necessary to secure the network in exchange for payment in protocol tokens. Cryptocurrencies are therefore a specific use of the blockchain, where the ledger is used to record user transactions. Although this blockchain is described in the media as a revolution, it is based on several pre-existing technologies that form the blockchain:

Peer-to-Peer and Decentralized Databases

The networks on which cryptocurrencies are based are peer-to-peer networks, notorious for sharing hacked files. Each user on the network is called a node. All databases are fully or partially duplicated across the network's various nodes.

Asymmetric cryptography

Well-known in computer science, asymmetric cryptography allows users to identify and authenticate themselves, and to decode messages using two keys generated by an algorithm. The public key is used to encrypt data and send messages to the holder of the private key, who will be the only one able to read them. It also allows verification that a message has been signed by a specific person. The private key is used to decrypt data encoded with the public key or, conversely, to sign a document to authenticate the signature using the public key. On the Bitcoin network, for example, the public key allows network members to verify who initiated the transaction. Only the owner of the bitcoins is authorized to carry out transactions. To do this, they use their private key to sign transactions and to validate, for example, a change of ownership.

Hash functions

A hash function allows a unique character string (called a hash) to be generated from a digital element, allowing it to be uniquely identified. Each—even minimal—modification of the element will result in a modification of this character string. Hashes allow two elements to be quickly compared. Indeed, only two identical elements will have the same hash, making it easy to identify from the hash whether a document has been modified or not. Bitcoin is based on the SHA256 hashing algorithm. It is this hash that allows network members to verify that the database has not been corrupted.

Mining

Mining creates currency and secures the network by validating transactions. In practice, mining involves providing the network with computing power in exchange for compensation in the network's cryptocurrency (the protocol token). Each node in the network can provide the following services:

- Propagate information within the network
- Verify transactions
- Add a block

In Bitcoin, the difficulty of the calculations performed to create a block adapts to the network's computing power in order to maintain a pre-specified rate of issuing new currency units. A Bitcoin block is created on average every 10 minutes. The more miners there are, the greater the difficulty of the calculations required to maintain this 10-minute interval between each block. The total number of Bitcoins issued is limited to 21 million. The last of these will be issued around 2040. Solving a mathematical problem to add a block to the blockchain is called "proof of work" (usually referred to by its abbreviation: "PoW"). This algorithm is currently very secure but criticized due to its high energy consumption. Newer platforms are turning to new security mechanisms such as Proof of Stake (PoS), which was adopted in 2022 by Ethereum. In this protocol, only people with large amounts of protocol tokens can validate blockchain transactions. Proof of Stake is based on the principle that a validator with a large amount of cryptocurrency has no incentive to take action against the network because, if cheating were proven, the cryptocurrency would lose a significant amount of value.

Distributed consensus

Distributed consensus is the process by which a consensus is found among all the nodes on the network in order to validate the transaction. This algorithm is more or less power-consuming, depending on the degree of trust that the players place in each other (the greater the number of verifications, the greater the computing power required). Bitcoin is a public blockchain. As such, it assumes that all users can validate transactions and thus attempt to cheat on the network. It therefore uses proof-of-work, a highly secure, but also highly resource-intensive method, hence its high energy cost. The algorithm is a program launched by the miners. To help it evolve, most miners have to update their program to a new version. This process is known as a "fork". It is what the Ethereum blockchain did in the event known as "The Merge", which switched from a PoW (Proof of Work) algorithm to a PoS (Proof of Stake) algorithm in 2022[4].

1.4 The Smart Contracts

Since the environment is decentralized, there is no trusted third party to validate transactions. Anyone on a blockchain network can issue a transaction. When a transaction is issued by a user, how can we ensure that it is not fraudulent? The consensus algorithm allows all nodes to agree on how to validate transactions and

[4]For more information about "The Merge", readers can refer to https://www.coindesk.com/learn/what-is-the-merge-and-why-has-it-taken-so-long

to reach a consensus among all the nodes in the network to validate the transaction. This algorithm will be more or less resource-intensive depending on the degree of trust the actors have in each other (the more extensive the verifications, the more computing power it will require). Bitcoin is a public blockchain. It assumes that all users can validate transactions and therefore attempt to cheat on the network. It therefore uses proof-of-work. This method is very secure but also very resource-intensive, hence its high energy cost.

The algorithm is a program launched by miners. To make it evolve, the majority of miners must update their program to a new version. This is called a "fork." This is what the Ethereum blockchain did with the event called "The Merge," which, in 2022, transitioned from a PoW (Proof of Work) algorithm to a PoS (Proof of Stake) algorithm.

1.5 Decentralized Autonomous Organizations (DAO)

Web3 allows people to organize themselves around a common goal through rules written algorithmically in smart contracts. These organizations are called Decentralized Autonomous Organizations (DAOs). Since the rules are written algorithmically in a public blockchain, they are accessible to anyone with the skills to read them, and they are immutable. People in the organization therefore no longer need to trust each other to work together, since the rules are written and executed automatically by the algorithm. Since the rules of a DAO are immutable, it is important that they are bug-free, otherwise, a hacker could detect a security flaw and exploit it to steal money or privileges. This is what happened, for example, with the "The DAO" project, which connects project leaders and investors to enable them to raise money in a completely decentralized way. In the traditional financial system, banks have the power to cancel an illegitimate transaction after the fact, but in a blockchain network, once a transaction is sent, it is immutable unless the majority of the network's nodes make an update to cancel it. This is called a "hard fork", but it requires significant coordination among blockchain stakeholders, potentially endangering the blockchain as it often divides the community. Following The DAO hack, the community behind the Ethereum blockchain (on which The DAO was based) synchronized to allow the blockchain to 'roll back' via a major update. The community then split into two separate projects: Ethereum (ETH) and Ethereum Classic (ETC) and the network lost some of its miners.

1.6 Tokens

A token is something that has existed in the physical world for a long time. A token can represent an economic value or a right of access. We currently use them regularly in the form of gift cards, loyalty points, casino tokens, club membership cards, company access cards, bus tickets, concert tickets, or simply a euro coin. These different tokens are more or less secure to prevent counterfeiting, and require verification by a trusted third party (e.g. for a concert ticket, the concert organizer scans your ticket at the entrance to verify its validity). Bitcoin introduced a new type of token: cryptographic tokens. These tokens are simply records in the blockchain. They are therefore very secure. The tokens can represent a right of ownership, a vote, or a right of access to a product or service. They are represented by a record in the blockchain, linked to a unique address that represents an identity. Only the holder of the private key to this address can access the token's contents. Cryptocurrencies found on exchanges are primarily tokens. Not all of them represent money; in fact, there are potentially an infinite number of possible tokens, each with a specific use. They can be classified according to their technical functioning. Bitcoin is a native token, also called a protocol token. These are tokens linked to a public blockchain that allow the individuals who operate it to be compensated. Indeed, to function, the blockchain needs computing power to verify and validate transactions. It therefore encourages individuals to provide computing power in exchange for a token reward. This computing power is used to operate and secure the blockchain. There are also tokens that represent economic value, access rights, or voting rights. These tokens are called application tokens. They are based on a blockchain but are not used to compensate the individuals who operate it. The Ethereum blockchain, for instance, has only one native token (ETH) and multiple application tokens, each with its own purpose. They can be classified according to what they represent: an asset token, an access right token, or a voting right. They can be classified according to their fungibility. A token is fungible if it can be replaced by the same unit of the same type of token; it must also be able to be divided and aggregated. Conversely, a token is non-fungible if it cannot be exchanged for a token of the same type and generally cannot be divided or aggregated. A token that represents a unique work of art is non-fungible. In fact, it cannot be exchanged for the same value because the work is unique. Therefore, there is no equivalent exchange. Furthermore, a work of art cannot be aggregated with another work of art; it is non-fungible. Conversely, a bitcoin is exchangeable with another bitcoin without the user gaining or losing anything. Bitcoin can be divided or aggregated; it is fungible. They can be classified according to their transferability. A token representing a movie ticket may or may not be transferable depending on the ATM's policy. A token representing a driver's license will be unique and non-transferable. They can be classified according to their short-term stability. For a token to be used as a medium of exchange, short-term stability is a prerequisite for it to serve as a unit of account. In the traditional economy, it is the central bank that manages the issuance of its currency and ensures its stability. In decentralized finance, some tokens are designed with mechanisms to ensure price stability. These are called stablecoins. They can be classified according to their opacity. On the Bitcoin network, transactions are public and anyone can access them, but the information is anonymized. Indeed, all of a user's transactions are tied to one or more addresses belonging to them. With the advances in Big Data, if we have enough information to cross reference, we can potentially find out who the person is behind a Bitcoin address. It then becomes possible to trace all past transactions, and if a Bitcoin is used in an illegal transaction, it also becomes possible to "blacklist" it so that it is no longer accepted as a means of payment. Conversely, "privacy tokens" like Zcash or Monero allow for greater anonymity in exchanges and ensure greater data opacity. In addition, they can be classified according to their supply. Each token has its own issuance. Bitcoin's issuance is fixed and limited to 21 million, but tokens representing access to a resource will logically be limited by the access capacity and average frequency of access to that resource. For example, a token representing a concert ticket will necessarily be limited to the number of seats available in the venue. Finally, they can be classified according to their life cycle. A token such as a bus ticket can only be used once and is then destroyed. A token representing loyalty points can be automatically destroyed if it has not been used after an expiry date. Most tokens representing assets will not have an expiry date and can be traded indefinitely.

1.7 Perspectives

The technology associated with Bitcoin, in other words, blockchain, has been widely discussed. Although this technology serves as a means of disintermediation when used for cryptocurrencies, it can be used in various fields to optimize certain processes. Below are some examples of possible applications:

Finance: Reducing transaction costs

Today, the SWIFT network is an interbank network used by over 10,000 financial institutions worldwide. The network relies on trusted intermediaries, but some banks do not have direct access to this system. Thus, many international transfers must go through several banks, each of which charges a fee, making the transfer process longer and more expensive. On the blockchain, borders do not exist, and no intermediary is necessary to transfer funds from one address to another so no one can make a profit from the transfer process. Transaction costs are limited to the cost of operating the network.

Intellectual property

Today, to prove ownership of an invention, one must prove that the creation date was earlier. This is usually done by mail via a Soleau envelope. The blockchain would make allow the digital fingerprint (the hash) of the document or invention to be stored and to time-stamp it in order to obtain a result similar to the Soleau envelope but in a more secure manner (indeed, in the blockchain, it cannot be lost or falsified).

Education: Certification of diplomas

Blockchain allows a digital fingerprint of the diploma to be recorded and certified (via a signature made with the issuing institution's key). This makes it easy to verify candidates' diplomas. This initiative was developed by the University of Nicosia in Cyprus in 2014, also adopted by MIT in the United States and ESCP in France[5].

Life insurance: Automatic payment of premiums to beneficiaries

Blockchain would allow compensation to automatically be paid to the beneficiaries registered in the contract through a mechanism that allows the death of policy-holders to be verified at regular intervals via specific functions called "oracles". This would make the process very quick and very inexpensive.

Health: Traceability of medicines

There are numerous counterfeit medicines worldwide, mainly in Asia and Africa (1 in 10 medicines in developing countries, according to the WHO). Blockchain could help combat this phenomenon by ensuring the traceability of medicines and certifying their authenticity.

Central Bank Digital Currencies

On December 4, 2019, Francois Villeroy de Galhau, Governor of the Bank of France, stressed the need to develop a central bank digital currency (CBDC)[6]. China was the first major economy to move forward on the issue with testing of the digital yuan (e-CNY) from 2021. This digital currency was developed by the People's Bank of China (PBoC) to compete with private digital payment methods WeChat Pay and Alipay, which are widely used by Chinese users[7]. The Fed was also working on a Central Bank Digital Currency until Donald Trump ended the

[5]https://www.escp.eu/news/diplomes-MS-MSc-ESCP-certifies-sur-la-blockchain
[6]https://acpr.banque-france.fr/system/files/2024-11/20191204_discours_villeroy_de_galhau_fr.pdf
[7]https://www.nytimes.com/2021/03/01/technology/china-national-digital-currency.html

project in 2025[8]. Currently, central banks lack direct ties with consumers and businesses. A CBDC would potentially give central banks direct access to consumers' banking data, allowing them to track all consumer and business transactions, and to better control their monetary and fiscal policies. It would also be possible to prohibit the purchase of certain goods to a specific public (e.g. block the purchase of cigarettes to under 16s).

Legal tender, the case of El Salvador

In 2021, El Salvador declared Bitcoin a "legal tender," meaning it must be accepted for all payments in the same way as banknotes. Salvadorans are a significant diaspora in the United States, sending a very large sum of money back to their home country every year, representing 20% of El Salvador's GDP, or more than $6 billion. According to their president, establishing Bitcoin as legal tender would potentially allow Salvadorans to save on bank fees during their transactions. Since Bitcoin is highly volatile, when paying in Bitcoin through the Chivo app (El Salvador's official wallet), the user can choose to immediately convert it to its dollar equivalent to stabilize the value of their transaction. In 2021, the National Bank of El Salvador announced that it held 550 Bitcoins in its coffers[9] (or more than 21 million euros at the time of this announcement in 2021). In December 2024, El Salvador announced that it held 5,980 Bitcoins, or nearly 600 million dollars [10]. On December 18, 2024, El Salvador reached an agreement with the IMF to end Bitcoin as legal tender in El Salvador in exchange for a $1.4 billion loan to strengthen its fiscal stability. Salvadoran merchants can now voluntarily accept

[8] source: https://shorturl.at/gF0Cm

[9] https://www.latribune.fr/opinions/tribunes/au-salvador-la-bataille-du-bitcoin-se-gagnera-sur-son-adoption-891878.html

[10] https://bitcoin.gob.sv/address/32ixEdVJWo3kmvJGMTZq5jAQVZZeuwnqzo

Bitcoin as payment but are no longer required to do so. The government also agreed to gradually reduce its crypto portfolio. Beyond the Blockchain technology and its broad applications, a deeper understanding of how cryptocurrency prices are formed is crucial. The subject of Market microstructure delves into the rules governing how buyers and sellers interact, how information is incorporated into prices, and how liquidity is provided and consumed. By examining these details, we can better understand the dynamics of price discovery, transaction costs, and overall market efficiency.

2 Market Microstructure

In this section, we examine how the cryptocurrency market functions[11]. We begin by describing its unique organization, highlighting the various types of exchanges, and the distribution of market capitalization across different cryptocurrencies, exchanges, and market participants. Particular attention will be paid to price formation mechanisms (Kyle 1985; Easley and O'Hara 1987), transaction costs (Stoll 1978; Roll 1984), and the importance of liquidity (Amihud and Mendelson 1986; Grossman and Miller 1988; Amihud 2002; Pástor and Stambaugh 2003) in this market. We also address the issue of market efficiency, assessing the extent to which cryptocurrency prices reflect available information and whether arbitrage opportunities exist. This analysis will give us a better understanding of the unique characteristics of the cryptocurrency market and lay the foundation for further developments in our research.

[11]To know more about cryptocurrency microstructure, readers can refer to https://shorturl.at/gD1wA

2.1 Market organization

Capitalization and the dominance of cryptocurrencies

The total market capitalization of cryptocurrencies represents the combined value of all cryptocurrencies in circulation, calculated by multiplying the price of each asset by its circulating supply. Total market capitalization is presented in Figure 1. As of December 2024, the total market capitalization of cryptocurrencies was estimated at approximately $3.7 trillion. The market is highly volatile, and its capitalization has increased significantly over time. In fact, it was around $260 billion in January 2020.

Figure 1: Volume and total capitalization (source: coinmarketcap.com)

This capitalization is distributed among more than 20,000 cryptocurrencies but in a very uneven way. Indeed, the 10 most capitalized cryptocurrencies represent over 82% of the total capitalization, with the top 100 generally taken as a market index. This distribution evolves over time. Bitcoin was the first known cryptocurrency and it dominates the market with 57.7% of total capitalization (December 2024). Ethereum, the blockchain at the origin of smart contracts, continues to develop since many web 3.0 applications are created via this blockchain, which can explain its gain in market share over time.

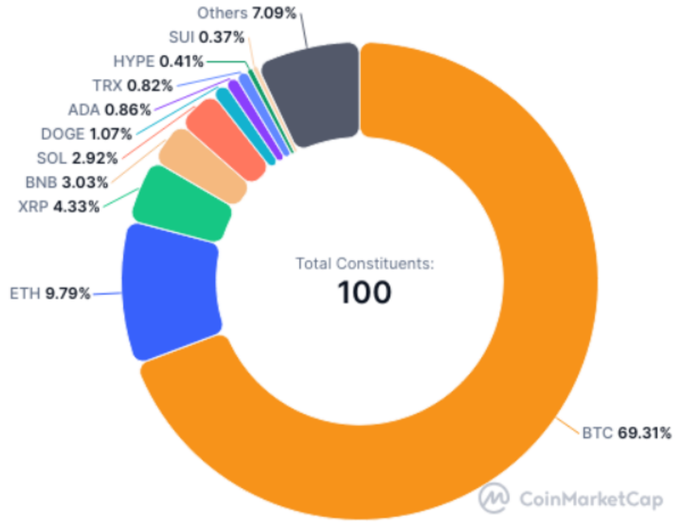


Figure 2: Market share of diverse cryptocurrencies, excluding stablecoins (source: coinmarketcap.com, 26 May 2025)

Rank (by capitalisation)	Name	Symbol	capitalization ($)	Dominance
1	Bitcoin	BTC	$2,174,945,221,047	63.36%
2	Ethereum	ETH	$307,301,190,523	8.951%
3	Tether	USDT	$152,762,877,211	4.4485%
4	Ripple	XRP	$135,610,448,417	3.9537%
5	Binance Coin	BNB	$94,995,968,005	2.7718%
6	Solana	SOL	$91,560,454,166	2.6674%
7	Circle	USDC	$61,565,319,220	1.7936%
8	Dogecoin	DOGE	$33,595,931,544	0.9787%
9	Cardano	ADA	$26,859,077,463	0.7827%
10	Tron	TRX	$25,810,998,985	0.7519%

Source: coinmarketcap.com

Table 1: Market share of the 10 most capitalized cryptocurrencies (May 2025)

Primary and secondary markets

The cryptocurrency market is similar to traditional financial markets in that they can be divided into two activities: the primary market and the secondary market. The primary market is the 'new' market and is where securities are issued. In traditional markets, intermediaries can purchase securities (equities, bonds...) through a bank or a company at the time of their creation. The exchanges that take place provide financing to the companies, states, or communities that issue the securities. The primary market enables the allocation of resources, connecting those who provide and those who seek capital. In the cryptocurrency market, securities are issued via an ICO (Initial Coin Offering), generally through a private company. The tokens issued are then exchanged for other tokens (often Ethereum). Unlike a stock, which allows you to own a share of the company's capital, when you buy a cryptocurrency, you buy the token and the rights associated with it (which differ from one token to another depending on its function). The tokens issued during the ICO represent a pre-payment for the future funded service. With the funding from investors, the company issuing the tokens must then develop its service. ICOs allow project launchers to avoid the traditional system, which would generally not lend as much money at such an early stage of the project and would sometimes ask for shares in the company in exchange. ICOs are therefore more like a crowdfunding process than an IPO, since all the people who own the exchanged tokens can participate in the process to gain access to the company's service or product, but do not generally obtain a share of the issuing company's capital. ICOs are still very risky for investors and project leaders due to the lack of legal frameworks worldwide. In Europe, the MICA Directive has strictly regulated

ICOs since December 30, 2024 [12].

The secondary market, the “second-hand” market, connects supply and demand, those who want to sell their already issued securities and those who want to buy them. This is what is generally referred to by individuals as “the stock market” when referring to traditional financial markets. In the cryptocurrency market, the secondary market is conducted on exchanges. These platforms are open 24/7, so there are no closing hours as in traditional markets, and trading is continuous.

Volumes and dominance of trading platforms

Exchanges match supply and demand, set prices, and clear and settle cryptocurrency transactions. They act as intermediaries in cryptocurrency exchanges, charging fees for these services. They may be centralized or decentralized.

- **Centralized exchange platforms**

  A centralized exchange platform (CEX for “Centralized EXchange”) is managed by a company that has full control over the exchange. Cryptocurrency exchanges are therefore handled by a third party on the platform, which has its own funds and is subject to the regulatory framework of the countries in which it operates. The regulations generally impose obligations such as the implementation of a KYC (“Know your customer”) system, or compliance with AML (Anti Money Laundering) directives in order to secure the platform’s users and avoid illegal activities. Each platform user has an account with an associated wallet. Users can generally buy cryptocurrencies in fiat currency directly from the platform and/or deposit their cryptocurrencies

[12] source: https://www.amf-france.org/en/news-publications/depth/mica

there. Exchanges are generally recorded in the database, but transactions on the blockchain are only really carried out when withdrawals/deposits are made from these wallets, which means very low transaction costs for traders and rapid transaction execution.

Rank	Name	Volume /24H in Billions ($)	Market share (%)
1	Binance	11.84	22.47%
2	Bybit	2.87	5.45%
3	Bitget	2.47	4.69%
4	HTX	2.44	4.63%
5	OKX	2.24	4.25%

Source: https://coinmarketcap.com/

* Volume on 24h the 26/05/2025

Table 2: Top 5 Centralized Exchanges by Volume (December 20, 2024)

Centralized exchanges represent the largest market share, but the distribution of activity reported in Table 2 is very uneven. Indeed, Binance is the market leader with a volume that represents approximately 4 times its largest competitor. Binance would represent more than 20% of all crypto exchanges (CEX and DEX combined) according to cryptorank.io. The second market share is attributed to the Bybit platform with only 5.45% of the market share[13].

- **Decentralized exchange platforms**

  Decentralized exchanges (DEXs) are trading platforms coded as smart con-

[13]source: https://cryptorank.io/exchanges the 26/05/2025

tracts and uploaded to a blockchain. Once deployed on the blockchain, they are theoretically impossible to stop. Since the exchange is a smart contract, it is also impossible to comply with regulations in place in different countries. DEXs can therefore be used in countries where cryptocurrency trading is prohibited (where users cannot access CEXs). Every transaction made on a DEX is carried out directly on the blockchain.

Rank	Name	Volume(24h) ($)	Market Share (in %)
1	Pancake Swap V3	7.23B	13.74%
2	Shadow	624M	1.18%
3	Hyperliquid	414.2M	0.79%
4	Uniswap V3 (Ethereum)	367.3M	0.70%
5	Aerodrome	317.5M	0.60%

Source: https://cryptorank.io/exchanges/dex

Volume the 25/05/2025

Table 3: Top 5 Decentralized Exchanges by Market Share

The leaders among decentralized platforms are reported in Table 3. The main CEX is Pancake Swap V3 with a 13.74% market share. Note that this ranking includes version numbers for the exchange names because once uploaded to the blockchain, it is impossible to delete a smart contract. Taking Uniswap as an example, after V2, a new version (V3) was launched to improve the liquidity rules of the old one. The majority of users now use version 3, but some users may still use version 2.

Although significant amounts of money are traded on decentralized exchanges, their market share is negligible compared to the main centralized platforms.

Market players

The world of cryptocurrencies attracts both retail and professional investors. On the one hand, there are retail cryptocurrency investors who represent a variety of motivations and profiles. Among them, three main profiles can be distinguished, each with its own characteristics and motivations:

- Speculators, who seek high returns, often invest for the short term and hope to profit from price fluctuations. Their behavior is often influenced by the fear of missing out (FOMO) on an opportunity, and they often have low risk aversion (Chohan and Kerckhoven 2023; Choudhary et al. 2024).
- Th Enthusiasts, for whom blockchain technology and the potential for disruption are of major interest. They are convinced of the long-term potential and tend to hold on to their cryptocurrencies, betting on their future appreciation (Srivastava, Singh, and Rana 2024).
- Investors who already invest in other markets and want to add Bitcoin to their portfolio to improve their diversification (J. H. Kim 2022) [14].

Retail investors are often young and financially educated. They are on average willing to take greater risks, regularly change their positions, and are more prone to behavioral biases (Hackethal et al. 2022). They are also sensitive to ESG investments (Ciaian, Cupak, et al. 2022) and obtain their information through online media and social networks (K. T. Kim and Fan 2025; Hackethal et al. 2022). They appear more inclined to adopt a trend-following strategy (Auer and Tercero-Lucas 2022).

[14] Exemple: https://www.blackrock.com/us/financial-professionals/insights/bitcoin-unique-diversifier

On the other hand, institutional players, who generally arrived in the crypto ecosystem later, have injected significant liquidity into the market. Some financial institutions offer derivative products such as futures, options, and exchange-traded funds (ETFs) based on cryptocurrencies. These products offer institutional investors more familiar and regulated ways to gain exposure to cryptocurrencies.

As reported in Figure 3, between 2018 and 2022, the assets under management (AUM) of crypto funds began growing. The arrival of Bitcoin ETFs pushed this adoption further as aggregate assets under management totaled over 125 billion in May 2025 [15].

[15] https://shorturl.at/7YFWj

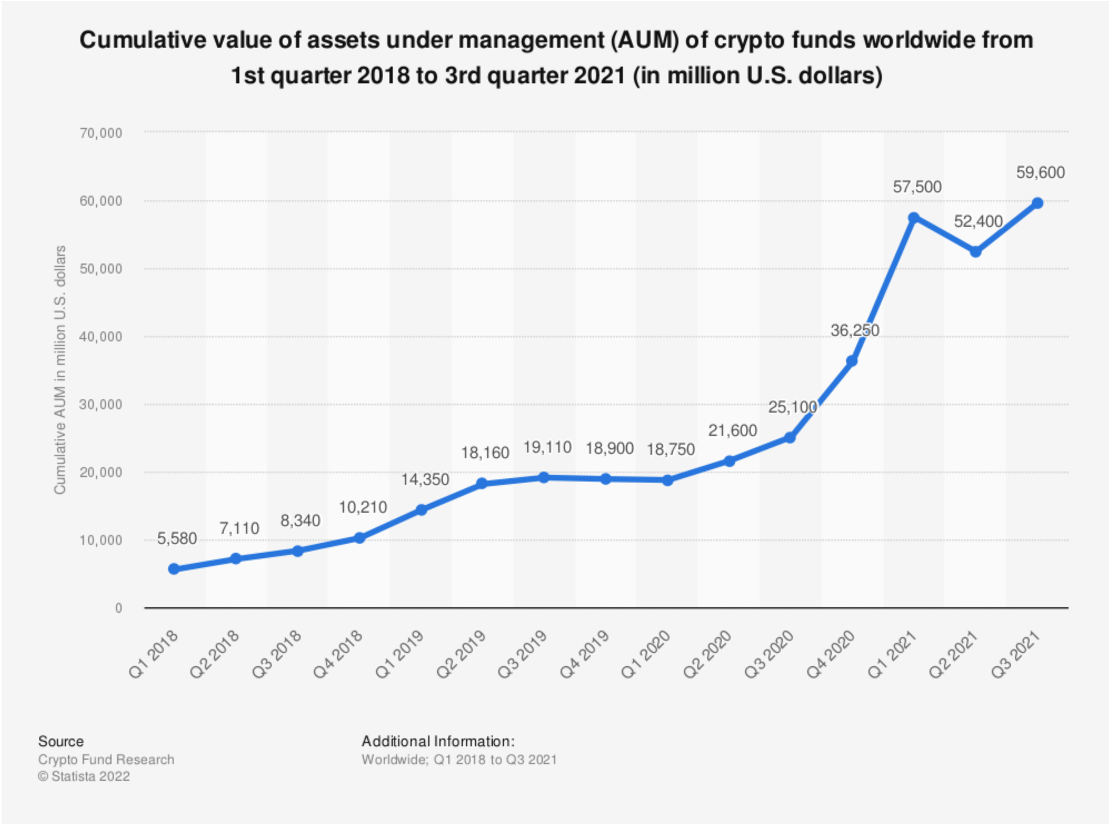


Figure 3: Crypto assets under management (source: Crypto Fund Research - Statista 2022)

A good example of this institutional adoption is BlackRock's iShares Bitcoin Trust (IBIT) which reached $50 billion in assets under management (AUM) in just 227 trading days, breaking the previous record of 1,323 days for an ETF to reach that milestone [16]. While institutional adoption has rapidly increased since the launch of ETFs, retail investors still account for around 88% of circulating Bitcoin supply. Institutional investors may be traditional structures that have adapted their operations to crypto-assets or structures created specifically for this investment

[16] https://www.investopedia.com/spot-bitcoin-etf-biggest-winners-and-losers-one-year-on-8771158

category.

Among institutional investors, we can cite several categories:

- Crypto funds: these funds aim to generate high returns by exploiting the volatility of cryptocurrencies through a range of active strategies (Huang, Lin, and P. Wang 2022), with some speculating on the short term while others invest for the long term. Examples of these crypto funds include Pantera Capital, Polychain Capital, Elwood Asset Management, and Ark Invest. This type of investor has been able to generate alpha over the past few years (Bianchi and Babiak 2022).
- Brokers: offer over-the-counter trading services and can also act as market makers on several platforms. One example is the French company Woorton. Brokers often have intuitive interfaces and educational resources, making them ideal for those new to the cryptocurrency world. They generally offer more comprehensive and accessible customer service than traditional exchanges, such as buying or selling cryptocurrencies at fixed prices, thereby avoiding the complexity of order books on exchanges.
- Lending platforms: these are platforms that allow users to lend for interest and borrow using their digital assets. Examples of this category include platforms such as Nexo and BlockFi.
- Payment service providers: these allow businesses to be paid in cryptocurrencies and then exchange their cryptocurrencies for fiat currency. Examples of this category include BitPay and Coinbase.

Approval of the first crypto ETFs (Exchange Traded Funds) by regulators marked a major turning point in cryptocurrencies' integration into the traditional financial system. In 2021, the first Bitcoin futures ETFs were launched, allowing investors to gain exposure to Bitcoin futures contracts. However, it was only in January 2024 that investors could invest in 'spot' ETFs, where companies directly hold Bitcoin. The first Bitcoin spot ETFs were approved by the SEC in January 2024 and are managed by major investment firms such as Ark Invest, BlackRock, Bitwise, and Grayscale.

2.2 Price formation mechanisms

Cryptocurrency price formation mechanisms are based on market principles similar to those of traditional assets, but with specific features related to their decentralized structure and 24/7 operation.

On crypto trading platforms, buy and sell orders are collected and organized by price in the order book, which allows interaction between buy (ask) and sell (bid) orders. Each transaction results from the meeting of a buyer and a seller willing to trade at a given price. The order book is viewable by platform users and updated in real time. The number of orders placed for each price on the bid and ask sides is called "market depth." Each trading pair has its own order book. To trade Bitcoin (BTC) for Tether (USDT), for example, there is a specific order book for that pair (BTC/USDT) which includes all orders to buy Bitcoin and to sell Bitcoin for USDT. Buy orders are grouped by price, and the total volume per price level at which buyers are willing to buy can be seen in the order book. Sell orders are also grouped by price in the order book, so that the total volume per price level at which sellers are willing to sell can also be viewed.

The minimum possible price fluctuation size between bid and ask orders per pair is called the “tick size”. This is defined by the exchange, and its size allows liquidity to be optimized [17]. The impact of tick size is a widely studied topic in the literature as it influences trading strategies, especially in markets with order books, which rely on the price and the order arrival in the book to determine which order will be executed first. A tick size that is too large reduces liquidity and increases the bid-ask spread, but a tick size that is too small increases order fragmentation (Harris 1994). Many studies show that a smaller tick size generally results in a smaller bid-ask spread and lower transaction costs (Harris 1994; Porter and Weaver 1997) even if the impact on market depth and the level of transparency in the liquidity supply is complex (Bourghelle and Declerck 2004). In the cryptocurrency market, tick sizes are generally small, leading traders to rely on the order book to position themselves in order to be executed before large orders (Dyhrberg, Foley, and Svec 2018) and thus to profit from these orders.

The Binance exchange platform allows two main types of order to be placed:

- A market order is one that will be executed immediately at the best available price in the order book. The buyer using a market order utilizes market liquidity; they are then called a “taker.” Depending on the requested volume and the depth of the available order book, the price at which the order will be executed may differ from the market price; this is called price “slippage.”
- A limit order is one that waits for the market price to reach the desired limit before being executed. The buyer using a limit order thus gives the market liquidity; they are then called a “maker”. Makers generally benefit from lower

[17] https://www.binance.com/en/support/announcement/detail/2e11b3d512fa4027b221b7b518fafe74

trading fees because they allow the exchange to function more efficiently by providing liquidity, but the execution speed of this type of order depends on market liquidity.

The order book allows traders to view limit orders. This allows them to identify potential imbalances between supply and demand and to identify short-term trends. It also allows them to calculate the size of orders to be placed to avoid creating excessive price slippage.

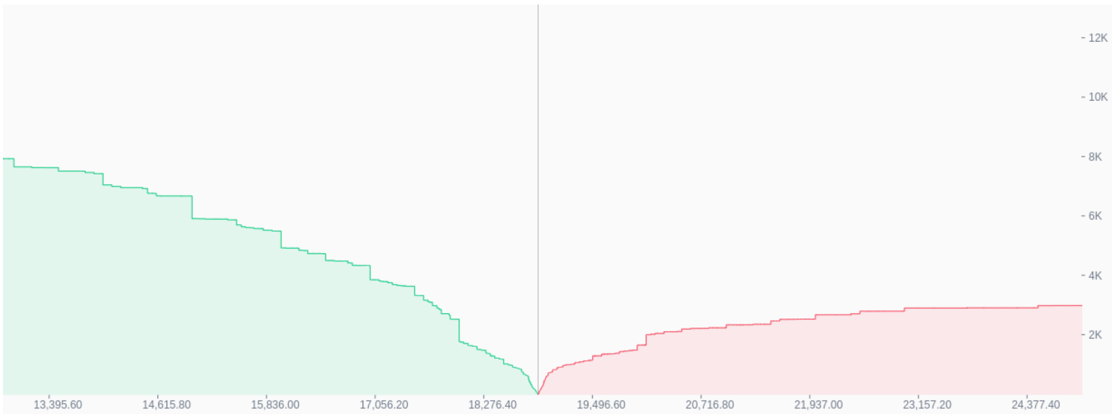


Figure 4: Example of an order book for the BTC/USDT pair (source: binance.com)

Wei (2018) highlights the role of liquidity in the cryptocurrency market. Liquidity helps to reduce volatility and improve market efficiency. This liquidity is dependent on current events. As Yue, S. Zhang, and Q. Zhang (2021) showed, liquidity tends to decrease after a negative news announcement and increase after a positive one. In 2023, for example, following legal action by the SEC, Binance saw part of its liquidity quickly disappear, resulting in price gaps of +3.5% compared to other platforms [18].

[18]source: https://shorturl.at/r74On

Despite their innovative nature, crypto markets share common fundamentals with traditional stock exchanges, while being subject to increased volatility.

2.3 Transaction costs

Cryptocurrency transaction costs vary depending on the blockchain network and exchange. Below is a detailed description of the different types of fees and their mechanisms:

- Network fees: these fees are paid to process a transaction on a specific blockchain. They compensate miners (Proof of Work) or validators (Proof of Stake) for their work securing and recording transactions.
- Exchange fees: platforms apply different types of commission for buying, selling, or withdrawing cryptocurrencies.
- Conversion fees: some platforms, such as Coinbase, allow direct exchanges between two cryptocurrencies in exchange for a platform commission.

Liquidity costs are required in addition to the fees charged by the network or platform. The concept of liquidity is important for financial markets; liquidity is defined as the ability to buy or sell assets on the market without causing a significant change in the asset's price and without waiting too long for the order to be executed. To assess whether a crypto market is liquid, three indicators are generally studied: 24-hour trading volume, order book depth, and bid-ask spread.

- The 24-hour trading volume is the total amount of an asset traded over a 24-hour period.
- The order book depth is the market's ability to absorb large buy and sell

orders. A depth indicator is generally calculated by adding the buy or ask volumes representing a certain percentage of the order book (e.g.: the ask volume at 10% represents the sum of all limit orders whose price is between the market price and the market price -10%).

- The bid-ask spread is the difference between the highest bid price and the lowest ask price offered in the order book. It is generally used as a liquidity indicator; the lower the spread, the greater the asset's liquidity. The bid-ask spread can also be considered as a cost for immediate order execution.

The liquidity provided by individual traders in the market is sometimes insufficient, which is why other means are sometimes used to provide liquidity. On centralized exchanges, market makers generally provide the liquidity. A market maker continuously provides both a "bid" (buy) price and an "ask" (sell) price for a financial asset. They buy at their bid price and sell at their ask price, creating a two-sided market profiting from the "bid-ask spread". By doing this, they ensure that buyers and sellers can always have a counterparty for their trades. On decentralized exchanges, order books are replaced by automated market makers (AMMs). These are smart contracts that use mathematical formulas, generally based on the quantities available in the liquidity pool (formulas vary depending on the AMM). With automated market makers, users buy and sell their tokens to smart contracts that act as intermediaries, eliminating the need for supply to match demand. The liquidity used by AMMs utilizes a fund called a liquidity pool. These liquidity pools are provided by users called liquidity providers (LPs). LPs receive a portion of trading fees in exchange for providing their liquidity. This protocol allows anyone with liquidity to become a market maker; this is one of the innovations of decentralized finance (DeFi). Barbon and Ranaldo (2024) compare these two market-making systems and show that on centralized platforms (Binance and Kraken), costs are generally lower than on decentralized platforms (Uniswap).

2.4 Volatility

ryptocurrencies have been noted for their sometimes extreme volatility. Volatility is a key characteristic of cryptocurrencies; it represents the intensity and frequency of changes in an asset's price over a given period. It can be interpreted as the uncertainty or risk associated with an investment. The volatility of cryptocurrencies is higher than in traditional markets (Baur and Dimpfl 2021). This significant volatility attracts investors seeking potential quick gains. Such speculative activities, fueled by fear of missing out (FOMO) and the search for high returns, can significantly amplify price fluctuations (R. J. Shiller 2005; Cheah and Fry 2015; Baur, K. Hong, and A. D. Lee 2018; R. J. Shiller 2019).

Cryptocurrency volatility can be attributed to several factors, including speculation (Cheah and Fry 2015; Baur, K. Hong, and A. D. Lee 2018), current events (Katsiampa 2019), and geopolitical factors (Aysan et al. 2019). The market is still young, and its regulations frequently change, generating uncertainty regarding its regulations and taxation. This uncertainty also contributes to volatility (Q. Wang et al. 2023). Furthermore, as the market is technologically driven, it is subject to cybersecurity risks. Hacks have a significant impact on market volatility, for example (W. Liu 2019; Corbet, Cumming, et al. 2020).

Bitcoin is often seen as a hedge against economic and financial uncertainty. In times of crisis, investors may turn to Bitcoin as an alternative to traditional assets, such as stocks and bonds (Demir et al. 2018). Research has found that cryptocurrencies are also more correlated with each other in times of uncertainty (Antonakakis, Chatziantoniou, and Gabauer 2019; Borri and Shakhnov 2020).

The cryptocurrency market exhibits varying degrees of long-term dependency (Bariviera 2021). Significant spillover effects are observed, indicating strong interconnectedness (Corbet, Meegan, et al. 2018) between cryptocurrencies. This strong interconnectedness is problematic in the event of bad news, as it propagates fast and affects the entire market (Shahzad et al. 2021).Market connectivity fluctuates depending on the level of uncertainty: it is high during periods of high uncertainty, but low during periods of low uncertainty (Antonakakis, Chatziantoniou, and Gabauer 2019). Research highlights a link with traditional assets as well as a risk of spillover of cryptocurrency volatility into non-digital assets (Milunovich 2018; Kurka 2019), although this risk is relatively low. Other research finds no cointegration between cryptocurrencies and traditional assets, instead highlighting the ability of cryptocurrencies to serve as a diversification tool (Corbet, Meegan, et al. 2018; Giudici and Abu-Hashish 2019; Gil-Alana, Abakah, and Rojo 2020).

2.5 Market efficiency

The theory of informational efficiency of markets (the "Efficient Market Hypothesis" or EMH) was first proposed by Fama (1965) as "a market in which prices fully and always reflect available information." He then introduced it in 1970 in an article entitled "Efficient Capital Markets: A Review of Theory and Empirical Work." It is one of the most influential theories of the last 50 years. According to Eugene Fama, there are three forms of informational efficiency:

- The strong form: all past information, both public and private, is instantly incorporated into an asset's price. All market participants would therefore have access to the same information and would interpret this information in the same way. Prices fully reflect the market's value. Technical analysis would therefore be useless.
- The semi-strong form: only past and public information is instantly incorporated into the asset's price. Some participants may have access to private information. Prices can therefore differ, making it possible to "beat the market" through information that others do not have (insider trading).
- The weak form: only past information is incorporated into prices. It would therefore be impossible for participants to make predictions about the asset's returns based on past information.

Although the cryptocurrency market is relatively new, many studies have focused on the informational efficiency of these new markets. For example, some researchers have shown that Bitcoin is an efficient market in the weak sense, but there is no consensus on this notion: Kristoufek (2018) and Tran and Leirvik (2020) show that the Bitcoin market has only been efficient since 2017, while other authors such as Garcia et al. (2014), Cheung, Roca, and and (2015), and Frunza and Guegan (2018)argue that Bitcoin has not always been an efficient market in this period.

The efficient market theory has been strongly challenged by behavioral finance, which considers that individuals constantly adapt to their environment. To adapt quickly, they use simple heuristics and therefore do not think “rationally”. Individuals are thus subject to behavioral bias that can explain the anomalies of the efficient market theory. The two theories are often contrasted. However, previous studies have shown that the Bitcoin market has sometimes been efficient and sometimes not.

Andrew Lo used the Adaptive Markets Hypothesis (Lo and R. Zhang 2024) to posit that market efficiency is not a static state but a dynamic process, where participants, influenced by both rational and behavioral biases, constantly adapt to the changing market conditions. Lo and Zhang’s approach is based on the principles of evolution (competition, adaptation, and natural selection). The authors explain that market efficiency depends on the number of participants in the market. Indeed, a small number of participants competing for an abundant resource will result in an inefficient market. However, a large number of competitors for a scarce resource will lead to a more efficient market. They believe that the risk/reward relationship in the market is dependent on several factors and that this relationship evolves over time, along with the factors on which it depends, arguing that there are arbitrage opportunities and that investment strategies can produce different results over time as the market evolves.

Several studies indicate that the market alternates between periods of efficiency and periods of inefficiency. Regarding the weak form of efficiency[19], Urquhart (2016) shows that the bitcoin market was inefficient between August 1, 2010 and July 31, 2013 and then efficient until July 31, 2016. Kristoufek (2018) claims that the Bitcoin market was mainly inefficient between 2010 and 2017. Bariviera (2017) also finds 2 periods, before and after 2014. Tran and Leirvik (2020) argue that the bitcoin market was only efficient from 2017 and that it was mainly inefficient before. The studies indicate that the bitcoin market efficiency depends on its maturity. Indeed, the market becomes more efficient over time (Köchling, Müller, and Posch 2019) yet inefficiencies remain (Hashemi Joo, Nishikawa, and Dandapani 2020; Krückeberg and Scholz 2020). Indeed, testing the semi-strong form of efficiency[20], Hashemi Joo, Nishikawa, and Dandapani (2020) used an event study to show that prices do not immediately reflect information. In fact, during significant events in the crypto ecosystem, the authors suggest that there is an opportunity for abnormal returns as the event can take up to 6 days to be reflected in prices. Additionally, Krückeberg and Scholz (2020) show that arbitrage opportunities increase over time for Bitcoin, but the authors do not consider platform fees or the timing needed to take advantage of potential arbitrage to exploit these price inefficiencies between platforms. Hattori and Ishida (2021) show that there are arbitrage opportunities during Bitcoin crashes between the Bitcoin market and its futures market. Grobys and Sapkota (2019) highlight the existence of a momentum effect (Jegadeesh and Titman 1993a) in the cryptocurrency market but

[19]Prices reflect all past information (historical prices, transaction volumes). It is impossible to make abnormal profits based on technical analysis alone.

[20]Prices reflect all publicly available information (financial reports, news, event announcements). It is impossible to make abnormal profits based on publicly available information.

suggest that the market grows more efficient over time.

Shifts between periods of efficiency and inefficiency, alongside the persistent presence of arbitrage opportunities and momentum effects, indicate that these markets are not yet fully mature. Additionally, the decentralized nature of cryptocurrencies raises significant concerns about investor safety and illicit activities. Consequently, a comprehensive and adaptive regulatory framework is essential to protect market participants from manipulation and fraud and to foster a safer, more transparent industry.

3 Regulation of cryptocurrencies

3.1 At global level

Regulation at global level is very disorganized. While some countries like Nepal, Iraq, and China have banned cryptocurrency trading and mining, others like Egypt have banned their banks and financial institutions from trading or only offering Bitcoin.

The collapse of the centralized exchange FTX in November 2022 highlighted the lack of cryptocurrency platform regulation. FTX, then the third-largest cryptocurrency exchange in the world, went bankrupt after revelations that customer funds were being commingled with the company's own funds via an affiliate company, Alameda Research,. These practices led to a massive loss of trust, leading to cascading withdrawals and a significant collapse of the crypto market. This event served as a catalyst to accelerate the implementation of stricter regulatory frameworks worldwide. In Europe, the MiCA (Markets in Crypto-Assets) regulation, adopted in 2024, was designed precisely to prevent such scandals.

3.2 At European level

On June 19, 2018, the EU's 5AMLD Directive (the 5th Anti-Money Laundering Directive) was published in the Official Journal of the European Union. Since January 10, 2020, all EU countries are required to transpose this directive into their domestic legal framework. This directive expands the scope of previous directives to cover service providers related to the exchange of virtual currencies and cryptocurrencies, as well as providers of electronic wallet services. The most significant change is the obligation to implement Know Your Customer (KYC) processes to prevent fraudulent transactions. This requires service providers to identify actual cryptocurrency holders and to verify their identity, thereby making transactions non-anonymous. They are also required to implement a transaction monitoring system that automatically detects transactions that pose a high risk of money laundering or terrorist financing. On September 24, 2020, the European Commission proposed the "Market in Crypto-assets" regulation, abbreviated to MiCa. Its aim is to harmonize national legislation, protect investors, regulate and support the development of tokens in Europe, and ensure financial stability. This regulation defines the concept of crypto-assets and crypto-asset service providers (CASPs). It regulates CASPs through new obligations and licenses, as well as defines penalties for noncompliance with these obligations.

This text aims to cover several areas:

- The supply and exchange of crypto-assets
- The supply and exchange of stablecoins

- Cryptoasset services
- The prevention of market abuse.

The text was adopted in December 30, 2024, resulting in only authorized crypto-asset service providers (CASPs) being able to operate. To obtain authorization, companies must comply with a number of obligations common to all digital asset services, such as having their headquarters in an EU member state or implementing a policy to prevent conflicts of interest. Depending on the service they offer, they must comply with additional obligations. For instance, for service providers ensuring custody of crypto-assets on behalf of their clients, the regulation governs the contractual agreement with the client (the parties' responsibilities, security, fees) and only authorizes subcontracting to service providers also authorized under the MiCa Regulation.

The MiCa Regulation also regulates ICOs (crypto fundraising):

- Obligation for issuers to publish a white paper describing the entire project: the characteristics of the offering (number of tokens issued, rights and responsibilities of each party, etc.), the main elements relating to the technology used, as well as the main risks of the project.
- Obligation to notify the authorities of their home country (e.g., in France: the Autorité des Marchés Financiers) and to send them the white paper before the ICO, as well as to notify them of any changes.
- Obligation to provide a right of withdrawal for investors when participating in an ICO.
- Obligation to ensure a high level of cybersecurity to safeguard investor funds.

- Penalties for non-compliance with the above obligations.

The MiCA proposal represents a step forward in terms of regulating digital finance players, but leaves a gap regarding decentralized finance (DeFi), which is a system that operates in a completely decentralized manner within the blockchain via smart contracts.

With MICA in force across the EU, the fight against money laundering and terrorist financing is also strengthened via the Transfer of Funds Regulation (TFR), which regulates digital asset transfers. For each transfer between two regulated platforms, the client's originating platform must transmit KYC information, and the platform receiving the funds must, in return, provide the beneficiary's KYC information. This information must be stored so that it can be consulted by the authorities.

3.3 At French level

France took action to regulate crypto-asset markets ahead of European regulations, with two main goals: to protect savers and to combat crime while boosting innovation. Transactions carried out by individuals in a non-professional context are governed by Article 150 VH bis of the General Tax Code, which covers capital gains from digital assets. From a tax perspective, capital gains are exempt if the annual amount of sales does not exceed €305[21]. They are also exempt when the digital asset held is exchanged for another digital asset without a cash payment (this is the case when currencies are traded on an exchange platform such as Binance, in particular). Apart from these exemptions, the single flat-rate deduction

[21] source: https://www.economie.gouv.fr/cedef/regime-fiscal-cryptomonnaies

of 12.8% applies, to which are added social security contributions of 17.2%.

For professionals, the law of December 9, 2016 on transparency, the fight against corruption, and the modernization of economic life (the “Sapin II” Law) meant that firms offering bitcoin intermediation services were considered as intermediaries in various goods. As such, they had to be registered with the AMF. Since May 22, 2019, the PACTE (“Plan d’Action pour la Croissance et la Transformation des Entreprises”[22]) law has established a new regulatory framework and a status has been created for digital asset service providers (PSAN)[23].A service provider is considered a PSAN if it provides at least one of the following digital asset services.

Registration is mandatory for these two services:

- Custody of cryptocurrency on behalf of third parties.
- Purchase/Sale/Exchange of cryptocurrency for real currency or other cryptocurrencies;

Registration is optional for the following digital asset services:

- Operation of an exchange;
- Reception and transmission of orders on behalf of third parties;
- Portfolio management on behalf of third parties;
- Digital asset advice;
- Guaranteed or unguaranteed investment of digital assets.

[22]meaning ”Action Plan for Business Growth and Transformation” in english.
[23]source: https://www.amf-france.org/fr/actualites-publications/actualites/actifs-numeriques-lamf-modifie-son-reglement-general-et-met-jour-sa-doctrine-sur-les-psan

The AMF verifies the service provider's compliance with anti-money laundering and countering the financing of terrorism (AML/CFT) regulations. It seeks the opinion of the Prudential Supervision and Resolution Authority (ACPR).

Once authorized, PSANs must comply with certain requirements, including professional liability insurance and minimum capital requirements. They must have at least one effective manager, sufficient human resources, a resilient IT system, internal control mechanisms, and complaints handling procedures. They must guarantee to limit conflicts of interest and have procedures in place to combat money laundering and terrorist financing. From a tax perspective, sales made for professional purposes fall under the BIC framework and are subject to income tax.

Regulating cryptocurrencies is a complex issue; governments and institutions must find solutions to protect users from fraud, scams, and hacks. They must also combat money laundering and terrorist financing and find ways to tax these activities in order to ensure this new sector contributes to the national economy. All of this must be done while striking a balance between protecting investors from risks and preserving innovation.

4 Risks associated with cryptocurrencies

4.1 Risks related to regulation

As we saw earlier, the cryptocurrency market is still young and its regulations are still under construction. Regulations not only reassure investors but can also slow down innovation. It is therefore not surprising that, when a new announcement is made regarding regulations, investors reevaluate their positions and price movements appear. Indeed, Y. Liu and Tsyvinski (2021) indicate that cryptocurrency returns are lower on days when there are regulatory announcements.

The authors also indicate that prices only react to negative events such as bans or restrictions on use, and not to positive ones such as progress in the establishment of the regulatory framework or the legal recognition of cryptocurrencies as a specific asset.

Shanaev et al. (2020) notes the importance of taking events concerning regulation into account to anticipate returns. The authors show that the cryptocurrency market reacts negatively when regulations become more oppressive (implementation of anti-money laundering policies, closing a platform, opening an investigation, banning an ICO). Conversely, the market reacts positively when regulations support cryptocurrency development and the “let it be” (e.g., approval of a new exchange platform).

4.2 Risks related to the technological aspect

Cryptocurrencies theoretically eliminate the need for “trusted third parties” as the blockchain is secure and decentralized. The notion of “trusted third party” encompasses any person or institution that, due to their status, creates a trusted environment to ensure a transaction. This vision is achievable provided that users exchange and store their cryptocurrencies themselves in order to respect the principle of decentralization. However, in practice, for reasons of simplicity, users generally store their cryptocurrencies in wallet systems managed by private companies (e.g. Ledger or Trezor). Cryptocurrency trading is also largely conducted via centralized exchanges managed by private companies. Clients of these exchanges deposit their cryptocurrencies into the exchange’s wallets. Consequently, intermediaries currently exist to manage their users’ cryptocurrencies, and these intermediaries are vulnerable to hacks and mismanagement.

Among the security concerns regularly reported in the press, hacking is a significant risk in the cryptocurrency sector due to the direct losses it causes and its repercussions on the market. Indeed, Corbet, Cumming, et al. (2020) show that hacks and fake ICOs are generally followed by a rise in volatility of the targeted cryptocurrency and an increase in general correlation between cryptocurrencies. W. Liu (2019) showed that investor attention to this topic (via Google searches for the keywords “Bitcoin Hack”) is negatively correlated with returns up to 6 weeks after the event. This aspect specific to cryptocurrencies therefore needs to be taken into account by investors.

However, it’s not the blockchain that is usually hacked. Indeed, scandals regularly occur regarding hacks of exchange platforms and wallets that store users' private keys. As yet, it is not the blockchain technology itself that has been hacked, but rather the platforms that offer services around the technology. However, the general public is generally unable to distinguish between the two, wrongly associating hacks with cryptocurrencies. The most infamous scandal was the closure of MtGox in 2014 [24]. A victim of hacking, the platform lost 744,408 BTC, the equivalent of over €250 million at the time. More recently, the TerraUSD stablecoin was hacked in 2022 [25] subsequently losing 87% of its value in 4 days, dragging the cryptocurrency LUNA down with it, which fell 99.9% on May 13, 2022. Or the hack of FTX[26] which suffered a loss equivalent to half a billion Euros 24 hours after the company was declared bankrupt.

Another major security risk for cryptocurrency holders is the risk of losing their private key. When cryptocurrency holders do not use an intermediary, they must store their private key on a personalized storage medium (hard drive, paper). As with cash in their wallet, if they lose it, the user is solely responsible for the loss. Once the money is withdrawn from the bank, the bank is no longer responsible for the loss. It is the same for Bitcoin. If a user loses their private key, their Bitcoins are no longer usable, and the only way to unlock them is to find the storage medium where they saved their private key.

As with any digital information, there is also the risk of phishing; hackers try to recover users' private keys by sending them a fake email from a website they use.

[24]https://www.lesechos.fr/2014/02/mais-ou-sont-passes-les-850000-bitcoins-geres-par-mtgox-292828

[25]https://www.cointribune.com/terrausd-ust-le-modus-operandi-de-lauteur-du-hack-de-800-millions-de-dollars-explique/

[26]https://www.lemonde.fr/pixels/article/2022/11/14/cryptomonnaies-ftx-confirme-avoir-ete-victime-d-un-piratage-apres-avoir-fait-faillite_6149793_4408996.html

This email then redirects them to the hacker's website (which copies the site they usually use) where they are asked to enter their private key. The hackers then save it and use it fraudulently.

Academic research has shown that these events have an impact on cryptocurrency prices. Indeed, Y. Liu and Tsyvinski 2021 show that Google searches for the keywords 'Bitcoin Hack' are negatively correlated with returns up to 6 weeks after the event. Corbet, Cumming, et al. 2020 note that hacks and fake ICOs are generally followed by an increase in the volatility of the targeted cryptocurrency and a rise in general correlation between cryptocurrencies.

4.3 The risks of scams and bankruptcies

With the growing adoption of cryptocurrencies, scams are on the rise. Indeed, many scams involve recovering the victim's private key in order to extract a cryptocurrency transaction. Among them, many online sites offer the use of their trading robots, exploiting investors' greed by promising often surreal returns. These sites are actually Ponzi schemes. The most well-known is undoubtedly the BitConnect case, which allegedly managed to extract up to $24 million from its users. These scams capitalize on investors' lack of knowledge and their greed.

Another common type of scam is the NFT scam. This involves selling a project with NFTs and then, once the money is raised, delivering users something that does not correspond to what was sold, or even disappearing without a trace. Among this type of scam, we can cite the famous "Evolved Apes"[27] or the "Onecoin"

[27]source: https://shorturl.at/EhYEq

project[28].

In addition to scams, crypto projects can also go bankrupt. Most bankruptcies to date have had a localized effect on the crypto ecosystem. In May 2022, the collapse of the Terra cryptocurrency resulted in the loss of $50 billion. This bankruptcy had very little effect on traditional finance, although it did cause the demise of major crypto players, including the bankruptcy of Celsius [29] and 3 Arrows Capital [30]. The crisis in the crypto ecosystem that followed the FTX bankruptcy was so significant that it had repercussions on traditional finance. Indeed, in November 2022, FTX, the second largest centralized exchange in the world, declared bankruptcy due to mismanagement. The FTX bankruptcy was the first case of contagion of a crisis originating from the cryptocurrency sector to the traditional finance sector. The bankruptcy led to a major crisis in the crypto ecosystem, resulting in numerous bankruptcies which led to repercussions in the world of traditional finance in 2023 via the bankruptcy of Silvergate Bank which had lent a lot of money to crypto projects.

4.4 Cryptocurrencies and the environment

Bitcoin and many other cryptocurrencies consume a significant amount of energy due to their use of the Proof of Work protocol. The University of Cambridge provides an index to track the amount of energy consumed by Bitcoin (the Bitcoin Cambridge Electricity Consumption Index[31]. The university reports that Bitcoin's electricity consumption in August 2023 was 129 TWh, comparable to the electric-

[28]source: https://shorturl.at/nnQSa
[29]source: https://shorturl.at/G4cZH
[30]source: https://shorturl.at/i5VJb
[31]https://ccaf.io/cbnsi/cbeci

ity consumption of Sweden, Norway, the United Arab Emirates, Argentina, and the Netherlands. Cryptocurrency miners are banned in some countries, such as China, and numerous press articles denounce the excessive use of electricity and computer equipment[32]. While blockchain applications may have a social and economic interest, some researchers like Truby (2018) denounce the environmental cost of the technology and propose regulatory solutions to enable its development in a more environmentally friendly manner.

Some blockchain players are not waiting for regulations to move forward on the issue, with some miners migrating to places where electricity is cleaner, particularly in Iceland, but they are a very small minority. Others certify that they use surplus energy, but do not indicate the proportion of renewable energy production they use.

Some cryptocurrencies opted for a less energy-intensive consensus protocol and have decided to adapt and change their consensus mechanism. This is the case for Ethereum, which changed its protocol to adopt Proof of Stake, or Tezos (and many others) which have used this system since their inception.

The electricity consumption of cryptocurrencies is therefore diverse, as is the means of producing this electricity. Although the sector's impact on global warming is relatively low compared to other sectors such as industry, this factor could be taken into account by regulators by prohibiting the mining of 'proof of work'-type cryptocurrencies, for example.

The varying environmental impact of cryptocurrencies and the new risks involved compared to traditional assets adds additional complexity to their evaluation. The

[32]source: https://shorturl.at/b9ANU

discussion naturally leads to a broader question: how do we define the intrinsic value of these digital assets? Unlike traditional financial instruments with established valuation models, cryptocurrencies challenge conventional approaches, making it essential to understand what truly influences their price.

5 Lack of consensus on core value

Since cryptocurrencies are digital assets, a legitimate question is how their fundamental value should be defined. Traditional finance is based on the assumption that markets are efficient and investors are rational. Within this framework, an asset's fundamental value is considered its intrinsic value, determined by objective factors such as future cash flows, profits, and interest rates (Gordon 1959; Campbell and R. Shiller 1988). Defining this fundamental value is difficult. Indeed, when we have measurable criteria, their consideration can vary from one analyst to another. Other characteristics, such as reputation or positive impact on society, are subjective and more complex to measure. Furthermore, the valuation criteria will vary depending on the type of asset. For example, different criteria are used to evaluate a railway company and a technology company. In fact, a train company valuation relies heavily on tangible assets like infrastructure, along with stable, predictable cash flows from long-term contracts and regulated tariffs. In contrast, valuing a technology company emphasizes future growth potential, user base, and scalability, often with less focus on immediate profitability or physical assets. It is therefore important to understand what influences an asset's price in order to evaluate it.

While traditional financial assets often derive their value from established metrics, the unique nature of cryptocurrencies requires a different approach to understand their price determinants. Academics have explored various factors beyond conventional financial analysis. This section delves into the various influences on cryptocurrency prices, moving from the foundational aspects of their production to the psychological and behavioral elements that drive market dynamics.

5.1 Production factors

Unlike gold, real estate, or stocks, cryptocurrencies are not physical assets and do not have accounting systems to estimate their value. However, studying their production factors reveals a common facet found across many cryptocurrencies. While each cryptocurrency has its own unique functionality and purpose, several common variables can generally be extracted. Thus, many studies have used variables extracted from blockchains using a Proof-of-Work mechanism. This data is easily accessible for Bitcoin. As the leading cryptocurrency in terms of market capitalization, the majority of studies using these variables have data from the Bitcoin network. Miners play a vital role in the Bitcoin ecosystem; they are the ones who verify transactions and create new bitcoins on the blockchain. Without miners, Bitcoin could not function as there would be no one to validate transactions and secure the network. The miners allocate their computing power to validate the transactions in progress on the network and thus secure the blockchain in the hope of obtaining bitcoins in exchange. They compete to solve mathematical problems; the first to find the solution is the one who can write the pending transactions into the blockchain. The person in question will then be compensated via newly issued bitcoins and will also receive the fees attached to the various transactions that he or she has validated. Around 2040, all the bitcoins will have been issued, so there is uncertainty about whether Bitcoin will still be able to attract miners after this date. Indeed, Easley, O'Hara, and Basu (2019) used a theoretical model to show that the relationship between miners and users is a complex equilibrium. Indeed, if transaction fees become too high, or if there is too much waiting time to validate a transaction, then users leave the network. Conversely, if transaction fees are not high enough, miners also leave the network. Pagnotta (2021) used another theoretical model to identify a feedback mechanism between the computing power offered by miners and the price. The more computing power increases, the more secure and efficient the blockchain becomes, and the more the price goes up. The increase in price and security attracts new users, who also increase the price, which in turn attracts new miners. This link between the increase in price and the rise in the number of miners is also confirmed by Kristoufek (2015), Hayes (2017), and Bhambhwani, Delikouras, and Korniotis (2021). Although these factors are significantly correlated with the long-term increase in the price of bitcoin, they do not explain the short-term variations (Y. Liu and Tsyvinski 2021) and one might wonder if they are not simply the consequence of the massive adoption of cryptocurrencies in recent years.

5.2 The network effect and speculation

In recent years, the mass adoption of Bitcoin has been a major factor in explaining its price (Kristoufek 2015; Hayes 2017; Bhambhwani, Delikouras, and Korniotis 2021). Its adoption can be measured by the increase in the number of miners or in the number of wallets (i.e. addresses) on the network. The metrics are significant factors in modeling the price of Bitcoin, yet, as we saw earlier, Bitcoin is neither a store of value nor a means of payment. So why its adoption? Research on Bitcoin has focused on its speculative aspect. Baur, K. Hong, and A. D. Lee (2018) argue that Bitcoin is not used as a medium of exchange or currency but is primarily used as a speculative vehicle. Cheah and Fry (2015) argue that the price of Bitcoin depends mainly on speculation and that its fundamental value is $0.

Bitcoin price movements would thus be linked to its adoption by users (higher demand). Y. Liu and Tsyvinski (2021) analyze the effect of cryptocurrency adoption by users, taking the increase in several variables into account: the number of wallets, the number of active addresses, the number of transactions, and the number of payments, showing that these variables are important drivers of cryptocurrency prices. Kristoufek 2015 argues that the growth in Bitcoin use for real transactions is linked to an increase in price. Aalborg, Molnár, and Vries (2019) show that the number of unique addresses in the Bitcoin network is positively correlated with Bitcoin returns on both a daily and a weekly horizon. These studies are consistent with the results of Ciaian, Rajcaniova, and Kancs (2016) which already showed in 2016 that demand (measured by the number of addresses and the number of transactions on the network) was positively correlated to the price of Bitcoin. Bhambhwani, Delikouras, and Korniotis 2021 note that the more mature the cryptocurrency, the more the network size (measured by the number of users on the blockchain) is a significant factor in modeling the price.

5.3 Behavioral finance

Research has shown that when an asset is difficult to value, it tends to be more sensitive to behavioral variables (Kumar 2009). It thus seems interesting to turn to behavioral finance to explain these price movements.

Behavioral finance challenges the assumption of investor rationality. It recognizes that emotions, cognitive biases, and psychological factors can influence investment decisions and lead to deviations between market price and fundamental value (R. Shiller 1981; Baker and Wurgler 2006).

As with traditional financial markets, anomalies exist in the efficiency of cryptocurrency markets. Indeed, Y. Liu and Tsyvinski 2021 show that the momentum effect measured by cumulative returns over one week can predict Bitcoin's returns up to 4 weeks in advance. They also show a reversal effect from 8 weeks.

Tzouvanas, Kizys, and Tsend-Ayush 2020 demonstrate that a J/K Momentum strategy (Jegadeesh and Titman 1993b) based on a 7-day window allows for abnormal returns. They also show that this effect depends on investor attention: cryptocurrencies with less user attention are more likely to be subject to the momentum effect than cryptocurrencies with more attention. This relationship between investor attention and momentum is consistent with research on the stock market (H. Hong, Lim, and Stein 2000).

The well-known work by Barber and Odean (2008) on the stock market explains that due to their limited attention span, investors primarily invest in stocks that initially catch their attention. Thus, an increase in attention to a stock generally predicts an increase in the price of said stock. In the cryptocurrency market, the impact of investor attention on prices has been measured by numerous proxies. A popular measure in the literature is a Google search intensity. Kristoufek (2013) shows that an increase in Google search intensity is associated with an increase in an upward or downward trend. Bouoiyour and Selmi (2015) and Philippas et al. (2019) show that Google searches have a positive and significant impact on Bitcoin returns. Urquhart 2018 suggests that high volatility and/or unusually high volume attracts investor attention, while Nasir et al. (2019) show that a Google search shock has a positive effect on Bitcoin returns for a week. Y. Liu and Tsyvinski (2021) also use Google searches as a proxy for investor attention, with the difference between searches this week compared to the previous four weeks used as a measure. They argue that this measure can predict cumulative future returns over a one- to six-week horizon.

n 2025, information research is also largely done via social networks (Mai, Bai, and Shan 2015; Philippas et al. 2019). Using the bitcointalk.org forum to reflect investor attention, Ciaian, Rajcaniova, and Kancs (2016) note that the number of posts is positively correlated with the price of Bitcoin. Mai, Bai, and Shan (2015) show that not all message origins have the same impact. Messages posted by the silent minority seem to have a greater effect on the price of Bitcoin, and messages on internet forums seem to have a more significant effect at daily frequency, while Twitter messages seem to have a more significant effect at hourly level. This is particularly true when tweets come from prominent figures (Ante 2023).

The role of investor sentiment and emotions on asset prices is also widely recognized (R. J. Shiller 2005; Kumar and C. M. Lee 2006; Taffler 2018). Notably, Taffler (2018) extensively explores the impact of emotions on financial decision-making, arguing that cognitive and emotional biases are not deviations from rationality but are inherent to human judgment and can significantly influence market behavior and asset pricing. He suggests that investment evokes a powerful emotional conflict between the excitement of potential gains and the anxiety of potential losses. Various indicators have been used to measure investor sentiment empirically: proxy variables can be used (internet searches, technical indicators, put/call ratios) as in Baker and Wurgler (2006). A popular method since Tetlock (2007) is to use an algorithm to analyze textual elements such as news articles, social media posts, or even messages on Internet forums to infer user sentiment. Thus, research has shown that Bitcoin is particularly sensitive to the sentiment conveyed in the news (Sapkota 2022) and on social networks (Smuts 2019; Guégan and Renault 2021; López-Cabarcos et al. 2021).

Thus, having established the influence of attention, sentiment, and emotions on the Bitcoin market, we can explore in more detail how these factors are manifested through modern information channels such as YouTube or Reddit. The following chapters examine Bitcoin market dynamics. More specifically, we focus on how these factors contribute to its returns, volatility, and the formation of bull run periods. Chapter 2 uses YouTube data on Bitcoin returns to investigates the impact of investor sentiment and attention. It examines how the sentiment expressed in different categories of videos influences price movements. Chapter 3 explores the relationship between investor sentiment in Reddit discussions and Bitcoin volatility. The chapter uses natural language processing and artificial intelligence to differentiate the impact of sentiment from various investor groups (classified by their Blockchain competency) on Bitcoin volatility. Finally, in Chapter 4, we develop a new model that integrates behavioral and emotional factors like fear, joy, and information appetite to more accurately identify Bitcoin bull runs. This sequential exploration highlights the central role of human behavior, sentiments, and emotions in shaping the Bitcoin market.

Bibliography

Aalborg, H. A., P. Molnár, and J. E. de Vries (2019). "What can explain the price volatility and trading volume of Bitcoin?" *Finance Research Letters* 29, pp. 255–265.

Amihud, Y. (2002). "Illiquidity and stock returns: cross-section and time-series effects". *Journal of Financial Markets* 5.1, pp. 31–56.

Amihud, Y. and H. Mendelson (1986). "Asset pricing and the bid-ask spread". *Journal of Financial Economics* 17.2, pp. 223–249.

Ante, L. (2023). "How Elon Musk's Twitter activity moves cryptocurrency markets". *Technological Forecasting and Social Change* 186, p. 122112.

Antonakakis, N., I. Chatziantoniou, and D. Gabauer (2019). "Cryptocurrency market contagion: Market uncertainty, market complexity, and dynamic portfolios". *Journal of International Financial Markets, Institutions and Money* 61, pp. 37–51.

Antonopoulos, A. M. (2017). *Mastering Bitcoin: programming the open blockchain.* Second edition. Sebastopol, CA: O'Reilly. 371 pp.

Auer, R. and D. Tercero-Lucas (2022). "Distrust or speculation? The socioeconomic drivers of U.S. cryptocurrency investments". *Journal of Financial Stability* 62, p. 101066.

Aysan, A. F., E. Demir, G. Gozgor, and C. K. M. Lau (2019). "Effects of the geopolitical risks on Bitcoin returns and volatility". *Research in International Business and Finance* 47.C, pp. 511–518.

Baker, M. and J. Wurgler (2006). "Investor Sentiment and the Cross-section of Stock Returns". *The Journal of Finance* 61.4, pp. 1645–1680.

Barber, B. and T. Odean (2008). “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors”. *Review of Financial Studies* 21.2, pp. 785–818.

Barbon, A. and A. Ranaldo (2024). *On The Quality Of Cryptocurrency Markets: Centralized Versus Decentralized Exchanges.*

Bariviera, A. F. (2017). “The inefficiency of Bitcoin revisited: A dynamic approach”. *Economics Letters* 161, pp. 1–4.

Bariviera, A. F. (2021). “One model is not enough: Heterogeneity in cryptocurrencies’ multifractal profiles”. *Finance Research Letters* 39, p. 101649.

Baur, D. G. and T. Dimpfl (2021). “The volatility of Bitcoin and its role as a medium of exchange and a store of value”. *Empirical Economics* 61.5, pp. 2663–2683.

Baur, D. G., K. Hong, and A. D. Lee (2018). “Bitcoin: Medium of exchange or speculative assets?” *Journal of International Financial Markets, Institutions and Money* 54, pp. 177–189.

Bhambhwani, S. M., S. Delikouras, and G. M. Korniotis (2021). “Blockchain Characteristics and the Cross-Section of Cryptocurrency Returns”, p. 53.

Bianchi, D. and M. Babiak (2022). “On the performance of cryptocurrency funds”. *Journal of Banking & Finance* 138, p. 106467.

Borri, N. and K. Shakhnov (2020). “Regulation spillovers across cryptocurrency markets”. *Finance Research Letters* 36, p. 101333.

Bouoiyour, J. and R. Selmi (2015). “What Does Bitcoin Look Like?” *Annals of Economics and Finance* 16.2, pp. 449–492.

Bourghelle, D. and F. Declerck (2004). “Why markets should not necessarily reduce the tick size”. *Journal of Banking & Finance* 28.2, pp. 373–398.

Campbell, J. and R. Shiller (1988). *Stock Prices, Earnings and Expected Dividends.* NBER Working Papers 2511. National Bureau of Economic Research, Inc.

Cheah, E.-T. and J. Fry (2015). "Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin". *Economics Letters* 130, pp. 32–36.

Cheung, A. (.-K., E. Roca, and J.-J. S. and (2015). "Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices". *Applied Economics* 47.23, pp. 2348–2358.

Chohan, U. W. and S. v. Kerckhoven, eds. (2023). *Activist retail investors and the future of financial markets: understanding YOLO capitalism.* Routledge international studies in money and banking. Abingdon, Oxon ; New York, NY: Routledge.

Choudhary, S., R. Bondia, V. Srivastava, and P. Chandra Biswal (2024). "Uncovering the Bitcoin investment behavior: An emerging market study". *Investment Management and Financial Innovations* 21.4, pp. 35–48.

Ciaian, P., A. Cupak, P. Fessler, and d. Kancs (2022). *Environmental-Social-Governance Preferences and Investments in Crypto-Assets.*

Ciaian, P., M. Rajcaniova, and d. Kancs (2016). "The economics of BitCoin price formation". *Applied Economics* 48.19, pp. 1799–1815.

Corbet, S., D. J. Cumming, B. M. Lucey, M. Peat, and S. A. Vigne (2020). "The destabilising effects of cryptocurrency cybercriminality". *Economics Letters* 191, p. 108741.

Corbet, S., A. Meegan, C. Larkin, B. Lucey, and L. Yarovaya (2018). "Exploring the dynamic relationships between cryptocurrencies and other financial assets". *Economics Letters* 165, pp. 28–34.

De Filippi, P. (2022). *Blockchain et cryptomonnaies.* Deuxième édition mise à jour. Paris: Que sais-je?

Demir, E., G. Gozgor, C. K. M. Lau, and S. A. Vigne (2018). "Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation". *Finance Research Letters* 26, pp. 145–149.

Dyhrberg, A. H., S. Foley, and J. Svec (2018). "The Impact of Tick Sizes on Trader Behavior: Evidence from Cryptocurrency Exchanges". *SSRN Electronic Journal.*

Easley, D. and M. O'Hara (1987). "Price, trade size, and information in securities markets". *Journal of Financial Economics* 19.1, pp. 69–90.

Easley, D., M. O'Hara, and S. Basu (2019). "From mining to markets: The evolution of bitcoin transaction fees". *Journal of Financial Economics* 134.1, pp. 91–109.

Fama, E. F. (1965). "The Behavior of Stock-Market Prices". *The Journal of Business* 38.1, p. 34.

Frunza, M. and D. Guegan (2018). "Is the Bitcoin Rush Over?" *SSRN Electronic Journal.*

Garcia, D., C. J. Tessone, P. Mavrodiev, and N. Perony (2014). "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy". *Journal of The Royal Society Interface* 11.99, p. 20140623.

Gil-Alana, L. A., E. J. A. Abakah, and M. F. R. Rojo (2020). "Cryptocurrencies and stock market indices. Are they related?" *Research in International Business and Finance* 51, p. 101063.

Giudici, P. and I. Abu-Hashish (2019). "What determines bitcoin exchange prices? A network VAR approach". *Finance Research Letters* 28, pp. 309–318.

Gordon, M. J. (1959). “Dividends, Earnings, and Stock Prices”. *The Review of Economics and Statistics* 41.2, p. 99.

Grobys, K. and N. Sapkota (2019). “Cryptocurrencies and momentum”. *Economics Letters* 180, pp. 6–10.

Grossman, S. J. and M. H. Miller (1988). “Liquidity and Market Structure”. *The Journal of Finance* 43.3, pp. 617–633.

Guégan, D. and T. Renault (2021). “Does investor sentiment on social media provide robust information for Bitcoin returns predictability?” *Finance Research Letters* 38, p. 101494.

Hackethal, A., T. Hanspal, D. M. Lammer, and K. Rink (2022). “The Characteristics and Portfolio Behavior of Bitcoin Investors: Evidence from Indirect Cryptocurrency Investments”. *Review of Finance* 26.4, pp. 855–898.

Harris, L. E. (1994). “Minimum Price Variations, Discrete Bid–Ask Spreads, and Quotation Sizes”. *The Review of Financial Studies* 7.1, pp. 149–178.

Hashemi Joo, M., Y. Nishikawa, and K. Dandapani (2020). “Announcement effects in the cryptocurrency market”. *Applied Economics* 52.44, pp. 4794–4808.

Hattori, T. and R. Ishida (2021). “The relationship between arbitrage in futures and spot markets and Bitcoin price movements: Evidence from the Bitcoin markets”. *Journal of Futures Markets* 41.1, pp. 105–114.

Hayes, A. S. (2017). “Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin”. *Telematics and Informatics* 34.7, pp. 1308–1321.

Hong, H., T. Lim, and J. Stein (2000). “Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies”. *Journal of Finance* 55.1, pp. 265–295.

Huang, X., J. Lin, and P. Wang (2022). “Are institutional investors marching into the crypto market?” *Economics Letters* 220, p. 110856.

Jegadeesh, N. and S. Titman (1993a). “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”. *The Journal of Finance* 48.1, pp. 65–91.

Jegadeesh, N. and S. Titman (1993b). “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”. *The Journal of Finance* 48.1, pp. 65–91.

Katsiampa, P. (2019). “Volatility co-movement between Bitcoin and Ether”. *Finance Research Letters* 30, pp. 221–227.

Kim, J. H. (2022). “Analyzing diversification benefits of cryptocurrencies through backfill simulation”. *Finance Research Letters* 50, p. 103238.

Kim, K. T. and L. Fan (2025). “Beyond the hashtags: social media usage and cryptocurrency investment”. *International Journal of Bank Marketing* 43.3, pp. 569–590.

Köchling, G., J. Müller, and P. N. Posch (2019). “Price delay and market frictions in cryptocurrency markets”. *Economics Letters* 174, pp. 39–41.

Kristoufek, L. (2013). “BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era”. *Scientific Reports* 3.1, p. 3415.

Kristoufek, L. (2015). “What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis”. *PLOS ONE* 10.4. Ed. by E. Scalas, e0123923.

Kristoufek, L. (2018). “On Bitcoin markets (in)efficiency and its evolution”. *Physica A: Statistical Mechanics and its Applications* 503, pp. 257–262.

Krückeberg, S. and P. Scholz (2020). “Decentralized Efficiency? Arbitrage in Bitcoin Markets”. *Financial Analysts Journal* 76.3, pp. 135–152.

Kumar, A. (2009). “Hard-to-Value Stocks, Behavioral Biases, and Informed Trading”. *The Journal of Financial and Quantitative Analysis* 44.6, pp. 1375–1401.

Kumar, A. and C. M. Lee (2006). “Retail Investor Sentiment and Return Comovements”. *The Journal of Finance* 61.5, pp. 2451–2486.

Kurka, J. (2019). “Do cryptocurrencies and traditional asset classes influence each other?” *Finance Research Letters* 31, pp. 38–46.

Kyle, A. S. (1985). “Continuous Auctions and Insider Trading”. *Econometrica* 53.6, p. 1315.

Liu, W. (2019). “Portfolio diversification across cryptocurrencies”. *Finance Research Letters* 29, pp. 200–205.

Liu, Y. and A. Tsyvinski (2021). “Risks and Returns of Cryptocurrency”. *The Review of Financial Studies* 34.6, pp. 2689–2727.

Lo, A. W. and R. Zhang (2024). *The adaptive markets hypothesis: an evolutionary approach to understanding financial system dynamics*. Clarendon lectures in finance. Oxford: Oxford University Press. 1 p.

López-Cabarcos, M. Á., A. M. Pérez-Pico, J. Piñeiro-Chousa, and A. Šević (2021). “Bitcoin volatility, stock market and investor sentiment. Are they connected?” *Finance Research Letters* 38, p. 101399.

Mai, F., Q. Bai, and J. Shan (2015). “The Impacts of Social Media on Bitcoin Performance”. *Thirty Sixth International Conference on Information Systems, Fort Worth*, p. 16.

Milunovich, G. (2018). “Cryptocurrencies, Mainstream Asset Classes and Risk Factors: A Study of Connectedness”. *Australian Economic Review* 51.4, pp. 551–563.

Nakamoto, S. (2008). “Bitcoin: A Peer-to-Peer Electronic Cash System”, p. 9.

Nasir, M. A., T. L. D. Huynh, S. P. Nguyen, and D. Duong (2019). “Forecasting cryptocurrency returns and volume using search engines”. *Financial Innovation* 5.1, p. 2.

Pagnotta, E. S. (2021). “Decentralizing Money: Bitcoin Prices and Blockchain Security”. *The Review of Financial Studies*. Ed. by I. Goldstein, hhaa149.

Pástor, Ľ. and R. F. Stambaugh (2003). “Liquidity Risk and Expected Stock Returns”. *Journal of Political Economy* 111.3, pp. 642–685.

Philippas, D., H. Rjiba, K. Guesmi, and S. Goutte (2019). “Media Attention and Bitcoin Prices”. *Finance Research Letters* 30, pp. 37–43.

Porter, D. C. and D. G. Weaver (1997). “Tick Size and Market Quality”. *Financial Management* 26.4.

Roll, R. (1984). “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market”. *The Journal of Finance* 39.4, pp. 1127–1139.

Sapkota, N. (2022). “News-based sentiment and bitcoin volatility”. *International Review of Financial Analysis* 82, p. 102183.

Shahzad, S. J. H., E. Bouri, S. H. Kang, and T. Saeed (2021). “Regime specific spillover across cryptocurrencies and the role of COVID-19”. *Financial Innovation* 7.1, p. 5.

Shanaev, S., S. Sharma, B. Ghimire, and A. Shuraeva (2020). “Taming the blockchain beast? Regulatory implications for the cryptocurrency Market”. *Research in International Business and Finance* 51, p. 101080.

Shiller, R. (1981). “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71.3, pp. 421–36.

Shiller, R. J. (2005). *Irrational exuberance*. 2nd ed. Currency/Doubleday. 304 pp.

Shiller, R. J. (2019). *Narrative economics: how stories go viral & drive major economic events*. Princeton: Princeton University Press. 377 pp.

Smuts, N. (2019). “What Drives Cryptocurrency Prices?: An Investigation of Google Trends and Telegram Sentiment”. *ACM SIGMETRICS Performance Evaluation Review* 46.3, pp. 131–134.

Srivastava, R., D. K. Singh, and N. P. Rana (2024). “Analysis of barriers to investment and mining in cryptocurrency for traditional and tech-savvy investors: A fuzzy approach”. *Technology in Society* 77, p. 102546.

Stoll, H. (1978). “The Supply of Dealer Services in Securities Markets”. *Journal of Finance* 33.4, pp. 1133–51.

Taffler, R. (2018). “Emotional finance: investment and the unconscious”. *The European Journal of Finance* 24.7, pp. 630–653.

Tetlock, P. C. (2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. *The Journal of Finance* 62.3, pp. 1139–1168.

Tran, V. L. and T. Leirvik (2020). “Efficiency in the markets of crypto-currencies”. *Finance Research Letters* 35, p. 101382.

Truby, J. (2018). “Decarbonizing Bitcoin: Law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies”. *Energy Research & Social Science* 44, pp. 399–410.

Tzouvanas, P., R. Kizys, and B. Tsend-Ayush (2020). “Momentum trading in cryptocurrencies: Short-term returns and diversification benefits”. *Economics Letters* 191, p. 108728.

Urquhart, A. (2016). “The Inefficiency of Bitcoin”. *Economics Letters* 148, pp. 80–82.

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

Voshmgir, S. (2020). *Token economy: how the Web3 reinvents the internet.* Second edition. Berlin: BlockchainHub Berlin.

Wang, Q., Q. Huang, X. Wu, J. Tan, and P. Sun (2023). “Categorical uncertainty in policy and bitcoin volatility”. *Finance Research Letters* 58, p. 104664.

Wei, W. C. (2018). “Liquidity and market efficiency in cryptocurrencies”. *Economics Letters* 168, pp. 21–24.

Yue, W., S. Zhang, and Q. Zhang (2021). “Asymmetric News Effects on Cryptocurrency Liquidity: an Event Study Perspective”. *Finance Research Letters* 41, p. 101799.

Chapter 2: Bitcoin Returns and YouTube News: A Behavioral Time Series Analysis

Pierre Fay, David Bourghelle, Fredj Jawadi

Univ. Lille, ULR 4999 - LUMEN, F-59000 Lille, France

Link to the published version:

https://www.tandfonline.com/doi/abs/10.1080/00036846.2024.2387870

Abstract

This study investigates whether investor's sentiment and attention information collected via YouTube can improve bitcoin return forecasts. Accordingly, we collected daily data over the period 2017-2023, covering calm and turbulent periods marked by different types and episodes of emotions. Unlike previous studies, we used YouTube videos to propose two sentiment proxies: investor attention to YouTube (daily number of YouTube video views) and investor sentiment on YouTube (number of positive and negative videos on YouTube). Interestingly, we break down both attention and sentiment per subject. Econometrically, we assess lead-lag effects between sentiment/attention and bitcoin return using causality tests and Vector Auto-regressive (VAR) model. We also evaluate the forecasting power of YouTube attention/sentiment data using a deep learning LSTM model. Our study shows two main results. First, we find lead-lag effects between bitcoin returns and per subject investor's attention and sentiment proxies. Second, we show that our deep learning LSTM model relying on the information provided by attention and sentiment supplants benchmark Buy and Hold Strategy to forecast future bitcoin returns.

Keywords: YouTube investor Sentiment, YouTube investor attention, bitcoin returns, VAR Model, LSTM Model.

JEL: C2, F10, G10.

1 Introduction

Following the informational Efficient Market Hypothesis (EMH) of Fama (1965, 1970), the price of a financial asset follows a random walk process and the returns of this financial asset approach a white noise. However, the EMH has failed to explain various financial stock market crises (1987 stock crash, 2000 dotcom bubble, 2007 subprime crises, etc), opening the door for an alternative model allowing further dependence for these returns. For a standard financial asset, economic fundamentals are mainly used to explain this dependence across returns, given that, at equilibrium, the price of an asset should converge toward its fundamental (Samuelson 1965). Chartist techniques can also be used to characterize and reproduce the return dependence dynamics. For cryptocurrencies, the fundamental analysis is less credible as they do not have an explicit fundamental value. Rather, alternative behavioral factors appear to drive the cryptocurrency prices, yielding different episodes of market ups and downs and raising the issue of crypto returns forecasting. Basically, investor sentiment and emotions appear to be key drivers, in line with the behavioral economics and finance theory expounded by Amos Tversky, Daniel Kahneman, Richard Thaler, and also extended by Robert Shiller and Richard Taffler through the irrational exuberance hypothesis (Shiller 2005), animal spirits hypothesis (Akerlof and Schiller 2009), the narrative economy (Shiller 2019), and the emotional finance hypothesis (Taffler 2018), among others. Indeed, forgetting about rationality and conventional finance theory teaching, behavioral economists consider that investors frequently make cognitive errors and make wrong decisions due to their own biases since their self-control is limited. This behavior and the power of these psychological factors is strong and is more likely to appear when the financial market is volatile and therefore open to strong investor appetite for risky trading, as with cryptocurrencies. Accordingly, the excess volatility for cryptocurrencies in general and bitcoin in particular has attracted considerable attention from media, investors, and regulators over the past few years. This is particularly interesting to better explain the dynamics of bitcoin price and to explain its up and down episodes. In fact, several studies have been conducted to analyze the dynamics of cryptocurrency returns, but the conclusions are not unanimous. Tran and Leirvik (2019) showed that the bitcoin market alternates between inefficiency and efficiency. Other papers showed that bitcoin became more efficient at the end of the study period in question (Urquhart 2016; Tran and Leirvik 2019), while in practice, several serious corrections have characterized bitcoin prices (i.e., 2018 and 2021 bubbles). This raises various questions about bitcoin market efficiency and the cryptocurrency drivers. Corbet, B. Lucey, et al. (2019) and Koutmos (2023) found that the bitcoin market is particularly sensitive to sentiment, potentially playing a key role in explaining bitcoin price changes. Several papers pointed to a correlation between investors' attention and cryptocurrencies prices. However, there is a lack of consensus regarding the direction of this relationship (Garcia et al. 2014; Bouoiyour and Selmi 2015; Urquhart 2018; Philippas et al. 2019; Nasir et al. 2019; Liu and Tsyvinski 2021).

In the same context, Ciaian, Rajcaniova, and Kancs (2016) studied the impact of investor attention, measured by the intensity of discussion on internet forums, and reached the same conclusion. Wei and Koutmos (2023) shows that an increase in attention by new investors can push Bitcoin prices and induced extra noise in the market. Consequently, while the related literature shows the usefulness of investor sentiment and attention, their conclusions vary with regard to measuring these two variables, which remains challenging. Indeed, newspapers are far from being investors' only source of information as the latter rely more on web 2.0, internet forums, and social media that have become the main drivers of investor sentiment and attention on cryptocurrencies (Mai, Bai, and Shan 2015; Philippas et al. 2019).

Empirically, different proxies have been used to measure investor sentiment, including a survey using a proxy variable (Internet searches, technical indicators, put/call ratios, etc.) as in Baker and Wurgler (2006), and an algorithm on media items (press articles, messages on social networks, Internet forums, etc.) as in Tetlock (2007).

Other sentiment analysis methods were also applied. For example, Araci (2019) used FinBERT, which is a transformer-based deep learning technique used to perform sentiment analysis on financial texts. To assess the impact of specific subjects on cryptocurrencies in the media and on social networks, several studies have used a topic modeling approach, a text mining technique that extracts subjects and associated keywords from a corpus. For example, Corbet, Larkin, et al. (2020) showed that macroeconomic news is useful in forecasting bitcoin returns. Phillips and Gorse 2018; Uras, Vacca, and Destefanis 2020; Ortu et al. 2022 showed that the occurrence of some subjects can help to predict price movement. Bitcoin price is also affected by VIX (Su et al. 2023), supply chain pressure (Qin, Su, et al. 2023) and carbon emission (Qin, Wu, et al. 2023). Long, B. M. Lucey, and Yarovaya (2021) combined topic modeling techniques with various sources of information (Google Trends, Reddit, Cryptocompare, forums, and news) to improve bitcoin forecasting. Overall, while these previous studies either focused on sentiment or on attention but less on these of two variables simultaneously, always, they relied on some proxies that do not capture the full and continuous information about sentiment. Further, their conclusions are often inconclusive. The present paper aims to fill this gap and investigate at the same time the impact of investor and sentiment attention on the formation of bitcoin returns. Unlike previous related studies, our paper is the first study to use Youtube videos, the second most popular social network providing more relevant news about sentiment, to assess the relationship between bitcoin returns and sentiment/attention news. This is a relevant and new contribution enabling us to capture and visualize investor's emotions and attention, which is relevant for investment decisions. Our findings show two interesting results. We propose an original measure of investor attention obtained by assessing the number of Youtube video views. Further, we proxy investor sentiment using the sentiment of Youtube videos. These two proxies are particularly useful as they capture investors' attention as well as the personal opinion and feelings of the publisher who uses Youtube. Second, we study the effect of investor attention and sentiment provided by Youtube on bitcoin returns. The impact of overall sentiment and the effect of attention and sentiment of specific subjects are investigated separately.

Our main result points to the relevance of information provided by YouTube video to perform the forecast of bitcoin. From that perspective, unlike the main rules of Fama's(1965) efficient market hypothesis, the bitcoin price does not reflect instantaneously and fully the whole available information, notably the information in that video. This implies that it would be possible to forecast bitcoin using the information of YouTube video, which consequently suggests that the market is not informationnally efficient.

The reminder of this paper is organized into four sections. Section 2 presents our data and explains the methodology. Section 3 discusses the empirical results. Section 4 concludes.

2 Data and Methodology

Our study uses daily data and covers the period: 17 August 2017 - 30 June 2023, using 2,144 observations, which includes several bitcoin market overreactions and crashes. The prices and volume data were collected from Binance API, which is the leading cryptocurrency exchange platform by volume. Bitcoin returns are computed as the first difference of bitcoin prices in logarithm. As for investors' attention and sentiment data, we used the API of YouTube to search for videos related to the keyword "Bitcoin". The total number of videos gathered was 94,420. For each video, we extracted the number of views and their respective titles. We present hereafter the process used to extract a sentiment from the video, the classification of videos per subject, and the calculation of our daily variables from these data to proxy investors' attention and sentiment.

2.1 The Sentiment Data

Various methodologies were used to analyze sentiment related to financial texts from online newspapers or social networks like Twitter or Reddit. Unlike previous studies, we extracted investor sentiment through an analysis of Youtube videos, while investigating the sentiment of Youtube videos based on their titles. To this end, we adopted the deep learning technique called FinBERT (Financial Bidirectional Encoder Representations from Transformers) proposed by Araci (2019) for sentiment analysis. FinBERT is a transformer based deep learning technique based on the BERT model published by Devlin et al. (2019) from Google AI Language.

Unlike other algorithms, FinBERT uses a contextual embedding, enabling it to learn multiple representations for each word in the document. The use of FinBERT thus allows us to outperform other techniques used in other studies, including VADER. In practice, we classify our videos as positive, negative, or neutral using their titles as input for the model. To illustrate this, we report some of the classification results in Table 4.

Table 4: Sentiment analysis on Youtube video titles using FinBERT

Title of the video	Sentiment
Why Alexandria green new deal is bullish...	positive
3 trends show ethereum is on track for strong growth...	positive
4 things you need to know about four tokens	neutral
5 altcoins to look out for this summer	neutral
4 reasons why bitcoin price continues to crash	negative
$31 million in Ethereum liquidates in past 12 hours	negative

Note: Table 4 shows some examples of FinBERT sentiment analysis of Youtube video titles from our database.

Our database specifies the sentiment variable, taking 1 for positive, 0 for neutral, and -1 for negative. This encoding enables us to compute daily key statistics for the sentiment variable.

2.2 Classification of the videos by subject

Next, we classify each Youtube video in the dataset by subject, using their titles. We set up a list of keywords for each subject. The subjects and associated keywords were first derived from a Latent Dirichlet Allocation (LDA) analysis using a coherence score to find the appropriate number of subjects. Second, we manually filtered the most relevant keywords for each subject. The classification algorithm (code in Listing 1) uses the list of keywords to classify videos among the corresponding subjects. An example of words associated with each subject is reported in Table 5.

Table 5: List of subjects and example of associated keywords

Subject	Associated keywords example
Hacks	scam, phished, hack, pirate, attack, steal
Network activities	mining, addresses, miner, farm, pools
Bitcoin adoption	partnership, adoption, accepted
Institutional and Central banks	institutional, bank, cdbc
Nft and Metaverse	nft, metaverse, opensea, axies, sport
Personality	ceo, burry, musk, butterin
Ico	ico, funding, participate, venture, capital
Trading robot	bot, robot
Regulation	ban, regulation, watchdog, lawsuit, authority
Price predictions	breakout, predict, analysis, high, resistance
Tutorial	explained, how, understanding

Note: Table 5 shows the subjects and some examples of associated keywords used to classify our Youtube videos.

We clean up the text and apply several pre-processing steps before moving to the classification algorithm. For example, we remove stopwords, lower the text, and delete special characters. We also filter the text to obtain only names and adjectives. The classification algorithm then examines each video title: if a title contains one keyword associated with a subject, it assigns the subject to the video or assigns the video to the "non classified" category. A video may be associated with multiple categories.

2.3 Calculation of the independent variables

Our dataset is composed of 94,420 videos. For each video, we note the number of views it received, the date of publication, the sentiment of the video, and the video subject. From these data, we compute the daily independent variables The first variable is the total number of views ($V_{t,s}$) on day t and the subject s,

$$V_{t,s} = \sum_{i=0}^{n} v_{t,s,i} \tag{6.1}$$

where: $v_{t,s,i}$ is the number of views of the video i on day t and the subject s. n is the total number of videos on day t.

We also compute $V_{t,All}$ that denotes the total number of views on day t for all subjects combined.

The second variable is the number of positive videos noted $E^{+}_{t,s}$, and the number of negative videos noted $E^{-}_{t,s}$ on day t for the subject s, (that for All subjects s is noted as All).

We compute a third independent variable noted $V^{+}_{t,s}$ and $V^{-}_{t,s}$ to test for the effect of the number of views on positive and negative videos for day t and the subject

s respectively (that for All subjects s is noted All).

This variable is calculated as the sum of views of positive (or negative) views of the corresponding videos respectively:

$$V_{t,s}^{+} = \sum_{i=0}^{n} v_{t,s,i}^{+} \tag{6.2}$$

and

$$V_{t,s}^{-} = \sum_{i=0}^{n} v_{t,s,i}^{-} \tag{6.3}$$

where $v_{t,s,i}^{+}$ is the number of views of positive video i on day t for the subject s. n is the total number of videos on day t. $v_{t,s,i}^{-}$ is the number of views of negative video i on day t and the subject s.

3 Empirical Analysis

We carried out a multivariate analysis to assess the effect of investors' attention and sentiment on bitcoin return, checking for lead-lag effects between bitcoin returns and our variables. To this end, we ran Granger causality tests and set up Vector Autoregressive (VAR) specifications that we used to estimate the impulse-response functions.

3.1 Estimating the impact of Youtube attention on Bitcoin returns

We report in Table 6, all the subjects, the number of videos per subject, and the total number of views per subject, along with the respective percentage it represents. From Table 3, we can note the high number of views, suggesting a relevant attention dynamic on bitcoin via Youtube. Further, the subject (Price Prediction), while considering the number of videos or the number of views concentrates the highest values (74.12%, 47.77% among the total number of videos and number of views, respectively), shows greater attention of bitcoin investors. The subject (Tutorials) arrives in second position, suggesting a need to learn more about bitcoin.

Table 6: Subject distribution in our dataset

Subject (s)	Number of videos	Number of views
Price predictions	77 204 (74.12)	470 241 450 (47.77)
Not classified	13 169 (12.64)	262 859 665 (26.7)
Tutorial	4 044 (3.88)	115 865 176 (11.77)
Institutional and Central banks	2 648 (2.54)	26 121 103 (2.65)
Personality	2 623 (2.52)	47 201 515 (4.79)
Regulation	1 318 (1.27)	14 831 081 (1.51)
Network activities	1 273 (1.22)	23 610 774 (2.4)
ICO	522 (0.50)	3 640 385 (0.37)
NFT and Metaverse	479 (0.46)	7 236 131 (0.74)
Hacks	381 (0.37)	5 737 304 (0.58)
Bitcoin Adoption	360 (0.35)	5 440 403 (0.55)
Trading robot	136 (0.13)	1 677 501 (0.17)

Note: Table 6 shows the numbers of videos per subject in our dataset. Values in (.) denote the value in percentage.

From Table 6, we note that all the videos except the non-classified and the "tutorial" videos concern the state of the market. Among these videos, price prediction represents 74.12% of the videos in our dataset, showing the importance of the technical analysis narrative among bitcoin traders on Youtube. This finding is not unexpected as, unlike other social platforms where information is transmitted via a text, Youtube allows for the creation of videos, the ideal media to show a graph and to comment on it with technical analysis. Further, while the price prediction subject represents 74.12% of videos published on Youtube, it only accounts for 47.77% of the total views received. Interestingly, we plot the 30 rolling days mean of the number of views received by the "price predictions" subject in Figure 1. We can see a peak of activity during the 2018 and 2021 bitcoin bubbles, suggesting further evidence of linkage between bitcoin price movements and the attention of investors on such videos with a lead-lag relationship.

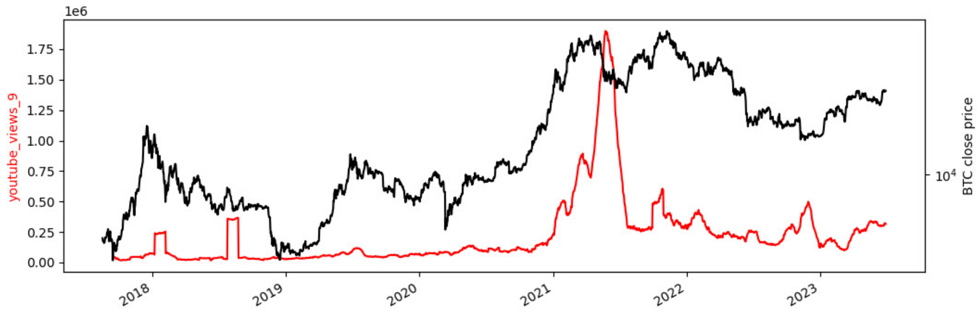


Figure 5: Number of views of "Price prediction" videos versus bitcoin price

We also noted a significant number of "tutorial" videos that explain how to buy the first Bitcoin, for example. "Tutorial" subject accounts for only 3.88% of bitcoin-related videos on Youtube, but these videos had received 11.77% of total views. In Figure 6, we can also see a rolling 30 days mean of the number of views received by the "tutorial" subject. These videos received more views in periods of bull and bear markets. Interestingly, we can observe a first major price move and spike in attention on tutorial videos during the 2018 bubble. It is possible that this phenomenon intensified due to increased adoption of cryptocurrency by the public between these two periods.

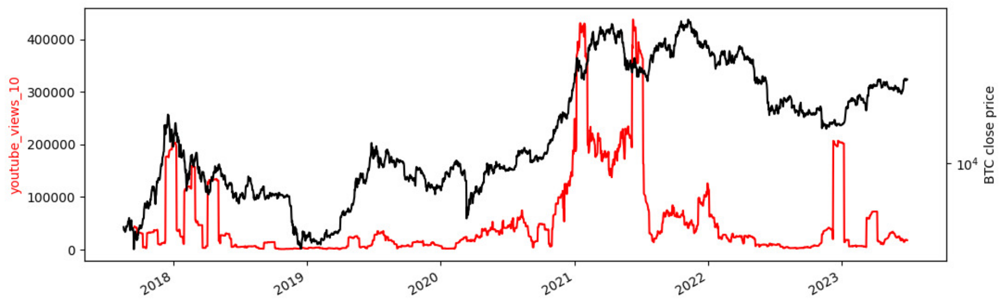


Figure 6: Number of views of “tutorial” videos versus bitcoin price

Unlike other social media platforms, YouTube gives us a direct measure of investor attention by publishing the number of times a video has been watched by users. Accordingly, we focused on the relationship between the number of views of bitcoin-related videos on YouTube and bitcoin returns.

Table 7: Results of stationarity tests

	ADF statistic (p-value)
$V_{Pricepredictions}$	-3.534 (0.007)
$V_{Tutorial}$	-6.912 (0.000)
$V_{InstitutionalAndCentralbanks}$	-12.901 (0.000)
$V_{Personality}$	-4.035 (0.001)
$V_{Regulation}$	-4.496 (0.000)
$V_{NetworkActivities}$	-46.08 (0.000)
V_{Ico}	-42.109 (0.000)
$V_{NftandMetaverse}$	-6.535 (0.000)
V_{Hacks}	-46.232 (0.000)
$V_{BitcoinAdoption}$	-13.07 (0.000)
$V_{Tradingrobot}$	-8.467 (0.000)
V_{All}	-3.407 (0.011)
r	-32.000 (0.000)

Note: Table 7 shows the results of the augmented Dickey–Fuller test (ADF) for our "number of views" variable, where the number of lags was specified using the Akaike information criterion. r denotes the bitcoin returns, while V denotes the number of views.

From Table 7, all the variables under consideration are stationary. We analyze the unconditional correlations between bitcoin returns and these proxies. We report the main results in Table 8.

Table 8: Unconditional Correlations between of V_s and bitcoin returns

	Correlation
$V_{Pricepredictions}$	-0.053
$V_{Tutorial}$	-0.017
$V_{InstitutionalAndCentralbanks}$	-0.009
$V_{Personality}$	-0.037
$V_{Regulation}$	-0.032
$V_{NetworkActivities}$	-0.015
V_{Ico}	-0.016
$V_{NftandMetaverse}$	-0.017
V_{Hacks}	-0.004
$V_{BitcoinAdoption}$	-0.002
$V_{Tradingrobot}$	-0.011
V_{All}	-0.035

Note: Table 8 shows the unconditional correlations between our variables V_s and bitcoin returns.

Accordingly, we noted that whatever the subject, the number of views is negatively but weakly and insignificantly correlated with bitcoin returns. To go further in the analysis of linkages between these variables, we check for causality relationships between the variables using a Granger Causality tests. We report the main results in Table 9 and obtain different findings.

Table 9: Granger causality test between V_s and bitcoin returns

Null hypotheses	F-statistic	p-value	p
$V_{Pricepredictions}$ does not Granger cause r	1.920	0.166	11
r does not Granger cause $V_{Pricepredictions}$	2.490	0.288	11
$V_{Tutorial}$ does not Granger cause r	9.222	0.002***	1
r does not Granger cause $V_{Tutorial}$	1.202	0.273	1
$V_{InstitutionalAndCentralbanks}$ does not Granger cause r	0.191	0.662	1
r does not Granger cause $V_{InstitutionalAndCentralbanks}$	1.970	0.161	
$V_{Personality}$ does not Granger cause r	6.081	0.048**	6
r does not Granger cause $V_{Personality}$	2.935	0.569	6
$V_{Regulation}$ does not Granger cause r	0.498	0.480	3
r does not Granger cause $V_{Regulation}$	1.657	0.437	3
$V_{NetworkActivities}$ does not Granger cause r	3.570	0.059*	1
r does not Granger cause $V_{NetworkActivities}$	5.044	0.025**	1
V_{Ico} does not Granger cause r	0.357	0.550	1
r does not Granger cause V_{Ico}	1.751	0.186	1
$V_{NftandMetaverse}$ does not Granger cause r	12.570	0.006***	3
r does not Granger cause $V_{NftandMetaverse}$	3.145	0.076*	3
V_{Hacks} does not Granger cause r	0.001	0.982	1
r does not Granger cause V_{Hacks}	0.217	0.642	1
$V_{BitcoinAdoption}$ does not Granger cause r	1.352	0.245	1
r does not Granger cause $V_{BitcoinAdoption}$	4.112	0.043**	1
$V_{Tradingrobot}$ does not Granger cause r	4.156	0.042**	1
r does not Granger cause $V_{Tradingrobot}$	11.349	0.246	1
V_{All} does not Granger cause r	17.844	0.037**	10
r does not Granger cause V_{All}	1.699	0.428	10

Note: F-Statistic denotes the statistic of the Fisher test and its p-value. ”p” denotes the lag number specified using the Partial Autocorrelation Function (PACF).

First, we found a unilateral causality relationship between the overall number of views (V_{All}) and Bitcoin returns (r) that is statistically significant at 5%, suggesting that overall investor attention does cause bitcoin returns, while bitcoin return does note Granger cause investor attention. Thus, the overall attention can be used to forecast Bitcoin returns. However, when considering disaggregated data and looking at the classification of our videos by subject, we obtain more significant causal relationships. This result confirms the importance of breaking down investor attention by subject when looking at the relationship between Youtube investors' attention and bitcoin returns. In particular, we found no causality relationship between bitcoin returns and V_{Hacks}, $V_{InstitutionalAndCentralbanks}$, V_{Ico}, $V_{Regulation}$ or $V_{Pricepredictions}$. For example, views related to Regulations or Institutions and Central Banks do not Granger cause bitcoin returns. This result suggests that investors consider that the regulation of crypto market and their adoption by financial institutions is even less possible, at least at present. Otherwise, the unilateral causality assumption of bitcoin returns is not rejected when considering views about Tutorials ($V_{Tutorial}$), Personality ($V_{Personality}$), Trading Robot ($V_{TradingRobots}$), Network Activities ($V_{NetworkActivities}$), NFT and Metaverse ($V_{NftandMetaverse}$).It suggests that the information related to investor attention on these Topics Granger causes bitcoin returns. We discuss these results in more detail hereafter.

Tutorial videos explain how to make an action like placing an order or selling a position. The related result shows that investor attention on Tutorial videos can provide useful information in forecasting future bitcoin returns.

Personality videos are videos talking about some very active personalities in the crypto internet space like Elon Musk or Vitalik Buterin among others. We can see that Bitcoin causality arising from $V_{Personality}$ is an interesting result. It is in line with the findings regarding Twitter, which explain how the tone of the world's wealthiest person can drive bitcoin returns.

Trading robots are videos promoting automatic crypto-trading robots, which often show unrealistic results selling the idea "get rich quick and easy". Attention on such videos can be seen as a proxy for investors' greed. The related result shows that investor attention on Trading robots videos can also provide useful information in forecasting future bitcoin returns.

Videos about Network activities are videos talking about the hashrate, transactions fees and about all the miners activities. Miners are people allocating processing power to verify and add transactions to a cryptocurrency's blockchain in exchange for a reward in Bitcoin. The related result shows a bilateral causality between investor attention on network activities and bitcoin return suggesting that consideration of investor attention (resp. bitcoin returns) to this subject helps to improve bitcoin return (resp. investor attention) forecasts.

NFTs and the metaverse began to receive significant attention in 2021, in part thanks to the success of the NFT art market, and Meta (formerly Facebook) famous project. Our results show a bilateral causality when observing investor's attention related to NFT and Metaverse ($V_{NftandMetaverse}$).

Overall, we found that the information obtained when capturing investors' attention using Youtube videos on subjects such as Personality, Network Activities, Trading Robot, Tutorials, and NFT Metaverse Granger cause bitcoin returns. These subjects can therefore help to improve the forecasting of bitcoin returns.

We will test this assumption hereafter.

In order to better assess these causality relationships, we ran a linear VAR model, allowing us to model the relations between these variables within a 2 equations system, for which each equation includes the lagged bitcoin return and a lagged value of the Youtube attention proxy. Taking the results of the Granger causality test, we only considered Youtube video subjects that had a significant lead-lag effect with the bitcoin return at the statistical level of 5%

For illustration, we set up a bilateral VAR specification, e.g., with one lag, between the bitcoin return r and the Number of views on "Tutorial" videos $V_{Tutorial}$ as:

$$\begin{cases} r_t = a_0 + a_1 r_{t-1} + a_2 V_{Tutorial,t-1} + e_{t1} \\ V_{Tutorial,t} = b_0 + b_1 V_{Tutorial,t-1} + b_2 r_{t-1} + e_{t2} \end{cases} \tag{6.4}$$

where: a_0, a_1, b_0 and b_1 denote the coefficients. e_{t1} and e_{t2} denote the error terms.

In practice, to set up the above VAR specification, the number of lags was selected using the Bayesian Information Criteria (BIC). Hereafter, we report the results of the VAR estimation for each subject under consideration. We also comment on the results of the estimate of the orthogonalized impulse response function (OIRF). Because the OIRF is sensible to the order, we checked that issue and we found no major difference in the results when changing the order of the variables.

For V_{All}, the results of the VAR model are reported in 10 and its respective OIRF function on Figure 7. Our results confirm a significant causality relationship between the overall attention on Youtube videos and bitcoin returns.

Table 10: Results of the linear VAR model for overall attention

	r		V_{All}	
$Constant$	0.668***	[13.887]	-0.025	[-0.818]
r_{t-1}	-0.052**	[-2.412]	0.013	[0.929]
r_{t-2}	0.046**	[2.107]	0.019	[1.387]
r_{t-3}	0.013	[0.582]	0.001	[0.089]
r_{t-4}	0.010	[0.473]	-0.002	[-0.157]
r_{t-5}	0.026	[1.180]	0.000	[0.036]
r_{t-6}	0.019	[0.866]	0.003	[0.222]
r_{t-7}	-0.020	[-0.938]	-0.013	[-0.937]
r_{t-8}	-0.029	[-1.336]	0.007	[0.530]
r_{t-9}	0.005	[0.239]	0.006	[0.471]
r_{t-10}	0.050**	[2.298]	0.004	[0.306]
$V_{All,t-1}$	0.027	[0.770]	0.143***	[6.568]
$V_{All,t-2}$	-0.020	[-0.565]	0.139***	[6.333]
$V_{All,t-3}$	0.023	[0.653]	0.089***	[4.019]
$V_{All,t-4}$	-0.012	[-0.327]	0.090***	[4.067]
$V_{All,t-5}$	0.062*	[1.743]	0.076***	[3.428]
$V_{All,t-6}$	0.016	[0.459]	0.067***	[3.009]
$V_{All,t-7}$	0.006	[0.171]	0.082***	[3.723]
$V_{All,t-8}$	-0.020	[-0.574]	0.036	[1.626]
$V_{All,t-9}$	-0.118***	[-3.393]	0.048**	[2.172]
$V_{All,t-10}$	-0.027	[-0.779]	0.065***	[2.960]
Number of observations		2134		
Log likelihood:		7189.582		
BIC		-12.263		
AIC		-12.375		
R2		0.0179		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant in each equation of the VAR model.

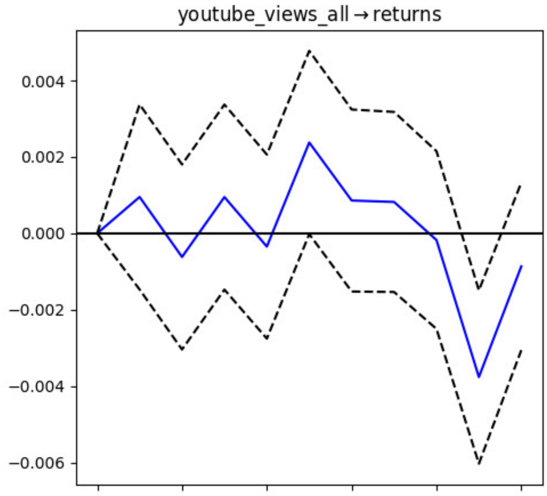


Figure 7: Orthogonalized Impulse response function of returns to investors' attention to all videos

In line with the results of the Granger causality tests, we found that when taking all the videos into account, overall attention (aggregate data) does significantly impact bitcoin returns. Interestingly, this impact is time-varying over time.

Next, we considered the estimation of VAR models for disaggregated data. For $V_{Tutorial}$ the VAR model results are reported in Table 11 and the OIRF is reported in Figure 8. There is no major difference in the relationship when changing the order. First, Investor attention to tutorials positively impacts bitcoin returns, confirming a lead-lag effect. Indeed, bitcoin returns can increase by more than 0.35% after a shock of one standard deviation on "Tutorial" videos, but this effect disappears after two days.

Table 11: Result of a linear VAR model using tutorial videos

	r		$V_{Tutorials}$	
$Constant$	0.748***	[48.492]	0.023*	[1.761]
r_{t-1}	-0.049**	[-2.297]	-0.020	[-1.096]
$V_{Tutorials,t-1}$	0.078***	[3.035]	0.032	[1.492]
Number of observations		2143.000		
Log likelihood:		6570.270		
BIC		-11.786		
AIC		-11.802		
R2		0.007		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. $Constant$ denotes the constant for the VAR model equations.

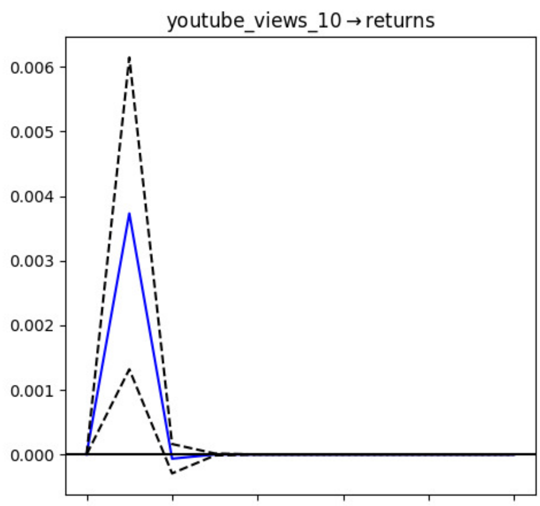


Figure 8: Orthogonalized Impulse response function of returns to "Tutorial" videos

For $V_{Personality}$, we reported the results of the VAR model estimation in Table 12. We plotted the OIRF in figure 9. When controlling for the order of variables, our result confirms the lead-lag unidirectional causality relationship, but unlike "Tutorials", a shock on investor attention to "Personality" videos implies a negative but time-varying and persistent reaction of bitcoin returns as the effect, even if it seems to pay off, does not completely disappear after 10 days. This result is interesting as it shows the power of "crypto personalities" over bitcoin returns.

Table 12: Results of a linear VAR model using personality videos

	r		$V_{Personality}$	
$Constant$	0.677***	[17.846]	-0.008	[-0.239]
r_{t-1}	-0.051**	[-2.352]	0.008	[0.452]
r_{t-2}	0.046**	[2.105]	-0.009	[-0.502]
r_{t-3}	0.009	[0.428]	0.001	[0.036]
r_{t-4}	0.008	[0.352]	-0.024	[-1.338]
r_{t-5}	0.025	[1.135]	0.017	[0.929]
r_{t-6}	0.016	[0.743]	0.023	[1.284]
$V_{Personality,t-1}$	-0.006	[-0.212]	0.095***	[4.372]
$V_{Personality,t-2}$	-0.054**	[-2.077]	0.110***	[5.075]
$V_{Personality,t-3}$	-0.020	[-0.766]	0.224***	[10.305]
$V_{Personality,t-4}$	0.024	[0.906]	0.105***	[4.831]
$V_{Personality,t-5}$	0.022	[0.859]	0.062***	[2.852]
$V_{Personality,t-6}$	-0.036	[-1.391]	0.079***	[3.643]
Number of observations		2138.000		
Log likelihood:		6592.505		
BIC		-11.749		
AIC		-11.818		
R2		0.010		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

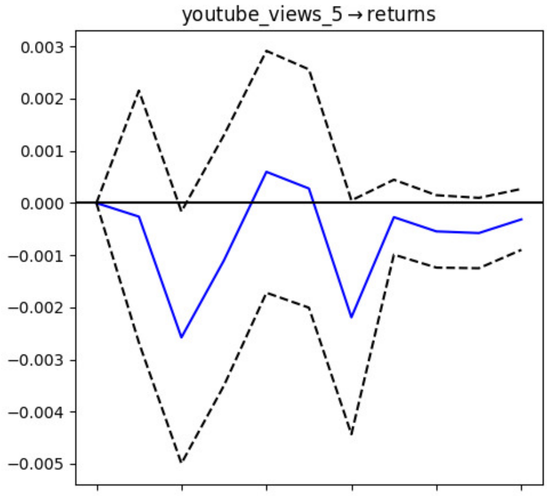


Figure 9: Orthogonalized Impulse response function of returns to investors' attention to "personality" videos

For $V_{Tradingrobots}$, the results of the linear VAR model (Table 13) also confirm the unidirectional causality relationship of these videos on bitcoin returns. Indeed, attention to such videos has a positive and significant effect on bitcoin returns, suggesting that investors remain sensitive to videos about bitcoin robots. Attention to such videos can be seen as a proxy for investor greed, which seems to have had a positive impact on bitcoin returns. The related OIRF is plotted in Figure 10 show a positive impact of 0.2%, but this disappeared after 2 days.

Table 13: Result of a linear VAR model with Trading robot videos

	r		$V_{TradingRobot}$	
$Constant$	0.749***	[48.519]	-0.008	[-0.951]
r_{t-1}	-0.050**	[-2.325]	0.014	[1.161]
$V_{TradingRobot,t-1}$	0.079**	[2.037]	0.011	[0.526]
Number of observations		2143		
Log likelihood:		7457.980		
BIC		-12.615		
AIC		-12.630		
R2		0.004		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

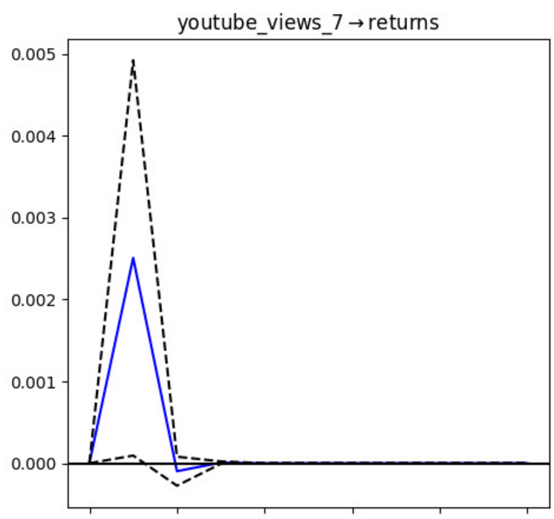


Figure 10: Orthogonalized Impulse response function of returns to investor attention on Trading robots" videos

For $V_{NetworkActivities}$, the results of the VAR model are reported in Table 14, and the impulse response function in Figure 11. While bitcoin returns significantly and negatively lead-lag the number of views on Network activities videos, the effect of the latter on bitcoin returns is positive and significant (but only at 10%) in the short-term. These results confirm the bidirectional causality relationship of attention to videos on bitcoin returns. It is interesting as it shows that narratives about miners have an impact and can be used to forecast bitcoin returns.

Table 14: Results of a linear VAR model using Network Activities

	r		$V_{NetworkActivities}$	
$Constant$	0.749***	[48.497]	0.021**	[2.519]
r_{t-1}	-0.050**	[-2.317]	-0.026**	[-2.244]
$V_{NetworkActivities,t-1}$	0.076*	[1.888]	0.003	[0.157]
Number of observations		2143		
Log likelihood:		7544.912		
BIC		-12.696		
AIC		-12.712		
R2		0.004		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

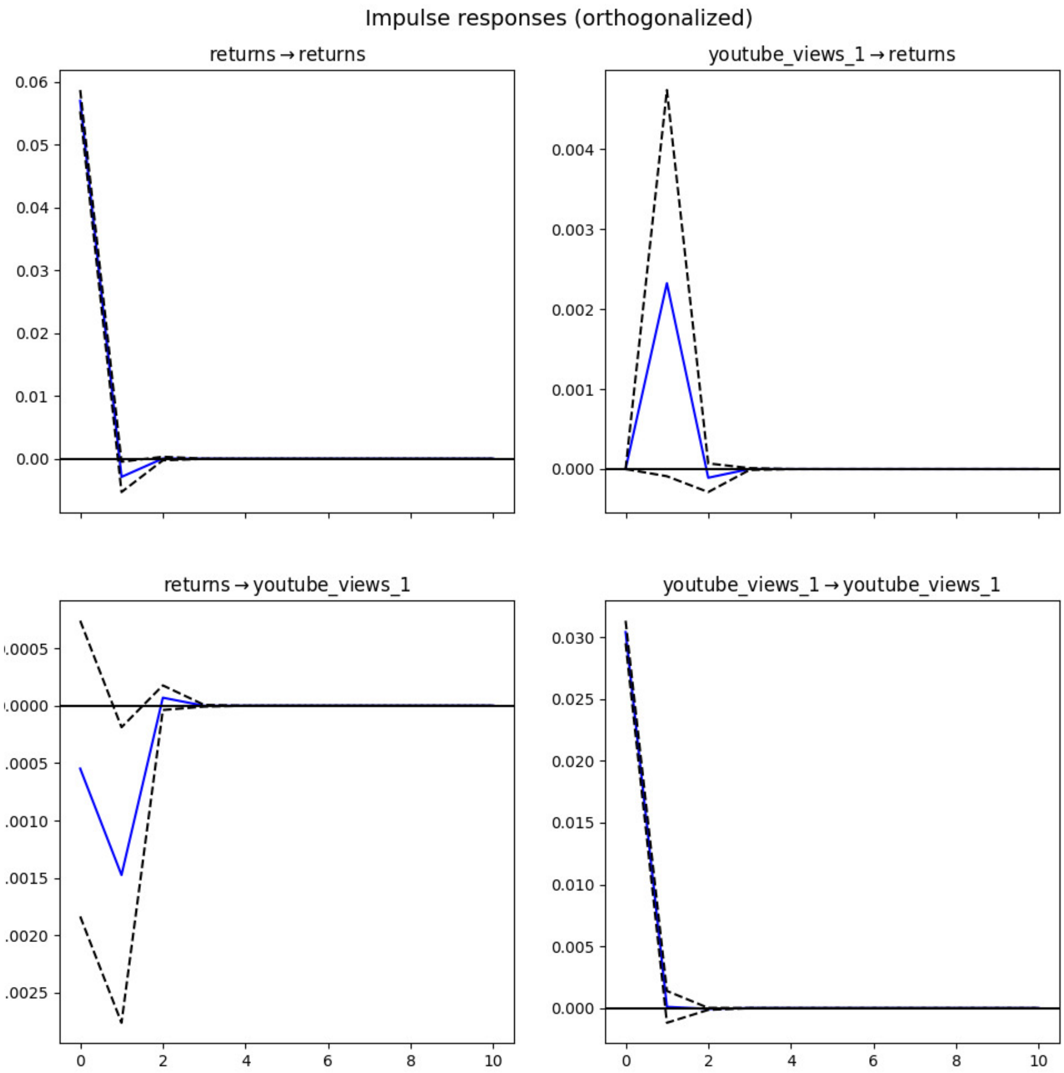


Figure 11: Impulse response function for "Network Activities" videos and returns

Finally, for $V_{NftAndMetaverse}$, the Granger analysis shows a weak bilateral causality relationship. We report the results of the VAR model in Table 15 and the related impulse response function in Figure 12. Accordingly, these results do not validate the bilateral relationship. Indeed, attention on such videos has a significant negative impact on Bitcoin returns for 7 days. The inverse relationship is not true.

Table 15: Results of a linear VAR model using NFT and Metaverse

	r		$V_{NftAndMetaverse}$	
$Constant$	0.714***	[26.227]	0.011	[0.942]
r_{t-1}	-0.052**	[-2.429]	-0.016	[-1.617]
r_{t-2}	0.044**	[2.016]	0.001	[0.069]
r_{t-3}	0.009	[0.419]	0.002	[0.193]
$V_{NftAndMetaverse,t-1}$	-0.054	[-1.103]	0.046**	[2.123]
$V_{NftAndMetaverse,t-2}$	-0.075	[-1.554]	0.053**	[2.433]
$V_{NftAndMetaverse,t-3}$	-0.136***	[-2.805]	0.053**	[2.448]
Number of observations		2141		
Log likelihood:		7937.513		
BIC		-13.040		
AIC		-13.077		
R2		0.011		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

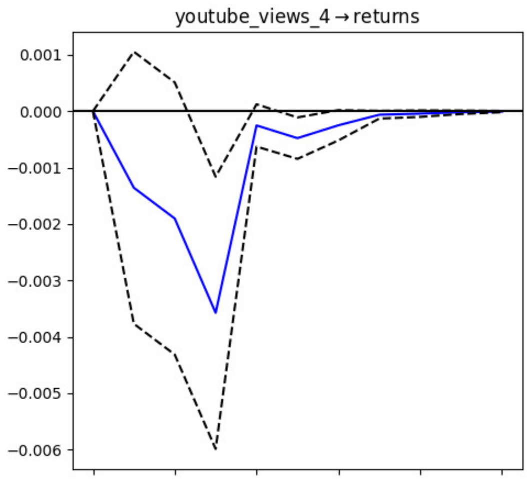


Figure 12: Impulse response function for “NFT and Metaverse” videos and returns

Overall, we found that investors' attention shows a lead-lag effect on bitcoin returns and help to improve the forecast of its future dynamics. However, the results vary with the video subject under consideration. While the intensity of views of videos related to subjects like "Tutorials", "Robot Trading", and "Network Activities" shows a positive and significant effect on bitcoin returns, the effect disappears after 1-2 days. Further, bitcoin returns react significantly and negatively to a shock on video views related to "Personality" and "NFT and Metaverse". Thus, the first contribution of this paper, shows the usefulness of per-subject investors' attention to explain bitcoin returns. In order to better characterize this relationship between investor's attention and bitcoin return, we explore in a second step the reaction of bitcoin return to investor's sentiment.

3.2 Estimating the impact of Youtube sentiment on Bitcoin returns

First, we analyze some statistics related to Youtube sentiment. Of the 94 420 videos, only 11 837 have been identified as positive or negative, the majority of Youtube titles have been identified by FinBERT as neutral. We report in 16, the number of days with positive and negative videos on Youtube related to the subjects under consideration.

We note that for most subjects, except for Regulation, the number of days with positive sentiment is higher than the number of days with negative sentiment. This result might suggest that crypto investors mistrust regulations.

Table 16: Number of days with positive and negative videos by subject on Youtube

Subject	Number of days with positive sentiment	Number of days with negative sentiment
Price predictions	3547	2816
Institutional and Central banks	198	135
Tutorials	162	94
Personality	138	118
Network activities	128	127
Regulation	103	183
Bitcoin adoption	56	19
Ico	41	27
Nft Metaverse	38	13
Hacks	23	37
Bot	5	0

Note: This table shows the number of positive and negative sentiment videos in our dataset.

Hereafter, we investigate the impact of the daily number of positive and negative videos on Bitcoin returns. In particular, we carry out the test only for subjects for which there is a minimum of 100 days of observations

We plot the rolling 30-day mean of daily negative sentiment of ”Price prediction” videos to obtain an overview of their interaction (Figure 13). We note that sentiment is time-varying and volatile. Further, while sentiment appears to be less correlated with bitcoin at the beginning of the period, we observe a more proactive relationship after 2021. This linkage seems to alternate between positive and negative and varies with the cycle phase: bear versus bull market.

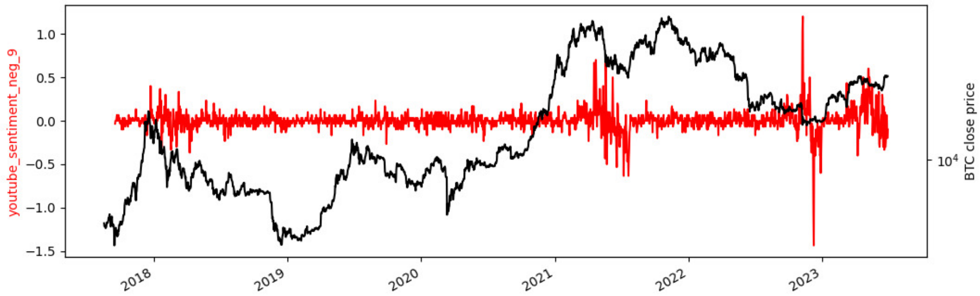


Figure 13: Daily negative sentiment of ”Price prediction” videos versus Bitcoin price

Next, we check the stationarity hypothesis for our sentiment variables using ADF tests. Results are reported in Table 17. All variables, except E^{-}_{All} , are stationary.

Table 17: Results of Unit Root Tests

	ADF statistic (p-value)
$E^{+}_{NetworkActivities}$	-6.048 (0.000)
$E^{-}_{NetworkActivities}$	-6.645 (0.000)
$E^{+}_{InstitutionalAndCentralbanks}$	-3.431 (0.01)
$E^{-}_{InstitutionalAndCentralbanks}$	-7.277 (0.000)
$E^{+}_{Personality}$	-3.268 (0.016)
$E^{-}_{Personality}$	-4.845 (0.000)
$E^{+}_{Regulation}$	-5.877 (0.000)
$E^{-}_{Regulation}$	-3.688 (0.004)
$E^{+}_{Pricepredictions}$	-11.694 (0.000)
$E^{-}_{Pricepredictions}$	-14.479 (0.000)
$E^{+}_{Tutorial}$	-5.182 (0.000)
$E^{-}_{Tutorial}$	-6.055 (0.000)
E^{+}_{All}	-10.185 (0.000)
E^{-}_{All}	-2.897 (0.046)
r	-32.000 (0.000)

Note: This table reports the results of the Augmented Dickey–Fuller (ADF) test for our sentiment variables.

Next, we compute the unconditional correlation reported in Table 18. Even if, in general, these correlations remain relatively low, the number of negative (positive) videos is negatively (positively) correlated with bitcoin returns, except for two subjects: Tutorial and Regulation.

Table 18: Unconditional correlation between Youtube sentiment and bitcoin returns

	Correlation
$E^{+}_{NetworkActivities}$	0.013
$E^{-}_{NetworkActivities}$	-0.067
$E^{+}_{InstitutionalAndCentralbanks}$	0.033
$E^{-}_{InstitutionalAndCentralbanks}$	-0.052
$E^{+}_{Personality}$	0.001
$E^{-}_{Personality}$	-0.029
$E^{+}_{Regulation}$	-0.005
$E^{-}_{Regulation}$	-0.021
$E^{+}_{Pricepredictions}$	0.101
$E^{-}_{Pricepredictions}$	-0.143
$E^{+}_{Tutorial}$	-0.017
$E^{-}_{Tutorial}$	-0.045
E^{+}_{All}	0.094
E^{-}_{All}	-0.109

Note: This table reports the unconditional correlations between Youtube sentiment proxies E^{-}_{s}, E^{+}_{s} and bitcoin returns.

To better investigate these linkages, we ran a Granger causality tests and report the main results in Table 19.

Table 19: Results of Granger causality test between E_s^+, E_s^- and bitcoin returns

Null hypotheses	F-statistic	p-value	p
$E^+_{NetworkActivities}$ does not Granger cause r	1.728	0.189	6
r does not Granger cause $E^+_{NetworkActivities}$	1.570	0.814	6
$E^-_{NetworkActivities}$ does not Granger cause r	0.002	0.965	1
r does not Granger cause $E^-_{NetworkActivities}$	7.136	0.008***	1
$E^+_{InstitutionalAndCentralbanks}$ does not Granger cause r	7.495	0.024**	4
r does not Granger cause $E^+_{InstitutionalAndCentralbanks}$	0.509	0.476	4
$E^-_{InstitutionalAndCentralbanks}$ does not Granger cause r	1.038	0.792	3
r does not Granger cause $E^-_{InstitutionalAndCentralbanks}$	14.520	0.000***	3
$E^+_{Personality}$ does not Granger cause r	6.681	0.035**	2
r does not Granger cause $E^+_{Personality}$	2.540	0.281	2
$E^-_{Personality}$ does not Granger cause r	6.942	0.031**	5
r does not Granger cause $E^-_{Personality}$	8.393	0.004***	5
$E^+_{Regulation}$ does not Granger cause r	0.865	0.834	6
r does not Granger cause $E^+_{Regulation}$	0.607	0.436	6
$E^-_{Regulation}$ does not Granger cause r	0.113	0.945	2
r does not Granger cause $E^-_{Regulation}$	23.571	0.000***	2
$E^+_{Pricepredictions}$ does not Granger cause r	1.243	0.265	20
r does not Granger cause $E^+_{Pricepredictions}$	13.301	0.102	20
$E^-_{Pricepredictions}$ does not Granger cause r	29.213	0.023**	20
r does not Granger cause $E^-_{Pricepredictions}$	28.478	0.000***	20
$E^+_{Tutorial}$ does not Granger cause r	1.505	0.471	3
r does not Granger cause $E^+_{Tutorial}$	1.939	0.379	3
$E^-_{Tutorial}$ does not Granger cause r	2.152	0.143	3
r does not Granger cause $E^-_{Tutorial}$	34.466	0.000***	3
E^+_{All} does not Granger cause r	1.409	0.235	20
r does not Granger cause E^+_{All}	4.817	0.090*	20
E^-_{All} does not Granger cause r	2.991	0.224	4
r does not Granger cause E^-_{All}	31.082	0.000***	4

Note: F-Statistic and p-value denote the Fisher test statistic and its p-value respectively. p is the number of lags. (***), (**) and (*) denote the significance at the statistical levels of 1%, 5%, and 10% respectively.

Overall, our results show that Bitcoin returns Granger cause the number of negative videos for of all the subjects tested. We also obtain a bilateral Granger causality between $E^{-}_{Personality}$ and Bitcoin returns, and between $E^{-}_{Pricepredictions}$ and Bitcoin returns. For positive sentiment, only $E^{+}_{InstitutionalAndCentralbanks}$ and $E^{+}_{Personality}$ Granger cause Bitcoin returns. These findings suggest that bitcoin would be more sensitive to negative than to positive sentiment. To better clarify these relationships between sentiment variables and bitcoin returns, we ran a linear VAR model which allows us to model the relationships within a 2-equation system. For each model, we select the number of lags using the Bayesian Information Criteria (BIC).

First, when considering the number of negative videos for all subjects on Youtube E^{-}_{All}, our results confirm a negative and significant lead-lag unidirectional relationship with bitcoin returns. VAR results are reported in Table 20 and its OIRF in Figure 14. We find that the number of negative videos reacts negatively to a shock on bitcoin returns, which suggests that when Bitcoin prices increase (resp. decrease), the number of negative videos on Youtube tend to decrease (resp. increase). This result confirms the intuition that negative returns have an impact on the overall negative sentiment on Youtube. It also shows that even if bitcoin returns can be useful to forecast negative sentiment on Youtube, the overall number of negative videos are not useful to forecast returns.

Table 20: Results of a linear VAR model for overall negative Youtube sentiment

	r		E_{All}	
$Constant$	0.707***	[21.856]	0.082***	[2.997]
r_{t-1}	-0.045**	[-2.059]	-0.130***	[-7.007]
r_{t-2}	0.049**	[2.205]	0.011	[0.573]
r_{t-3}	0.003	[0.139]	-0.004	[-0.233]
r_{t-4}	0.003	[0.159]	0.017	[0.908]
$E^{-}_{All,t-1}$	0.027	[1.053]	0.536***	[24.784]
$E^{-}_{All,t-2}$	-0.028	[-0.949]	0.104***	[4.212]
$E^{-}_{All,t-3}$	-0.039	[-1.319]	0.049**	[1.971]
$E^{-}_{All,t-4}$	0.025	[0.991]	0.182***	[8.460]
Number of observations		2140		
Log likelihood:		6592.953		
BIC		-11.773		
AIC		-11.821		
R2		0.007		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

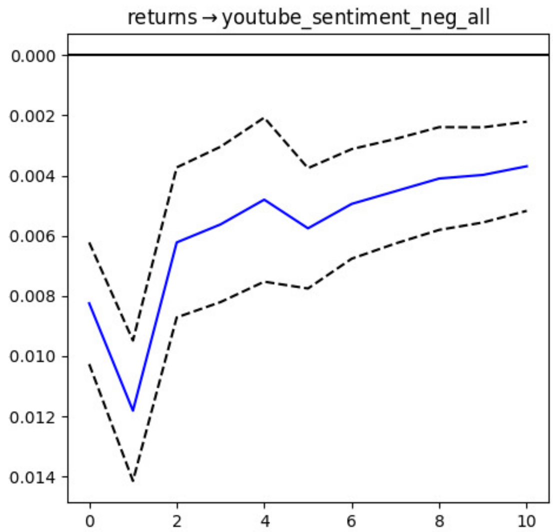


Figure 14: Orthogonalized Impulse response function of returns to overall Youtube negative sentiment

Thus, we consider the disaggregated measures of sentiment for negative videos about Personalities ($E^{-}_{Personality}$). The estimated results of the VAR model are reported in Table 21 with its respective OIRF in Figure 15. VAR results do not validate this relationship and indicate that $E^{-}_{Personality}$ is affected by bitcoin returns but that the inverse does not hold.

Table 21: Results of a linear VAR model for negative videos on “Personality”

	r		$E^{-}_{Personality}$	
$Constant$	0.696***	[19.850]	0.018	[0.846]
r_{t-1}	-0.052**	[-2.421]	-0.025*	[-1.903]
r_{t-2}	0.043**	[1.991]	0.002	[0.175]
r_{t-3}	0.007	[0.301]	-0.004	[-0.282]
r_{t-4}	0.007	[0.321]	-0.012	[-0.891]
r_{t-5}	0.022	[1.008]	0.015	[1.160]
$E^{-}_{Personality,t-1}$	-0.008	[-0.231]	0.100***	[4.678]
$E^{-}_{Personality,t-2}$	-0.056	[-1.600]	0.052**	[2.454]
$E^{-}_{Personality,t-3}$	-0.039	[-1.179]	0.271***	[13.328]
$E^{-}_{Personality,t-4}$	-0.014	[-0.399]	0.211***	[10.018]
$E^{-}_{Personality,t-5}$	0.006	[0.166]	0.134***	[6.251]
Number of observations		2139		
Log likelihood:		7262.356		
BIC		-12.387		
AIC		-12.446		
R2		0.009		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. $Constant$ denotes the constant parameters.

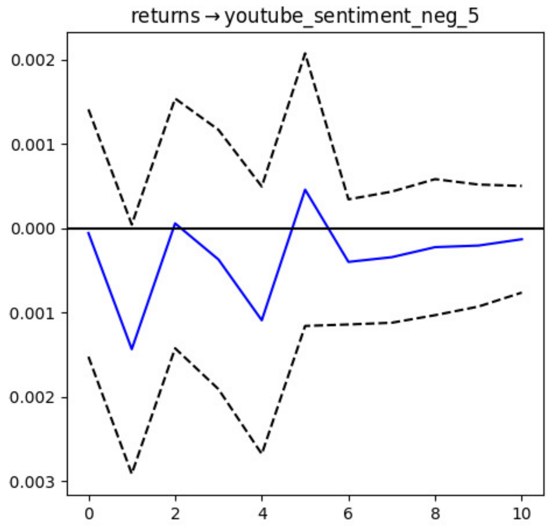


Figure 15: Orthogonalized Impulse response function of returns on sentiment about Personality

The same results as $E^{-}_{Personality}$ holds for $E^{-}_{PricePredictions}$, $E^{-}_{NetworkActivities}$, $E^{-}_{InstitutionalAndCentralbanks}$, $E^{-}_{Regulation}$, and $E^{-}_{Tutorials}$ (Table 28, 29, 30, 31, 32 and their respective OIRF in Figure 19, 20, 21, 22, 23 in the appendices).

For positive videos relationship with bitcoin returns, the Granger causality test indicates that $E^{+}_{InstitutionalAndCentralbanks}$ and $E^{+}_{Personality}$ should help to forecast bitcoin returns.

We run a VAR for $E^{+}_{InstitutionalAndCentralbanks}$ and bitcoin return and we report our VAR results in Table 22 and its OIRF in Figure 16. Our results point to a unidirectional lead-lag relationship from Institutional and Central banks positive videos to bitcoin returns, suggesting that when the number of such videos increases (resp. decrease), bitcoin returns increase (resp. decrease) the next day, then it decreases (increase) for the following 8 days. This result is interesting because it shows that bitcoin investors tend to overreact to positive news about institutional and central banks, which can explain a first increase in prices followed by a correction of the market

Table 22: Results of a linear VAR model for positive videos on “Institutional And Central banks”

	r		$E^{+}_{InstitutionalAndCentralbanks}$	
$Constant$	0.699***	[22.413]	0.017	[0.345]
r_{t-1}	-0.050**	[-2.287]	-0.018	[-0.540]
r_{t-2}	0.048**	[2.217]	0.033	[0.984]
r_{t-3}	0.012	[0.534]	-0.005	[-0.148]
r_{t-4}	0.011	[0.493]	-0.015	[-0.456]
$E^{+}_{InstitutionalAndCentralbanks,t-1}$	0.027*	[1.929]	0.149***	[6.908]
$E^{+}_{InstitutionalAndCentralbanks,t-2}$	-0.030**	[-2.111]	0.148***	[6.713]
$E^{+}_{InstitutionalAndCentralbanks,t-3}$	-0.008	[-0.530]	0.077***	[3.494]
$E^{+}_{InstitutionalAndCentralbanks,t-4}$	-0.009	[-0.623]	0.102***	[4.662]
Number of observations	2140			
Log likelihood:	5275.568			
BIC	-10.542			
AIC	-10.589			
R2	0.009			

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

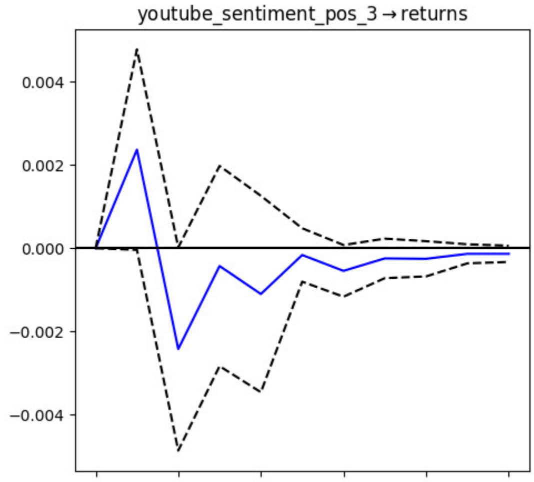


Figure 16: Orthogonalized Impulse response function of returns on positive sentiment about "Institutional And Central banks"

Finally, Granger causality results indicate that the number of positive videos about personality ($E^{+}_{Personality}$) Granger cause bitcoin returns. Accordingly, we estimate a VAR model for bitcoin returns and $E^{+}_{Personality}$ and report the main results in Table 23 and Figure 17. Our results indicate that $E^{+}_{Personality}$ is negatively related to bitcoin returns suggesting that when the number of positive videos about personality increases (resp. decreases), bitcoin returns tend to decrease (resp. increase) for the next 8 to 10 days.

Table 23: Results of a linear VAR model for positive videos on "Personality"

	r		$E^{+}_{Personality}$	
$Constant$	0.717***	[32.058]	0.032	[1.615]
r_{t-1}	-0.050**	[-2.306]	-0.006	[-0.307]
r_{t-2}	0.046**	[2.154]	-0.030	[-1.575]
$E^{+}_{Personality,t-1}$	-0.018	[-0.725]	0.233***	[11.050]
$E^{+}_{Personality,t-2}$	-0.052**	[-2.150]	0.214***	[10.136]
Log likelihood:		6493.116		
BIC		-11.703		
AIC		-11.729		
R2		0.008		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

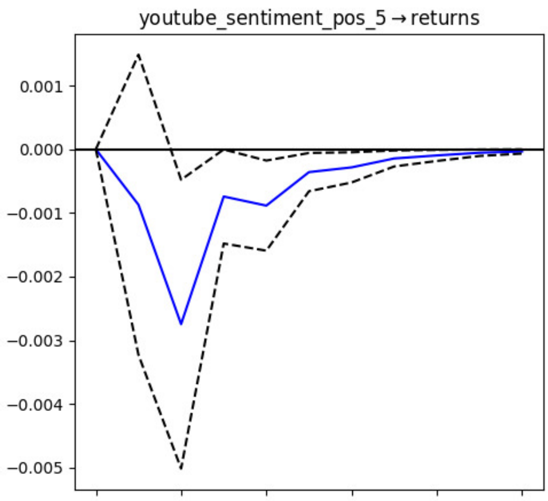


Figure 17: Orthogonalized impulse response function of returns on positive sentiment about "Personality"

Overall, both YouTube investors' attention and sentiment appear to drive bitcoin returns. However, the size and sign of these lead-lag effects depend on the subject under consideration.

Indeed, when looking at the attention variable, for aggregated subjects, investor attention has a significant effect on Bitcoin returns. This effect is mostly positive during the first 8 days. This result is consistent with the literature. In fact, many studies showed a positive effect of investor's attention on Bitcoin returns. Among them, Kristoufek (2013) showed that the overall attention on Bitcoin, measured by the number of views on Wikipedia, has a positive impact on Bitcoin prices. Liu and Tsyvinski (2021) also used Google Search as an attention proxy and showed that high investor attention predicts high future returns. Bouoiyour and Selmi (2015), Philippas et al. (2019) or Nasir et al. (2019) also showed a positive impact of investor's attention on Bitcoin returns but those papers are studying the overall attention on Bitcoin.

Unlike previous related studies, we contribute while breaking down the overall attention of investors on YouTube into different subjects. We show that, the relationship is not the same for all the subjects, and that not all subjects are significant. In fact, the number of views of videos related to "Tutorials", "Robot Trading", and "Network Activities" have a positive effect on bitcoin returns that fades after 1-2 days. We also show that Bitcoin returns react negatively to subjects like "Personality" or "NFT and Metaverse".

Regarding sentiment, we also distinguish positive and negative videos. For the overall sentiment, there is no significant causality relationship with the overall positive sentiment and bitcoin returns. However, we show that the overall negative sentiment is Granger caused by Bitcoin returns but that it has no significant effect on bitcoin returns.

The number of positive videos appears to drive bitcoin returns only for “Personality” and ”Institutional And Central banks” subjects. These results are intuitive and in line with previous studies. Indeed Auer and Claessens (2018) showed that the market reacts positively to events that help legally recognize Bitcoin as a particular asset. Shanaev et al. (2020) also showed that the market reacts positively when the news supports the development of cryptocurrencies. For ”Personality” videos, the number of positive videos is negatively correlated with Bitcoin returns. It may be surprising but, it can be explained by the fact that traders overreact quickly (during the day) to a tweet of a personality. A Youtube video takes time to be published so when it is released to comment on this event, the market is already correcting the price. For example, Ante (2023) showed that non-negative tweets from Elon Musk lead to significantly positive abnormal Bitcoin returns in less than 2 hours.

The number of negative videos is mainly driven by bitcoin returns. This result shows that on Youtube, negative videos are released in reaction to the bear market but they seem to have no significant impact on bitcoin returns.

Overall, these results are interesting but suffer from two limitations. On the one hand, although investor attention and investor sentiment can interact, they were investigated separately. On the other hand, the linear framework only captures linear linkages between bitcoin returns and these drivers. In the next section, we fill this gap while investigating the combined effect of investors’ attention and sentiment on bitcoin returns. In practice, we define two variables noted V_s^+ and V_s^- as the total number of views of positive and negative videos (respectively) to a given subject s. Interestingly, in so doing, we simultaneously test whether both YouTube attention and YouTube sentiment can help to explain the dynamics of bitcoin returns, indirectly enabling us to assess for nonlinear YouTube sentiment/attention effects.

3.3 Assessing the joint impact of Youtube sentiment and attention on bitcoin returns

We first check the stationarity of the combined variables, noted V_s^+ and V_s^- hereafter, and we report the results of the ADF test in Table 24. Except V_{All}^-, which was not stationary in the level but stationary in the first difference, we find that the assumption of the unit root is rejected for all the other series under consideration.

Table 24: Results of Unit Root Test

	ADF statistic (p-value)
$V^{+}_{Tutorial}$	-46.466 (0.000)
$V^{-}_{Tutorial}$	-45.759 (0.000)
V^{+}_{All}	-15.864 (0.000)
$V^{+}_{NetworkActivities}$	-12.425 (0.000)
$V^{-}_{InstitutionalAndCentralbanks}$	-10.298 (0.000)
$V^{+}_{InstitutionalAndCentralbanks}$	-8.05 (0.000)
$V^{-}_{Personality}$	-7.961 (0.000)
$V^{-}_{NetworkActivities}$	-5.976 (0.000)
$V^{+}_{Regulation}$	-5.877 (0.000)
$V^{+}_{Personality}$	-5.814 (0.000)
V^{-}_{All}	-5.108 (0.000)
$V^{-}_{Pricepredictions}$	-4.235 (0.001)
$V^{-}_{Regulation}$	-3.688 (0.004)
$V^{+}_{Pricepredictions}$	-3.223 (0.019)
r	-32.0 (0.000)

Note: This table presents the main results of the Augmented Dickey–Fuller test (ADF) for our "mixed" variables (M_s).

Next, we analyze the correlations of the mixed variables (V_s) with bitcoin returns. The main results are reported in Table 25.

Table 25: Unconditional Correlations between the variables M_s and bitcoin returns

	Correlation
$V^{-}_{Pricepredictions}$	-0.123
V^{-}_{All}	-0.093
$V^{-}_{Tutorial}$	-0.06
$V^{-}_{InstitutionalAndCentralbanks}$	-0.044
$V^{-}_{NetworkActivities}$	-0.027
$V^{+}_{Regulation}$	-0.019
$V^{+}_{Tutorial}$	-0.013
$V^{-}_{Personality}$	-0.011
$V^{-}_{Regulation}$	-0.005
$V^{+}_{NetworkActivities}$	0.008
$V^{+}_{Pricepredictions}$	0.026
$V^{+}_{InstitutionalAndCentralbanks}$	0.043
$V^{+}_{Personality}$	0.057
V^{+}_{All}	0.064

Note: This table shows the unconditional correlations between our M_s variables and bitcoin returns.

Accordingly, we note further evidence of weak linear correlation between our mixed variables and bitcoin returns whatever the subject under consideration. The most correlated variable is $V^{-}_{Pricepredictions}$. In the next step, we investigate the linkages between these variables through a Granger causality test analysis. We report the main results in Table 26 and reach different conclusions.

Table 26: Results of Granger causality test between V_s^+, V_s^- and bitcoin returns

Null hypotheses	F-statistic	p-value	p
$V^+_{NetworkActivities}$ does not Granger cause r	0.392	0.531	1
r does not Granger cause $V^+_{NetworkActivities}$	0.880	0.348	
$V^-_{NetworkActivities}$ does not Granger cause r	0.502	0.479	1
r does not Granger cause $V^-_{NetworkActivities}$	2.021	0.155	1
$V^+_{InstitutionalAndCentralbanks}$ does not Granger cause r	0.496	0.481	1
r does not Granger cause $V^+_{InstitutionalAndCentralbanks}$	11.3094	0.001***	1
$V^-_{InstitutionalAndCentralbanks}$ does not Granger cause r	0.779	0.377	1
r does not Granger cause $V^-_{InstitutionalAndCentralbanks}$	4.414	0.036**	1
$V^+_{Personality}$ does not Granger cause r	0.918	0.338	1
r does not Granger cause $V^+_{Personality}$	8.621	0.003***	1
$V^-_{Personality}$ does not Granger cause r	1.630	0.202	1
r does not Granger cause $V^-_{Personality}$	3.932	0.047**	1
$V^+_{Regulation}$ does not Granger cause r	0.595	0.441	1
r does not Granger cause $V^+_{Regulation}$	0.128	0.721	1
$V^-_{Regulation}$ does not Granger cause r	0.034	0.855	1
r does not Granger cause $V^-_{Regulation}$	1.596	0.207	1
$V^+_{Pricepredictions}$ does not Granger cause r	7.304	0.121	7
r does not Granger cause $V^+_{Pricepredictions}$	6.420	0.011**	7
$V^-_{Pricepredictions}$ does not Granger cause r	3.480	0.323	7
r does not Granger cause $V^-_{Pricepredictions}$	11.608	0.001***	7
$V^+_{Tutorial}$ does not Granger cause r	0.018	0.892	1
r does not Granger cause $V^+_{Tutorial}$	0.793	0.373	1
$V^-_{Tutorial}$ does not Granger cause r	0.878	0.349	1
r does not Granger cause $V^-_{Tutorial}$	16.055	0.000***	1
V^+_{All} does not Granger cause r	0.280	0.597	20
r does not Granger cause V^+_{All}	2.135	0.144	20
V^-_{All} does not Granger cause r	15.176	0.010**	7
r does not Granger cause V^-_{All}	27.641	0.000***	7

Note: F-Statistic and p-value denote the Fisher test statistic and its p-value respectively. p is the number of lags. (***), (**) and (*) denote the significance at the statistical levels of 1%, 5%, and 10% respectively.

Combining sentiment and attention through these mixed variables appears to point to some form of nonlinear relationship between sentiment, attention, and bitcoin returns.

In fact, we found that when examining sentiment and attention separately, overall negative videos have a unidirectional causality effect on bitcoin returns, but when the two variables are examined together, there is evidence of a bidirectional causality relationship. The results of the VAR linear model is reported in Table 27 with its OIRF in Figure 18.

These VAR results confirm the bilateral relationship: when there is a shock on both attention and the number of negative videos published, bitcoin returns tend to be unstable for the next 10 days but the first 3 days can lead to a feedback loop. In fact, a positive (resp. negative) shock in Bitcoin returns will reduce (resp. rise) the attention on negative videos and the number of these videos. Because of this drop (resp. rise) in both attention and number of negative videos, in the second and third day, returns tend to increase (resp. decrease) which in turn reduces (resp. rises) the attention on negative videos and the number of those videos.

Table 27: Results of a linear VAR model for All negative videos

	V^{-}_{All}		r	
$Constant$	0.682***	[16.667]	-0.036	[-1.198]
r_{t-1}	-0.050**	[-2.296]	-0.042***	[-2.645]
r_{t-2}	0.047**	[2.144]	-0.002	[-0.098]
r_{t-3}	0.007	[0.314]	0.039**	[2.451]
r_{t-4}	0.004	[0.169]	-0.010	[-0.649]
r_{t-5}	0.031	[1.434]	0.039**	[2.463]
r_{t-6}	0.023	[1.051]	0.047***	[2.936]
r_{t-7}	-0.016	[-0.751]	-0.014	[-0.900]
$V^{-}_{All,t-1}$	0.005	[0.180]	0.140***	[6.438]
$V^{-}_{All,t-2}$	-0.050*	[-1.649]	0.085***	[3.865]
$V^{-}_{All,t-3}$	-0.045	[-1.503]	0.046**	[2.081]
$V^{-}_{All,t-4}$	-0.019	[-0.643]	0.171***	[7.932]
$V^{-}_{All,t-5}$	0.092***	[3.061]	0.087***	[3.971]
$V^{-}_{All,t-6}$	0.004	[0.141]	0.057***	[2.590]
$V^{-}_{All,t-7}$	-0.040	[-1.351]	0.080***	[3.712]
Number of observations		2137		
Log likelihood:		6889.134		
BIC		-12.016		
AIC		-12.095		
R2		0.014		

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

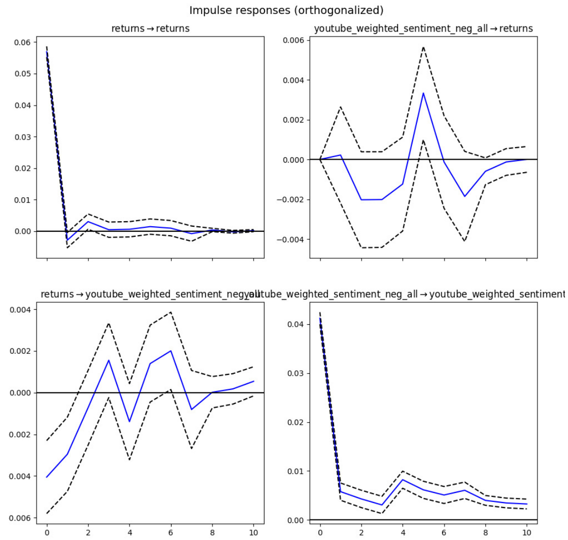


Figure 18: Orthogonalized impulse response function of bitcoin returns to the number of negative videos (all subject combined)

4 Conclusion

This paper studies whether data on investors' behavior obtained from YouTube can better forecast bitcoin returns. Unlike previous related studies, we contribute by extracting information on investors' attention and investors' sentiment using YouTube videos. We identify per-subject news related to investors' sentiment and attention that drive bitcoin returns. In addition, we use appropriate algorithms to break down the information provided by YouTube to extract dis-aggregated sentiment and attention data classified and class them per subject, offering a first study investigating the per-subject impact of investor sentiment and attention on YouTube on bitcoin returns. Further, while distinguishing attention-related YouTube videos from sentiment-related videos, we adopt a Causality analysis and VAR estimations to separately investigate the effects of attention and sentiment on bitcoin returns. We then compute mixed variables using YouTube videos concerning both attention and sentiment to simultaneously capture the combined effects of sentiment and attention on bitcoin returns, which has the advantage of reproducing further asymmetry, complexity and non-linearity in the reaction function of bitcoin returns to these factors. Our analysis yields interesting findings that help us to better explain the dynamics between bitcoin returns and investor sentiment. First, investor attention improves the forecast of Bitcoin returns even though the relationship varies with the video subject under consideration. For aggregated subjects, investor attention has a positive effect on Bitcoin returns. For dis-aggregated subjects, investor attention on videos about “Tutorials”, “Robot Trading”, and “Network Activities” have a positive effect on bitcoin returns that fades after 1-2 days. Further, bitcoin returns react negatively to attention on subjects like “Personality” or ”NFT and Metaverse”. Second, while looking at the relationship of the number of positive and negative videos on Bitcoin returns, our results show that only the number of positive videos on the subjects ”Personality” and ”Institutional and central banks” drive Bitcoin returns. After a positive shock on the number of positive videos about ”Institutional and central banks”, investors overreact to positive news which can lead to a first increase in prices followed by a correction of the market the following days. After a shock on the number of positive ”personality” videos, prices tend to decrease for the next few days. The number of negative videos on YouTube for the whole dataset and for each subject separately appears to be mainly driven by Bitcoin returns and are indeed useless to forecast Bitcoin returns. Finally, when combining sentiment and attention, we show a bidirectional causality relationship. When there is a shock on both attention and the number of negative videos published, it can trigger a feedback loop which can lead to sharp price moves. These findings have different implications for policymakers, regulators, through the identification of behavioral drivers related to YouTube (investors' sentiment and attention) that can help to better forecast the dynamics and up and down of cryptocurrencies as well as bubbles. A future extension of the present study would be to investigate the contribution of investor's attention and sentiment to forecast bitcoin volatility.

5 Disclosure statement and funding

No potential competing interest was reported by the authors.

No funding was received.

6 References

Akerlof, G. and R. Schiller (2009). "Animal Spirits - How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism".

Ante, L. (2023). "How Elon Musk's Twitter activity moves cryptocurrency markets". *Technological Forecasting and Social Change* 186, p. 122112.

Araci, D. (2019). "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". *ArXiv* abs/1908.10063.

Auer, R. and S. Claessens (2018). "Regulating cryptocurrencies: assessing market reactions", p. 15.

Baker, M. and J. Wurgler (2006). "Investor Sentiment and the Cross-section of Stock Returns". *The Journal of Finance* 61.4, pp. 1645–1680.

Bouoiyour, J. and R. Selmi (2015). "What Does Bitcoin Look Like?" *Annals of Economics and Finance* 16.2, pp. 449–492.

Ciaian, P., M. Rajcaniova, and d. Kancs (2016). "The economics of BitCoin price formation". *Applied Economics* 48.19, pp. 1799–1815.

Corbet, S., C. Larkin, B. M. Lucey, A. Meegan, and L. Yarovaya (2020). "The impact of macroeconomic news on Bitcoin returns". *The European Journal of Finance* 26.14, pp. 1396–1416.

Corbet, S., B. Lucey, A. Urquhart, and L. Yarovaya (2019). "Cryptocurrencies as a financial asset: A systematic analysis". *International Review of Financial Analysis* 62, pp. 182–199.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". *Association for Computational Linguistics*. Vol. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186.

Fama, E. F. (1965). "The Behavior of Stock-Market Prices". *The Journal of Business* 38.1, p. 34.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". *The Journal of Finance* 25.2, p. 383.

Garcia, D., C. J. Tessone, P. Mavrodiev, and N. Perony (2014). "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy". *Journal of The Royal Society Interface* 11.99, p. 20140623.

Koutmos, D. (2023). "Investor sentiment and bitcoin prices". *Review of Quantitative Finance and Accounting* 60.1, pp. 1–29.

Kristoufek, L. (2013). “BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era”. *Scientific Reports* 3.1, p. 3415.

Liu, Y. and A. Tsyvinski (2021). “Risks and Returns of Cryptocurrency”. *The Review of Financial Studies* 34.6, pp. 2689–2727.

Long, C., B. M. Lucey, and L. Yarovaya (2021). “”I Just Like the Stock” versus ”Fear and Loathing on Main Street” : The Role of Reddit Sentiment in the GameStop Short Squeeze”. *SSRN Electronic Journal*.

Mai, F., Q. Bai, and J. Shan (2015). “The Impacts of Social Media on Bitcoin Performance”. *Thirty Sixth International Conference on Information Systems, Fort Worth*, p. 16.

Nasir, M. A., T. L. D. Huynh, S. P. Nguyen, and D. Duong (2019). “Forecasting cryptocurrency returns and volume using search engines”. *Financial Innovation* 5.1, p. 2.

Ortu, M., S. Vacca, G. Destefanis, and C. Conversano (2022). “Cryptocurrency ecosystems and social media environments: An empirical analysis through Hawkes’ models and natural language processing”. *Machine Learning with Applications* 7, p. 100229.

Philippas, D., H. Rjiba, K. Guesmi, and S. Goutte (2019). “Media Attention and Bitcoin Prices”. *Finance Research Letters* 30, pp. 37–43.

Phillips, R. C. and D. Gorse (2018). “Mutual-Excitation of Cryptocurrency Market Returns and Social Media Topics”. *Proceedings of the 4th International Conference on Frontiers of Educational Technologies - ICFET ’18*. Moscow, Russian Federation: ACM Press, pp. 80–86.

Qin, M., C.-W. Su, Y. Wang, and N. M. Doran (2023). "COULD "DIGITAL GOLD" RESIST GLOBAL SUPPLY CHAIN PRESSURE?" *Technological and Economic Development of Economy* 30.1, pp. 1–21.

Qin, M., T. Wu, X. Ma, L. L. Albu, and M. Umar (2023). "Are energy consumption and carbon emission caused by Bitcoin? A novel time-varying technique". *Economic Analysis and Policy* 80, pp. 109–120.

Samuelson, P. (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly". *Industrial Management Review*, p. 41.

Shanaev, S., S. Sharma, B. Ghimire, and A. Shuraeva (2020). "Taming the blockchain beast? Regulatory implications for the cryptocurrency Market". *Research in International Business and Finance* 51, p. 101080.

Shiller, R. J. (2005). *Irrational exuberance*. 2nd ed. Currency/Doubleday. 304 pp.

Shiller, R. J. (2019). *Narrative economics: how stories go viral & drive major economic events*. Princeton: Princeton University Press. 377 pp.

Su, C.-W., S. Yang, M. Qin, and O.-R. Lobonţ (2023). "Gold vs bitcoin: Who can resist panic in the U.S.?" *Resources Policy* 85, p. 103880.

Taffler, R. (2018). "Emotional finance: investment and the unconscious". *The European Journal of Finance* 24.7, pp. 630–653.

Tetlock, P. C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". *The Journal of Finance* 62.3, pp. 1139–1168.

Tran, V. L. and T. Leirvik (2019). "Efficiency in the markets of crypto-currencies". *Finance Research Letters*, p. 101382.

Uras, N., S. Vacca, and G. Destefanis (2020). "Investigation of Mutual-Influence among Blockchain Development Communities and Cryptocurrency Price Changes". *Proceedings of the IEEE/ACM 42nd International Conference on Software En-*

gineering Workshops. ICSE '20: 42nd International Conference on Software Engineering. ICSEW'20. Seoul, Republic of Korea: Association for Computing Machinery, pp. 779–782.

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

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

Wei, W. C. and D. Koutmos (2023). "Investor Attention and Bitcoin Trading Behaviors". *Essays on Financial Analytics*. Ed. by P. Alphonse, K. Bouaiss, P. Grandin, and C. Zopounidis. Cham: Springer International Publishing, pp. 87–116.

1 Appendix

Table 28: Results of a linear VAR model for negative videos on Price Predictions

	r		$E^{-}_{PricePredictions}$	
$Constant$	0.550*	[1.868]	3.262***	[15.036]
r_{t-1}	-0.039*	[-1.766]	-0.122***	[-7.431]
r_{t-2}	0.047**	[2.106]	-0.009	[-0.541]
r_{t-3}	0.008	[0.347]	-0.020	[-1.234]
r_{t-4}	0.015	[0.672]	0.006	[0.351]
r_{t-5}	0.024	[1.079]	0.032*	[1.924]
r_{t-6}	0.022	[0.991]	0.016	[0.938]
r_{t-7}	-0.018	[-0.819]	0.012	[0.749]
r_{t-8}	-0.027	[-1.194]	0.014	[0.875]
r_{t-9}	0.017	[0.757]	0.009	[0.524]
r_{t-10}	0.044**	[1.968]	0.020	[1.204]
r_{t-11}	-0.017	[-0.741]	-0.000	[-0.007]
r_{t-12}	0.012	[0.542]	0.013	[0.791]
r_{t-13}	0.011	[0.479]	0.024	[1.481]
r_{t-14}	-0.013	[-0.589]	0.005	[0.293]
r_{t-15}	0.011	[0.510]	0.035**	[2.127]
r_{t-16}	-0.032	[-1.405]	0.005	[0.312]
r_{t-17}	0.030	[1.347]	0.007	[0.426]
r_{t-18}	0.004	[0.175]	-0.012	[-0.710]
r_{t-19}	0.012	[0.516]	0.035**	[2.145]
r_{t-20}	0.038*	[1.687]	0.004	[0.264]
$E^{-}_{PricePredictions,t-1}$	0.021	[0.698]	-0.525***	[-23.604]
$E^{-}_{PricePredictions,t-2}$	0.013	[0.387]	-0.487***	[-19.507]
$E^{-}_{PricePredictions,t-3}$	-0.024	[-0.648]	-0.438***	[-16.122]
$E^{-}_{PricePredictions,t-4}$	0.012	[0.312]	-0.315***	[-10.964]
$E^{-}_{PricePredictions,t-5}$	-0.001	[-0.025]	-0.300***	[-10.183]
$E^{-}_{PricePredictions,t-6}$	0.005	[0.120]	-0.252***	[-8.357]
$E^{-}_{PricePredictions,t-7}$	0.050	[1.214]	-0.231***	[-7.590]
$E^{-}_{PricePredictions,t-8}$	-0.000	[-0.004]	-0.228***	[-7.456]
$E^{-}_{PricePredictions,t-9}$	0.070*	[1.687]	-0.200***	[-6.505]
$E^{-}_{PricePredictions,t-10}$	-0.050	[-1.206]	-0.173***	[-5.646]
$E^{-}_{PricePredictions,t-11}$	0.016	[0.379]	-0.219***	[-7.146]
$E^{-}_{PricePredictions,t-12}$	0.027	[0.654]	-0.195***	[-6.361]
$E^{-}_{PricePredictions,t-13}$	0.002	[0.044]	-0.138***	[-4.534]
$E^{-}_{PricePredictions,t-14}$	0.061	[1.480]	-0.164***	[-5.433]
$E^{-}_{PricePredictions,t-15}$	0.019	[0.477]	-0.073**	[-2.459]
$E^{-}_{PricePredictions,t-16}$	-0.020	[-0.502]	-0.129***	[-4.436]
$E^{-}_{PricePredictions,t-17}$	-0.041	[-1.061]	-0.102***	[-3.631]
$E^{-}_{PricePredictions,t-18}$	-0.020	[-0.550]	-0.092***	[-3.435]
$E^{-}_{PricePredictions,t-19}$	-0.021	[-0.647]	-0.103***	[-4.232]
$E^{-}_{PricePredictions,t-20}$	-0.029	[-1.011]	0.018	[0.833]
Number of observations			2124	
Log likelihood:			6874.045	
BIC			-11.853	
AIC			-12.071	
R2			0.027	

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels

Table 29: Results of a linear VAR model for negative videos on "Network Activities"

	r		$E_{NetworkActivities}$	
$Constant$	0.750***	[48.341]	0.038***	[3.260]
r_{t-1}	-0.051**	[-2.336]	-0.044***	[-2.669]
$E^{-}_{NetworkActivities,t-1}$	0.001	[0.044]	0.198***	[9.354]
Number of observations	2143			
Log likelihood:	6786.224			
BIC	-11.988			
AIC	-12.004			
R2	0.003			

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

Table 30: Results of a linear VAR model for negative videos on "Institutional And Central banks"

	r		$E^{-}_{InstitutionalAndCentralbanks}$	
$Constant$	0.708***	[25.726]	0.098***	[3.259]
r_{t-1}	-0.049**	[-2.242]	-0.095***	[-4.017]
r_{t-2}	0.047**	[2.174]	-0.043*	[-1.792]
r_{t-3}	0.010	[0.461]	0.013	[0.536]
$E^{-}_{InstitutionalAndCentralbanks,t-1}$	0.006	[0.280]	0.101***	[4.650]
$E^{-}_{InstitutionalAndCentralbanks,t-2}$	-0.006	[-0.327]	0.137***	[6.366]
$E^{-}_{InstitutionalAndCentralbanks,t-3}$	-0.018	[-0.913]	0.098***	[4.548]
Number of observations	2141.000			
Log likelihood:	6001.289			
BIC	-11.232			
AIC	-11.269			
R2	0.005			

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

Table 31: Results of a linear VAR model for the number of negative videos on "Regulation"

	r		$E^{-}_{Regulation}$	
$Constant$	0.714***	[31.726]	0.089***	[4.590]
r_{t-1}	-0.048**	[-2.224]	-0.092***	[-4.955]
r_{t-2}	0.047**	[2.176]	-0.025	[-1.308]
$E^{-}_{Regulation,t-1}$	0.008	[0.299]	0.402***	[17.799]
$E^{-}_{Regulation,t-2}$	-0.007	[-0.270]	0.066***	[2.965]
Number of observations	2142			
Log likelihood:	6516.109			
BIC	-11.724			
AIC	-11.751			
R2	0.005			

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

Table 32: Results of a linear VAR model for negative videos on “Tutorials”

	r		$E^{-}_{Tutorials}$	
$Constant$	0.712***	[25.825]	0.102***	[3.478]
r_{t-1}	-0.050**	[-2.309]	-0.132***	[-5.710]
r_{t-2}	0.044**	[1.993]	0.019	[0.807]
r_{t-3}	0.010	[0.460]	-0.020	[-0.849]
$E^{-}_{Tutorials,t-1}$	-0.018	[-0.908]	0.151***	[6.988]
$E^{-}_{Tutorials,t-2}$	0.002	[0.114]	0.129***	[5.954]
$E^{-}_{Tutorials,t-3}$	-0.030	[-1.510]	0.082***	[3.812]
Number of observations	2141			
Log likelihood:	6067.206			
BIC	-11.293			
AIC	-11.330			
R2	0.007			

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters.

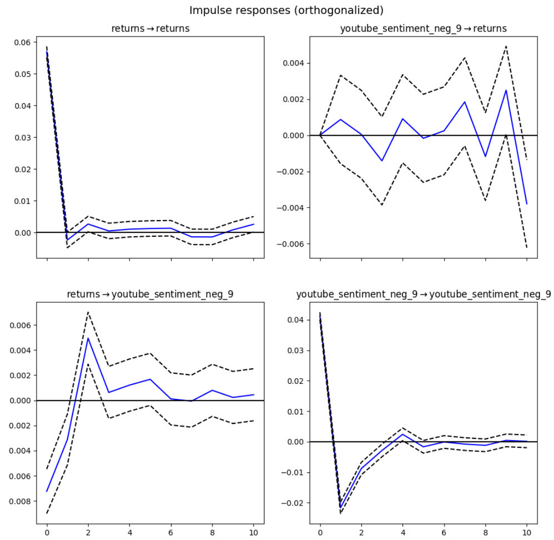


Figure 19: Orthogonalized Impulse response function of returns on negative sentiment about “Price Predictions”

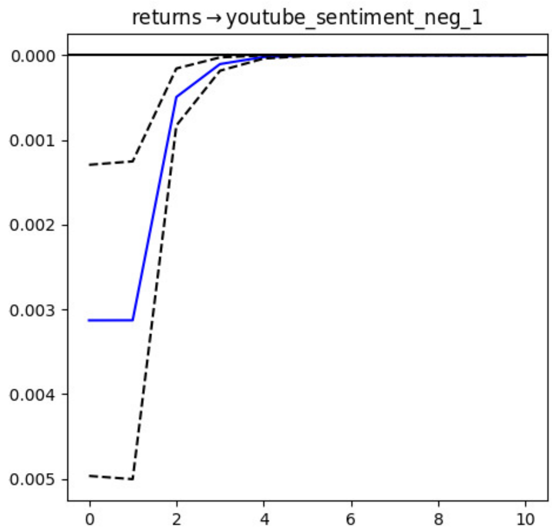


Figure 20: Orthogonalized Impulse response function of returns on sentiment about “Network Activities”

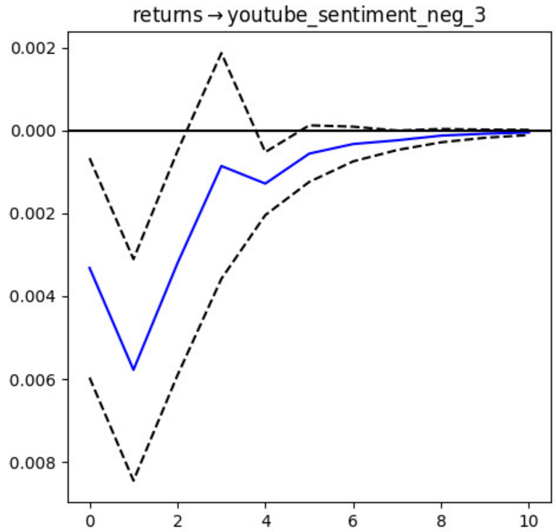


Figure 21: Orthogonalized Impulse response function of returns on negative sentiment about “Institutional And Central banks”

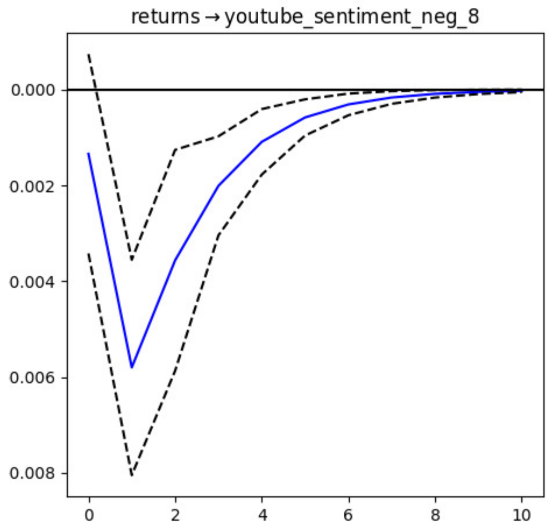


Figure 22: Orthogonalized impulse response function of returns on negative sentiment about “Regulation”

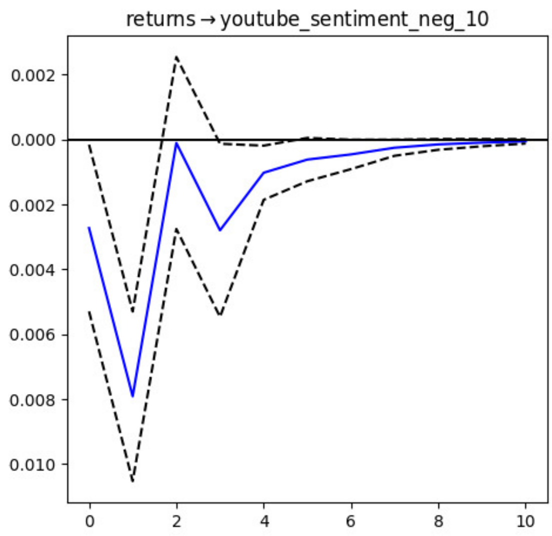


Figure 23: Orthogonalized impulse response function of returns on negative sentiment about “Tutorials”

Listing 1: Youtube video, topic classification sample code

```
import re
from code.datas.BaseDatas import BaseDatas


class SentimentClassifierGuided(BaseDatas):


def get_voca_id( self ):
''' Function to return dictionnary with associated keywords and the associate
    id for the topic '''
dico = {
'price_predictions': ['bulls', 'bears', 'rally', 'breakout', 'high', 'low', '
    recover', 'predict', 'volume', 'price', 'review', 'analysis', 'bullish', '
    bearish', 'resistance', 'support', 'level','bull', 'bear', 'run', 'rally',
```

'forecast', 'prediction', 'expect', 'hit', 'price', 'market', 'crash', 'fall',
'raise', 'break', 'technical', 'decline', 'prediction', 'uptrend', '
downtrend', 'dump', 'volatility', 'pumping', 'pattern', 'predict', 'charts', '
ath', "pump", "indicator", "indicators", 'plunges', 'plunge'],
'nft_metaverse': ['nfts', 'axies', 'sport', 'nft', 'metaverse', 'opensea', '
game', 'eternity', 'artist'],
'ico': ['ico', 'icos', 'launch', 'funding', 'venture', 'capital', 'vc', 'fund',
'participate'],
'bot': ['bot', 'robot', 'bots', 'robots'],
'adoption': ['own', 'partnership', 'adoption', 'accepted', 'partners'],
'network_activities': ['mine', 'supply', 'farm', 'parachains', 'mining', '
addresses', 'miner', 'hashrate', 'energy', 'gpu', 'block', 'pool', 'miners', '
pools'],
'institutional_central_bank': ["tesla", 'microstrategy', 'institutional', '
presidential', 'bank', 'cbdc', 'banks', 'tokens', 'fed', 'goldman', 'sachs',
'hsbc', 'jpmorgan', "visa", "mastercard", "paypal"],
'regulation': ["sec", "amf", 'banned', "court", 'lawsuit', 'authority', '
regulation', 'regulate', 'senator', 'legal', 'ban', 'regulates', 'tax', '
france', 'europe', 'federal', 'china', 'malta', 'watchdog', "regulator", "
regulators"],
'personality': ['ceo', 'michael', 'burry', 'elon', 'musk', 'vitalik', 'butterin
', 'saylor', 'elonmusk', 'warren', 'buffett', 'keiser', 'cuban', 'biden', '
kiyosaki', 'gates', 'rickards', 'thiel'],

```
'hack': ['scam', 'phished', 'hack', 'hacks', 'pirate', 'attack', 'attacks', '
    steal', 'stole', 'stolen','stoles', 'hacker', 'theft'],
'explanation': ['explains', 'explained', 'how', 'explain', 'understanding', '
    guide', 'tutorial'],
}
lda_topics = {
    'explanation': 10,
    'price_predictions': 9,
    'regulation': 8,
    'bot': 7,
    'ico': 6,
    'personality': 5,
    'nft_metaverse': 4,
    'institutional_central_bank': 3,
    'adoption': 2,
    'network_activities': 1,
    'hack': 0,
}

return dico, lda_topics


def run(self):
'''Assign a topic (ie. a topic code) to each Youtube video (if detected or
    assign -1)'''
```

```
dico, lda_topics = self.get_voca_id()
REPLACE_BY_SPACE_RE = re.compile('[/(){}\[\]\|@,;]')
BAD_SYMBOLS_RE = re.compile('[^0-9a-z #+_]')

sql = "SELECT _id, title \
       FROM youtube_btcusdt y \
       WHERE y.title is not null "
self.mycursor.execute(sql)
res = self.mycursor.fetchall()
addlist = list()
for value in res:
    id = value[0]
    title = value[1].lower()
    title = REPLACE_BY_SPACE_RE.sub(' ', title)
    title = BAD_SYMBOLS_RE.sub('', title)
    title = title.split(" ")
    subjects = list()
    topic_has_subject = False
    for itemo in dico.items():
        subject = itemo[0]
        keys = itemo[1]
        for key in keys:
            if key in title:
                if lda_topics[subject] not in subjects:
```

```
                subjects.append(lda_topics[subject])
            topic_has_subject = True

    if not topic_has_subject:
        addlist.append("('" + id + "',-1)")
    else:
        for sub in subjects:
            addlist.append("('" + id + "'," + str(sub) + ")")

for i in range(0, len(addlist), 500):
    print('page ' + str(i))
    to_save = addlist[i:i + 500]
    sql = "INSERT INTO youtube_btcusdt_lda (id_youtube,lda_classe)
        VALUES " + str(",".join(to_save))
    self.mycursor.execute(sql)
    self.mydb.commit()
```

Chapter 3: Does Blockchain Competent Investor's Sentiment Drive Bitcoin Volatility ? Further Evidence from Artificial Intelligence Tests.

David Bourghelle, Pierre Fay, Fredj Jawadi

IAE Lille University School of Management

Abstract

This paper extends the behavioral finance literature on the relationship between investor sentiment and volatility in two ways. On the one hand, we propose a natural language-processing method to measure Blockchain-competent investor sentiment while extracting information from Reddit forums. On the other hand, we extend a risk-free measure of volatility based on daily realized volatility (HARRV model: Heterogeneous Autoregressive Realized Volatility) to a nonlinear context using a Markov switching approach over the period 2018-2023, and we investigate the effect of Blockchain-competent investor sentiment on bitcoin volatility. Furthermore, using the state-of-the-art deep learning technique, FinBERT (Financial Bidirectional Encoder Representations from Transformers) to detect sentiment, we test whether Blockchain-Competent (BC) and Non-Blockchain-Competent (NBC) investors' comments on Reddit, and thus sentiment, have a different impact on Bitcoin's volatility. Our results show two key findings. First, we find that Blockchain-competent investor sentiment has a significant and nonlinear effect on bitcoin volatility. Second, the distinction between positive and negative sentiment, as well as Blockchain-competent investor sentiment and Blockchain-non-competent investor sentiment can help to improve bitcoin volatility forecasts.

Keywords: Artificial Intelligence, Sentiment analysis, Bitcoin volatility, HARRV Model, HAR-RV Markov Switching model, Forecast.

JEL: C2, F10, G10.

1 Introduction

Blockchain has been presented as a technological revolution, frequently compared to the Internet in the late 1990s. When the Internet was born in 1989, it was primarily used in few sectors, and it took nearly 10 years to develop user-friendly interfaces, leading to mass adoption and the introduction of new functionalities like email, e-commerce, and social media. In the 1990s, the adoption of Internet and these new Information and Communication Technologies boosted stock markets to unprecedented levels, significantly exceeding what could be justified by company earnings and fundamentals (Shiller 2000). This led to high overvaluation on most stock markets and significant volatility excess, followed by the dot-com bubble in 2000 and a major collapse. Shiller (2000) explains these stock market dynamics by extensive investor overconfidence observed after the 1990s, known as an irrational exuberance phenomenon.

During this period of high uncertainty, hugely optimistic earnings growth prospects were advanced to justify valuation levels reached by the shares of new-economy startups[1]. More generally, the rapid chain of several financial bubbles since the early 2000s is symptomatic of this logic. Since the early 1990s, there has been growing empirical evidence showing that the stock market is driven by investor psychology (Daniel, Hirshleifer, and Teoh 2002).

For cryptocurrencies, this observation is even more relevant, especially since no fundamental anchor value can really be calculated. Accordingly, hypotheses of speculative bubbles and crashes are even more significant and frequent[2].

[1] A. B. Perkins and M. C. Perkins (1999) and Ofek and Richardson (2003)
[2] See Chowdhury, Damianov, and Elsayed (2022)

In the prolific behavioral financial literature, hypotheses of market efficiency and rationality are strongly criticized. Alternatively, financial asset price patterns tend to be associated with cognitive bias (Tversky and Kahneman 1974) and market irrationality (Shiller 2005).

For example, Kumar (2009) showed that when an asset is hard-to-value, as for cryptocurrencies, it is subject to strong behavioral bias. For a new technological asset, even for a rational investor, these patterns may be explained by the time-varying uncertainty about its future productivity (Pástor and Veronesi 2009) or its prospective value (Lee, Li, and Zheng 2020). In fact, the technology improves over time, and its added value becomes clearer to investors who are more susceptible to buying or selling it, consequently producing price patterns.

To better understand the uncertainties that can cause these price patterns, we need to acknowledge the multiple specificities of cryptocurrencies. First, they are naturally highly volatile (Yermack 2013; Kristoufek 2023) and there is no consensus on a pricing methodology. Second, the liquidity of cryptocurrencies is limited as it is a relatively recent market with few buyers and sellers (Trimborn, Li, and Härdle 2019). Third, since cryptocurrency regulation is yet to be clarified, additional news about cryptocurrency regulation could significantly impact their prices (Auer and Claessens 2018). Fourth, the technology behind cryptocurrencies is still being developed. Accordingly, investors' vision of the future value of these assets evolves with the different developments occurring on the market (new technological advances, hacks, forks,Yermack 2013). Consequently, there is a high degree of uncertainty around cryptocurrencies that fuels discussion between investors, and some of these discussions involve investor sentiment.

A growing financial literature has highlighted the role of sentiment on asset pricing and market volatility (Audrino, Sigrist, and Ballinari 2020). Indeed, the effect of sentiment has been widely studied in the traditional financial market literature. D. Long et al. (1990) defined noise traders as individuals who make investment decisions based on factors other than fundamental analyses (emotional bias, speculation, or herd mentality). These noise traders can collectively influence asset prices, pushing them to deviate from their intrinsic values. Baker and Wurgler (2006) showed that when market sentiment is high, equities tend to be overvalued and their prices subsequently undergo a correction. Conversely, when sentiment is low, stocks tend to be undervalued, which then generates positive returns. In their seminal paper, Bollen, Mao, and Zeng (2011) used Twitter to extract investor sentiment, showing a strong link between investor sentiment on Twitter and stock market movements. Audrino, Sigrist, and Ballinari (2020) showed that the informativeness of sentiment on Twitter and StockTwits data for future volatility is particularly interesting when an unexpected announcement or breaking news occurs.

Regarding cryptocurrencies, Mao, Counts, and Bollen (2011) found that implied volatility has a statistically significant relationship with Twitter sentiment. Polasik et al. (2015) showed that the tone of online newspaper articles is positively correlated with Bitcoin price. Bourghelle, Jawadi, and Rozin (2022b) demonstrated the key role of sentiment and collective emotions when predicting bitcoin volatility using the Fear and Greed crypto index, while Sapkota (2022) showed that sentiment resulting from news has a long-term effect on bitcoin volatility. According to Öztürk and Bilgiç (2022), sentiment on the most influential accounts' tweets can be used to predict Bitcoin returns.

That is, while sentiment appears relevant, its measure is problematic, as it is not observed. Several proxies have been adopted, with the papers' conclusions dependent on the measure or proxies used. However, it is worth to note that among the various Internet platforms on which investors exchange information, Reddit is the most relevant. In fact, Reddit is an online forum divided into thematic communities. Reddit has the advantage of a massive user base that launches numerous financial discussions. Compared to most social media platforms, discussions on Reddit can be less restrictive and therefore more in-depth. Reddit is used by S. (Long et al. (2023) to show the important role the site played in transmitting positive sentiment during the bullish episodes of the famous GameStop rally. The same applies to cryptocurrencies. The role of Reddit sentiment on bitcoin returns and volatility has been demonstrated by several studies (Phillips and Gorse 2017; Bowden and Gemayel 2022). For example,Ortu et al. (2022) used VADER on Reddit posts to show that some thematic occurrences and associated sentiments can forecast price dynamics.

Sentiment analysis has traditionally been used in the behavioral finance literature to extract sentiment derived from various textual data such as news articles, social media posts, or even financial reports. In practice, sentiment analysis relies on algorithms to analyze natural language and identify sentiment indicators. Since the work of Tetlock (2007), the dictionary-based approach has historically been used to measure sentiment.

To understand the price dynamics of financial assets and cryptocurrencies, natural language processing (NLP) may be used, which is a field of artificial intelligence designed to understand human language. Among various natural language processing techniques, sentiment analysis aims to identify and extract the sentiments expressed in a text. It has a crucial role in a variety of applications, including marketing and the detection of hate speech. Various dictionaries have been used, as in Loughran and Mcdonald (2011) or Hutto and Gilbert (2015). Recently, thanks to the development of deep learning models, NLP has improved, and new techniques such as FinBERT (Financial Bidirectional Encoder Representations from Transformers) proposed by Araci (2019) have been adopted in the financial literature. This new technique has the advantage of offering better sentiment recognition performance as it allows a word to have diverse interpretations depending on the context.

While several previous papers have used aggregate measures of sentiment and highlighted the negative effect of sentiment on bitcoin's volatility (Oad Rajput, I. A. Soomro, and N. A. Soomro 2022), in reality, not all investors develop the same strategies or react in the same way to an event or piece of news. For bitcoin, Lee, Li, and Zheng (2020) showed the coexistence of speculators and technology-savvy investors who invest for very different reasons. Indeed, while speculators tend to follow a dynamic trading strategy, tech-savvy investors buy (resp. sell) when the price falls below (resp. rises above) a forward-looking value that is defined by supply and demand factors.

In this paper, we extend the previous literature on the relationship between bitcoin sentiment and volatility by proposing a NLP method to measure Blockchain-competent investors on Reddit forums, and we investigate their effect on price dynamics. Blockchain-competent investors are investors with technical knowledge of Blockchain technology, which underpins bitcoin. We argue that given their significant understanding of the technology, we expect them to react more discerningly to news on the market and to be less affected by other investors' sentiment.

This paper therefore tests whether Blockchain-competent investor sentiment on Reddit does drive bitcoin volatility. Further, we investigate whether the distinction between positive/negative sentiment and Blockchain-Competent (BC)/Non-Blockchain-Competent (NBC) investors can improve bitcoin volatility forecasting.

The remainder of this paper is organized into four sections. Section 2 presents the data and briefly recalls the methodology. Section 3 discusses the main empirical results. Section 4 concludes.

2 Data and Methodology

Our study uses daily data covering the period between August 1, 2018 and December 31, 2023 (1949 observations), which is appropriate as it contains several episodes of bubbles, crashes, and remarkable events (COVID-19, ban of miners from China, Terra-Luna cryptocurrency hack, FTX bankruptcy in November 2022, etc.).

Our data includes bitcoin prices collected from Binance API, the leading cryptocurrency exchange platform. We computed bitcoin returns as the logarithm difference of the closing prices.

$$r_t = \log(P_t) - \log(P_{t-1}) \quad \text{(A.1)}$$

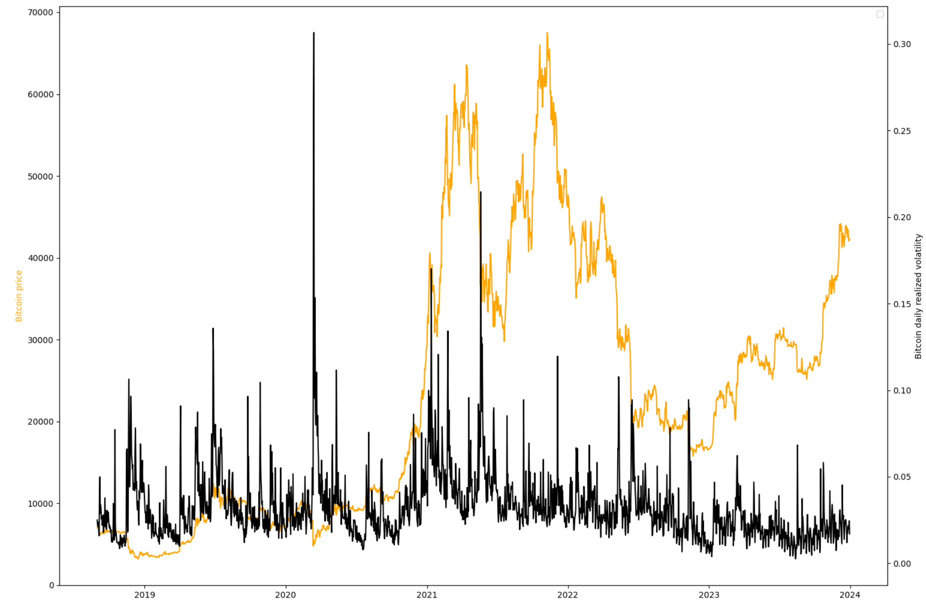
Next, we measure bitcoin volatility. Rather than relying on a volatility proxy using a parametric model such as a GARCH model, we used a risk-free model proxy, relying on the realized volatility (RV) computed as:

$$RV_t = \sqrt{\sum_{i=t-n}^{i=t} r_t^2} \tag{A.2}$$

Where: RV_t denotes the realized volatility at t. n is the number of 1-minute intervals (1440 for a one-day period, 10080 for a one-week period, 43200 for a one-month period). r_t is the log return for time t. Hereafter, we note RV_h for hourly realized volatility (for the last 60 minutes), RV_d for daily realized volatility (the last 24 hours), RV_w for weekly realized volatility (the last 7 days), and RV_M for monthly realized volatility (the last 30 days).

To provide an overview of the data, we report the bitcoin closing price and its realized volatility in Figure 24. Figure 24 shows that the bitcoin price exhibits several sharp changes, yielding further evidence of different volatility regimes.

Figure 24: Daily Bitcoin price and realized volatility



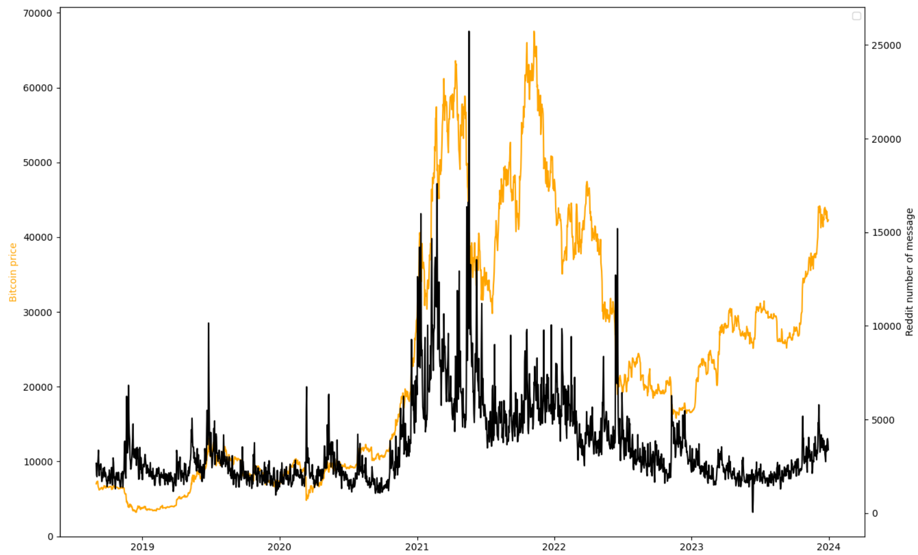
As for sentiment, we extracted investor sentiment from the 6,995,901 Reddit comments on the "Bitcoin" Reddit channel and 510,303 Reddit submissions during the study period. Reddit historical data was gathered through Reddit API.

Reddit is an online forum split into topic-specific communities. It has a massive user base and includes many financial discussions. Compared to most social media platforms, Reddit discussions can be more in-depth and less restricted. This is particularly useful as it allows us to capture a wider range of emotions and opinions than other social media. Messages on X (formerly Twitter), for example, even in the financial literature, are limited to 280 characters.

Figure 25 showing the daily number of Reddit message during the study period yields two main observations. First, users were especially active during the bull run (end of 2020 to first half of 2021). Second, during this period, major institutions such as Tesla and Mastercard announced investments in Bitcoin, signaling growing mainstream adoption.

We also extracted sentiment from 105,130 news headlines from popular crypto specialized websites[3].

Figure 25: Number of Reddit comments on VS Bitcoin price



2.1 Measuring Investor Sentiment

FinBERT proposed by Huang, Wang, and Yang (2023) is used in this paper to extract sentiment from text. In particular, we applied this technique to extract sentiment from news headlines and Reddit comments. It should be noted that FinBERT outperforms other sentiment classification methods such as Natural Language Processing algorithms of Loughran and Mcdonald (2011) using a dictionary approach. FinBERT, in fact, has the advantage of analyzing extracted words with

[3]The news headlines were extracted from the RSS feed of bitcoinist.com, cointelegraph.com, newsbtc.com, coindesk.com and u.today during the study period

regard to the context. We thus used FinBERT to classify our Reddit comments into one of the three following sentiments: Positive (+1), Neutral (0), or Negative (-1). In this way, FinBERT extracted a text and provided a sentiment signal.

For illustration, examples of sentiment classification are reported in Table 33.

Table 33: Examples of Sentiment analysis on Reddit comments using FinBERT

Comment	Sentiment
This is good for Bitcoin	positive
It's not exactly positive at the moment.	negative
How much bitcoin on those addresses in total?	neutral

Note: The table above shows examples of FinBERT sentiment analysis of Reddit comments extracted from our database.

We obtained 582,180 positive and 497,881 negative comments for Bitcoin, reflecting the impressive expression of sentiment and emotion via Reddit.

We discuss, hereafter, the process based on AI techniques by which FinBERT analyzes language and computes the sentiment indicator. FinBERT works by combining two techniques in particular: the BERT algorithm and a Neural Network Sentiment Classifier. First, BERT (Bidirectional Encoder Representations from Transformers)converts the text into a vector representation called "Embedding". This vector representation analyzes the words and the context of the sentence. The embedding is then fed into a trained Neural Network Sentiment classifier to make the final sentiment classification, based on calculating the probability of each sentiment class: positive, negative, or neutral. In practice, we assigned the class with the greater probability to the item (news or Reddit comments).

Figure 26: FinBERT process

First step: Converting Text with the BERT Algorithm

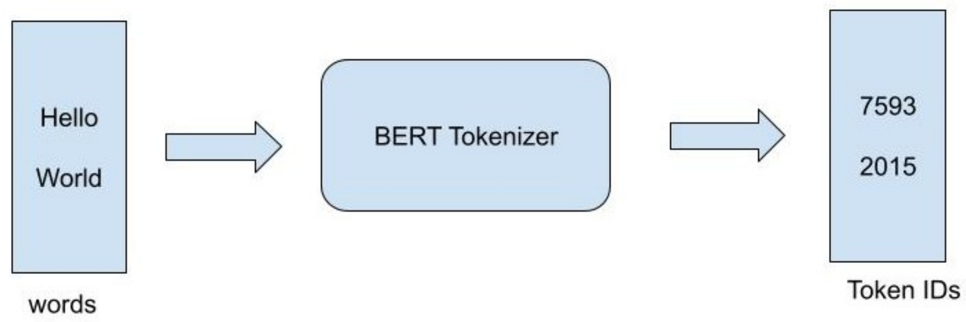
BERT (Bidirectional Encoder Representations from Transformers) is a powerful tool for converting text into embedding vectors. Embedding vectors are numerical representations of sentences that are not directly understandable by a human. The BERT algorithm captures the meaning and the relationships between words in a high-dimensional space. In fact, the similarities and relationships between words are encoded through the relative position of vectors in space. From this perspective, "car" and "bus" for example might have more cosine similarity than "car" and "cheese."

Figure 27: Generating the Embedding with BERT

BERT has been pre-trained by Google from a huge amount of text data. In fact, unlike alternative embedding methods (Glove, Word2vec), BERT representations are robust to further changes in the context of the text under consideration. Indeed, BERT analyzes the entire sentence at once, considering how each word relates to the others. This process is important as it helps to capture the nuances of language and further changes in word meaning depending on the context. We discuss, hereafter the three main steps related to the BERT algorithm.

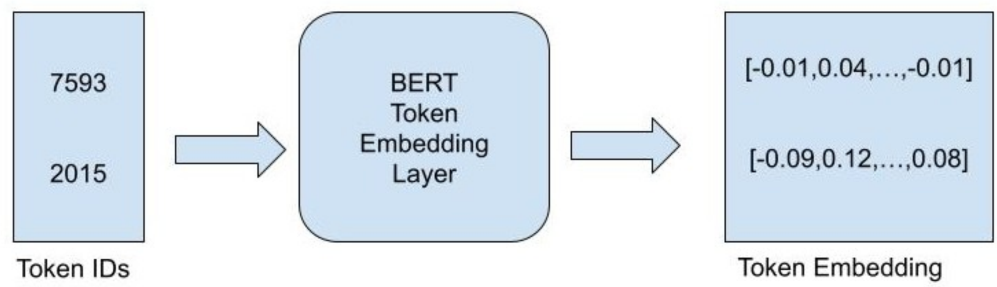
In the first step, words are converted into Token IDs. In particular, each word is associated with a unique integer.

Figure 28: Example of Changing Words to Token IDs



In the second step, the token IDs are converted into token embedding vectors, position embedding vectors, and segment embedding vectors.

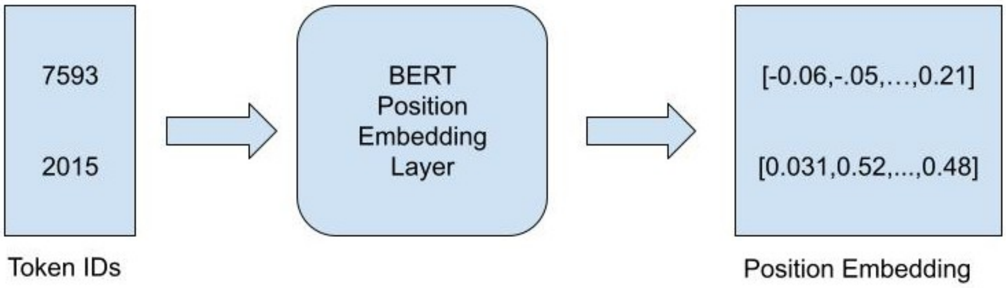
Figure 29: Token Embedding Process



In the token embedding layer, words are transformed into vectors. Two words are considered as similar by BERT if they show high cosine similarity. These embedding vector values are learned when the BERT model is trained by Google. Google uses 768-dimensional space for embedding. Each dimensional space represents a semantic category. Each value of an embedding vector is defined as the probability of the word belonging to one of the 768 semantic spaces.

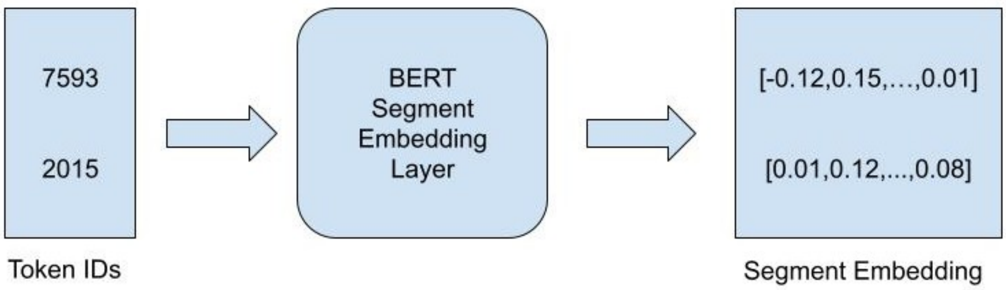
At the same time, the Token IDs are also passed to a Position Embedding Layer, where the main task is to represent the position of each token in the input sequence.

Figure 30: Position Embedding Process



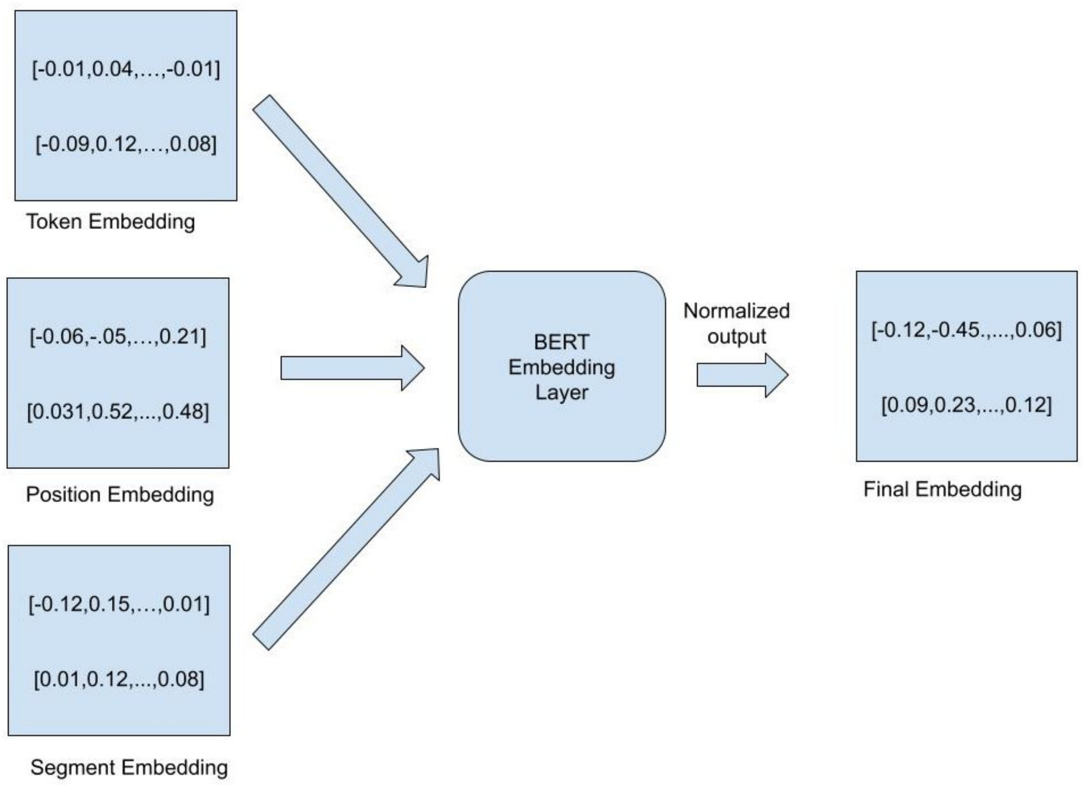
Token IDs are also fed into a segment embedding layer. Segment embedding helps to identify the order between sequences. Tokens of the first sentence have an embedding value of 0, whereas tokens of the second sentence have an embedding value of 1.

Figure 31: Segment embedding process



In the final step, the three embeddings are passed into the BERT Embedding Layer, which is responsible for computing the sum of embedding vectors. The result is normalized to produce a final embedding vector containing all information about the encoded sentence.

Figure 32: Final Embedding Process



This final embedding gives BERT a different representation of a word depending on its place in the text and the surrounding words. For more details on the training process and how embedding works, see Devlin et al. (2019).

For a better understanding of financial vocabulary, BERT was trained on Analyst Reports (1.1B tokens), Corporate Reports (2.5B tokens), and Earnings Call

Transcripts (1.3B tokens) using two strategies from the original BERT paper.

The first strategy is Masked Language Modeling (MLM) where FinBERT is trained to retrieve the masked words of a sentence containing random masked words. The second strategy is Next Sentence Prediction (NSP). FinBERT takes two sentences and is trained to recognize the order of the two sentences. This pre-training phase helps the model to understand the language syntax and semantics.

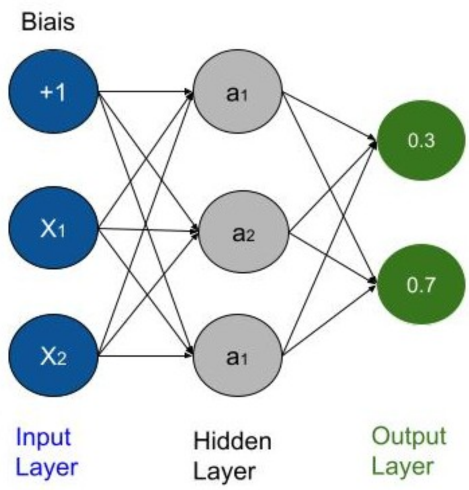
Second Step: Sentiment classification

BERT is then fine-tuned by adding a neural network classifier on top of its embedding layer to become FinBERT. The final BERT embedding vectors obtained in the previous step are fed into a neural network classifier to obtain the "positive", "neutral" or "negative" output.

Neural Networks are a type of Artificial Intelligence (AI) algorithm inspired by the way the human brain works. They are powerful tools for various applications such as image recognition, speech recognition, translation, and recommendation systems. A Neural Network consists of multiple layers of artificial neurons. In each layer, each neuron (called perceptron) is connected to all the other neurons in the next layer with parameters called 'weights'.

A Neural Network consists of multiple layers, each layer consisting of multiple artificial neurons. An example of a simple neural network with 3 input features is shown in Figure 10. The input layer (x1, x2, x3) represents the input features. Neurons in the hidden layer transform the values from the previous layer with a weighted linear summation (w1x1 + w2x2 + w3x3). One Neural Network can have multiple hidden layers. The result is then sent to a nonlinear activation function like a sigmoid function in the case of a classifier. The output layer receives the values from the hidden layer and transforms them into output values. In Figure 10, there are two neurons corresponding to two labels. Each neuron will output the probability of being in its respective label.

Figure 33: Example of a Two-label Neural Network Classifier with one Hidden Layer



As for a Machine Learning model, neural networks should be trained to execute a specific task (classification, regression). To this end, we give them input data and output targets. When training a neural network, the artificial neurons try to reduce the error by adjusting their weights parameters to the training data (w1, w2, w3 are the arrows linked to the neurons in 33). This process is called back-propagation.

FinBERT uses a classifier trained with 10,000 sentences from analyst reports. Following Huang, Wang, and Yang (2023), sentences are manually labeled as positive, negative or neutral. During the training process, the classifier receives the final embedding as input and the sentiment as the output target. This allows the network's artificial neurons to adjust their weights parameters to the training data to minimize errors. Once trained, the classifier will calculate the probability that a set of inputs matches each label.

The probability of being in one of the three sentiment categories is calculated as:

$$p(S|V) = \frac{e^{W_S \times V}}{\sum e^{W_S \times V}} \tag{A.3}$$

where, V is the embedding vector, S is the category label (ie. the sentiment label which can be neutral, positive, or negative).

In practice, we assigned the label with the highest probability to the text when using FinBERT.

2.2 Identification of Blockchain-Competent Users

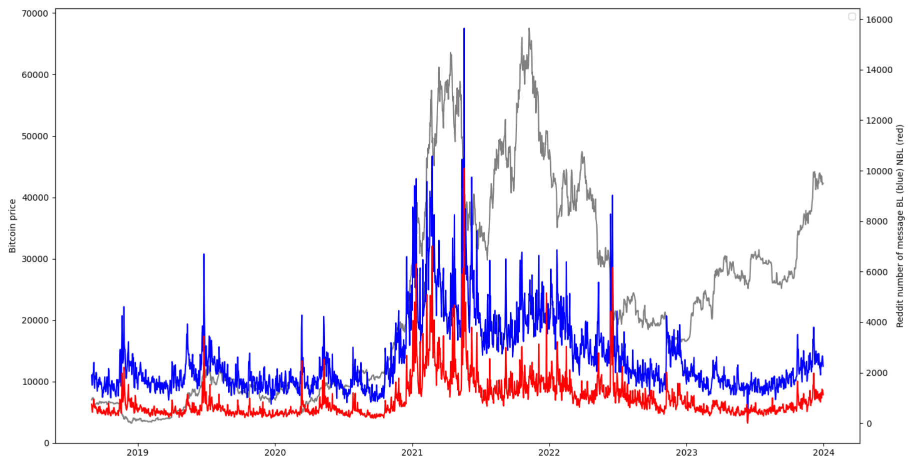
It is important to detect blockchain-literacy as a user level and not a message level because not every BC user's message is technical. Our measurement of sentiment should take all user's messages into account, and not only technical messages.

To detect blockchain-competent users, we manually constructed a dictionary of blockchain technical words extracted from the nouns most often used in Reddit submissions and Reddit comments.

We, then, flagged a user in our database as blockchain-competent if he or she used one of the words from the dictionary in at least one submission or comment in the database. When calculating the number of positive and negative comments for BC and NBC users, this allowed us to filter whether the author is blockchain-competent (BC) or not (NBC) for each comment.

We identified 591,645 Reddit users. Among these users, only 63,349 (10.7%) were identified as blockchain-competent (BC). BC users posted 5,108,784 comments, while non-blockchain-competent (NBC) users posted only 1,887,117 comments. This result is interesting as it shows that even if Reddit BC users account for just 10.7% of the Reddit Bitcoin channel users, they are responsible for 73% of the comments, which means that this population is constantly active on the channel. Figure 34 shows the number of comments from NBC (in blue) and BC (in red) users, along with the price of Bitcoin (in gray) during the study period. We can observe that BC users are more active. Both users seem to react at the same time, but the BC users' reactions appear to be amplified. That is, for both users, the reactions are more volatile round 2021.

Figure 34: Number of messages from BC and NBC users vs. Bitcoin price



With regard to comment sentiment, BC and NBC users post virtually almost the same percentage of positive and negative messages. In fact, BC users have 7.96% of positive messages, 84.9% of neutral messages, and 7.14% of negative messages, and NBC users have 9.28% of positive messages, 83.66% of neutral messages, and 7.06% of negative messages.

2.3 Methodology

We investigated the link between Blockchain Competency and Bitcoin Volatility. We identified Blockchain Competent users (BC) and Non Blockchain Competent Users (NBC) and extracted sentiment from their Reddit messages to obtain two daily variables of sentiment (number of positive and negative messages) for each group of users. We then analyzed the data in three steps. First, we tested whether splitting sentiment into BC/NBC components to capture different effects for each group helped us to improve the volatility modeling. Second, we tested whether our specification helped to improve the forecast of volatility. Finally, we tested whether sensitivity of the sentiment of BC/NBC users to other comments and news articles is the same for each group.

In practice, various models were used to forecast Bitcoin volatility. GARCH models are popular and have been extensively used in the literature (Dyhrberg 2016, Katsiampa 2017, Köchling, Schmidtke, and Posch 2020).The HAR-RV model proposed by Corsi (2009) is another popular volatility model, well-adapted to Bitcoin volatility modelling and forecasting (Sapkota 2022, Yu 2019, Yi, He, and Zhang 2022). Bergsli et al. 2022 show that the HAR-RV model outperforms GARCH models when forecasting Bitcoin realized volatility as it captures the long memory effect and the volatility clustering phenomenon present in cryptocurrency markets (see Jawadi et al. (2019) for extensions of the HAR model). The main advantage of using the HAR-RV model rather than a GARCH model is that it allows us to capture further heterogeneity in the data through combining different frequencies.

In fact, in his seminal paper, Corsi (2009) sets up the HAR-RV model in which the realized volatility depends on three volatility components: short-term/daily volatility (RV_d), medium-term/weekly volatility (RV_w) and the long-term/monthly volatility (RV_M). However, while Corsi (2009) used 5 days to represent a trading week and 22 days for the month, cryptocurrency markets are open 24/24 and 7 days/week; these components have to be adapted for this specific market using 7 days to represent the week and 30 days to represent the month, as in Sapkota (2022). Further, in line with the Mixture Distribution Hypothesis (MDH) literature, trading volume was used as a control variable. Interestingly, we introduce the daily average sentiment as a driver for daily realized volatility.

We present hereafter the different empirical specifications that will be estimated and put in competition to model the daily realized volatility ($RV_{d,t}$). Thus, we set up the following specification (Model 1):

$$\begin{aligned} RV_{d,t} = \beta_0 + \beta_1 RV_{d,t-1} + \beta_2 RV_{w,t-1} + \beta_3 RV_{M,t-1} \\ + \beta_4 VOL_{t-1} + \beta_5 S_{t-1} + \varepsilon \end{aligned} \tag{A.4}$$

where S and Vol denote sentiment and trading volume respectively.

First, we split the sentiment between positive and negative components, which has the advantage to capture further asymmetry related to the sentiment effect, yielding the following specification (Model 2).

$$\begin{aligned}RV_{d,t} = \beta_0 + \beta_1 RV_{d,t-1} + \beta_2 RV_{w,t-1} + \beta_3 RV_{M,t-1} \\ + \beta_4 VOL_{t-1} + \beta_5 S^{+}_{t-1} + \beta_6 S^{-}_{t-1} + \varepsilon\end{aligned} \tag{A.5}$$

Then, we decomposed these two sentiment components into BC and NBC users' components while distinguishing the positive from the negative sentiment (Model 3). This is particularly important to estimate their different effects on Blockchain Volatility.

$$\begin{aligned}RV_{d,t} = \beta_0 + \beta_1 RV_{d,t-1} + \beta_2 RV_{w,t-1} + \beta_3 RV_{M,t-1} \\ + \beta_4 VOL_{t-1} + \beta_5 S^{+}_{BC,t-1} + \beta_6 S^{+}_{NBC,t-1} + \beta_7 S^{-}_{BC,t-1} + \beta_8 S^{-}_{NBC,t-1} + \varepsilon\end{aligned} \tag{A.6}$$

Where β_0 is the constant; the other β coefficients capture the effects of regressors.

Interestingly, to account for further nonlinearity in the data, we test the null hypothesis of linearity against its alternative hypothesis of non-linearity. After rejecting linearity, we applied a regime switching approach to analyze the effect of our sentiment variables on realized volatility in a nonlinear framework.

The Markov Switching model introduced by Hamilton (1989) has been widely used in economics and finance. The underlying idea of this model is to specify the dynamics of the dependent variable as a switching process, allowing the asymmetrical behavior of the dynamics of the dependent variable to be captured. Hereafter, we applied this Markov Regression with the HAR model to analyze the realized volatility-sentiment relationship using a regime switching approach. Our Markov-Switching HAR-RV (Model 4) is defined as:

$$RV_{d,t} = \beta_0(S_t) + \beta_1(S_t)RV_{d,t-1} + \beta_2(S_t)RV_{w,t-1} + \beta_3(S_t)RV_{M,t-1}$$
$$+\beta_4(S_t)VOL_{t-1} \quad (A.7)$$
$$+\beta_5(S_t)S^+_{BC,t-1} + \beta_6(S_t)S^+_{NBC,t-1} + \beta_7(S_t)S^-_{BC,t-1} + \beta_8(S_t)S^-_{NBC,t-1} + \varepsilon$$

where the non-observed state variable $S_t \in \{0,1\}$ and the transition across regimes are specified according to the following probability design:

$$P(S_t = s_t | S_{t-1} = s_{t-1}) = \begin{bmatrix} p_{00} & p_{10} \\ 1 - p_{00} & 1 - p_{10} \end{bmatrix} \quad (A.8)$$

where p_{ij} is the switching probability from regime i, to regime j.

These probabilities are assumed to be non-time varying. Further, the above parameters are estimated using the Maximum Likelihood method.

Empirically, we first analyzed the relationship of the different sentiment measures with Bitcoin Volatility using our HAR-RV Markov Switching model on the whole dataset. Second, we separated our dataset, using 80% for training[4] and reserving 20% of the remaining data as a test set[5] for the forecasting evaluation. We compared the forecasts obtained for model (1) and model (2), and those for model (2) and model (3). Recall that model (1) is the HAR model augmented with volatility and average sentiment of the day. In model (2), the sentiment variable is split into two components: positive and negative sentiment. In model (3), we split the sentiment variable into positive/negative and BC/NBC components.

[4]Training set: from 31 August 2018 to 6 December 2022 (1559 days)
[5]Test set: from 7 December 2022 to 31 December 2023 (390 days)

Notably, we compared the forecasts of these different models using two loss functions: the Root of the Mean Squared Error (RMSE) and the Mean Absolute Error (MAE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{A.9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{A.10}$$

where the variable Yi denotes the relaized volatility.

The loss functions were estimated for each model and then we computed the ratio of these values for the two models under consideration. If the ratio is less than one, it suggests that our model with the split between BC/NBC users supplants the benchmark model (the model without the BC/NBC components) in terms of volatility forecasting. Otherwise, the benchmark model shows better forecasting performance. In addition, we applied the Diebold-Mariano (DM) statistical test, which tests the null hypothesis that the two forecasts have the same accuracy against its alternative hypothesis of significant difference. This test is required to double check that the difference between the estimated loss function (RMSE or MAE) for the two models under consideration is statistically significant.

Finally, to identify the driver of investor sentiment, we tested the effect of the number of positive and negative news articles on the sentiment of Reddit messages for our two groups. For each sentiment component, after double checking for multicollinearity, we regressed the sentiment on its past value to capture further momentum in the sentiment on both the other sentiment variable of the two groups and on the number of positive news articles (N_t^+) and the number of negative news articles (N_t^-). This yielded the following specifications:

$$S_{BC,t}^+ = \beta_0 + \beta_1 N_t^+ + \beta_2 N_t^- + \beta_5 S_{BC,t-1}^+ + \beta_6 S_{NBC,t}^+ + \beta_7 S_{BC,t}^- + \beta_8 S_{NBC,t}^- + \varepsilon \tag{A.11}$$

where β are the coefficients of the regression; $S_{BC,t}^+$ is the number of positive comments on Reddit from Blockchain Competent users for the day t; N_t^+ is the number of positive news articles for day t; N_t^- is the number of negative news articles for the day t; $S_{BC,t-1}^+$ is the number of positive comments on Reddit from Blockchain Competent users for the day before t; $S_{NBC,t}^+$ is the number of positive comments on Reddit from Non Blockchain Competent users for the day t, $S_{BC,t}^-$ is the number of negative comments on Reddit from Blockchain Competent users for the day t, $S_{NBC,t}^-$ is the number of negative comments on Reddit from Non Blockchain Competent users for the day t, and ε is the error term

$$S_{BC,t}^- = \beta_0 + \beta_1 N_t^+ + \beta_2 N_t^- + \beta_5 S_{BC,t}^+ + \beta_6 S_{NBC,t}^+ + \beta_7 S_{BC,t-1}^- + \beta_8 S_{NBC,t}^- + \varepsilon \tag{A.12}$$

where β are the coefficients of the regression; $S_{BC,t}^-$ is the number of negative comments on Reddit from Blockchain Competent users for the day t; N_t^+ is the number of positive news articles for the day t; N_t^- is the number of negative news articles for the day t; $S_{BC,t}^+$ is the number of positive comments on Reddit from Blockchain Competent users for the day t; $S_{NBC,t}^+$ is the number of positive comments on Reddit from Non Blockchain Competent users for the day t; $S^-_{BC,t-1}$ is the number of negative comments on Reddit from Blockchain Competent users for the day before t; $S^-_{NBC,t}$ is the number of negative comments on Reddit from Non Blockchain Competent users for the day t; and ε is the error term.

$$S^+_{NBC,t} = \beta_0 + \beta_1 N^+_t + \beta_2 N^-_t + \beta_5 S^+_{BC,t} + \beta_6 S^+_{NBC,t-1} + \beta_7 S^-_{BC,t} + \beta_8 S^-_{NBC,t} + \varepsilon \tag{A.13}$$

where β are the coefficients of the regression; $S^+_{NBC,t}$ is the number of positive comments on Reddit from Non Blockchain Competent users for the day t; N^+_t is the number of positive news articles for the day t; N^-_t is the number of negative news articles for the day t; $S^+_{BC,t}$ is the number of positive comments on Reddit from Blockchain Competent users for the day t; $S^+_{NBC,t-1}$ is the number of positive comments on Reddit from Non Blockchain Competent users for the day before t; $S^-_{BC,t}$ is the number of negative comments on Reddit from Blockchain Competent users for the day t; $S^-_{NBC,t}$ is the number of negative comments on Reddit from Non Blockchain Competent users for the day t; and ε is the error term.

$$S^-_{NBC,t} = \beta_0 + \beta_1 N^+_t + \beta_2 N^-_t + \beta_5 S^+_{BC,t} + \beta_6 S^+_{NBC,t} + \beta_7 S^-_{BC,t} + \beta_8 S^-_{NBC,t-1} + \varepsilon \tag{A.14}$$

where β are the coefficients of the regression; $S^-_{NBC,t}$ is the number of negative comments on Reddit from Non Blockchain Competent users for the day t; N^+_t is the number of positive news articles for the day t; N^-_t is the number of negative news articles for the day t; $S^+_{BC,t}$ is the number of positive comments on Reddit from Blockchain Competent users for the day t; $S^+_{NBC,t}$ is the number of positive comments on Reddit from Non Blockchain Competent users for the day t; $S^-_{BC,t}$ is the number of negative comments on Reddit from Blockchain Competent users for the day t; $S^-_{NBC,t-1}$ is the number of negative comments on Reddit from Non Blockchain Competent users for the day before t; and ε is the error term.

3 Empirical Analysis

3.1 Preliminary Analysis

First, we checked the linear correlation between our variables and reported the main results in Table 34. Accordingly, we observed that past realized volatility (RV components) is highly and positively correlated with current realized volatility, indicating further evidence of volatility persistence and long memory effects that can be reproduced by a HAR-RV framework. Second, aggregated sentiment is negatively correlated with Bitcoin realized volatility as in Oad Rajput, I. A. Soomro, and N. A. Soomro 2022. Interestingly, when allowing for asymmetrical effects by splitting sentiment into positive/negative and BC/NBC, we obtained a higher significant and positive correlation between sentiment and realized volatility of bitcoin. This result is in line with prospect theory (Kahneman and Tversky 1979), which posits that investors evaluate potential gains and losses differently, with greater sensitivity to losses than gains.

Table 34: Correlation between our variables and daily realized volatility

	$RV_{d,t}$
$RV_{d,t-1}$	0.706
$RV_{w,t-1}$	0.654
$RV_{M,t-1}$	0.503
S	-0.173
S^{+}_{t-1}	0.593
S^{-}_{t-1}	0.655
$S^{+}_{BC,t-1}$	0.589
$S^{-}_{BC,t-1}$	0.651
$S^{+}_{NBC,t-1}$	0.570
$S^{-}_{NBC,t-1}$	0.632

Note: This table shows the correlation between sentiment proxies and volatility over daily frequency.

Before going ahead and estimating our models, we checked that our variables were stationary using the Augmented Dickey–Fuller (ADF) test, and we reported the main results in Table 35. Accordingly, we showed that all of our variables are stationary except VOL, S^{+}_{BC} and S^{+}_{NBC} that are integrated of one order. Accordingly, we consider their first differences hereafter.

Table 35: Results of Unit Root Test

	p-value	ADF Statistic
RV_d	0.000	-8.494
RV_w	0.000	-5.041
RV_M	0.000	-4.729
ΔVOL	0.000	-12.705
S	0.000	-11.518
S^+	0.038	-2.969
S^-	0.047	-2.883
ΔS^+_{BC}	0.000	-14.993
ΔS^+_{NBC}	0.000	-13.716
S^-_{BC}	0.019	-3.217
S^-_{NBC}	0.021	-3.187

Note: This table above shows the main results of the ADF test.

3.2 Does Blockchain Competency Sentiment Drive Bitcoin Volatility?

We investigated the influence of sentiment on realized volatility under different hypotheses and using different specifications, and reported the main results in Table 36.In fact, in model (1), reported in the second column, we considered an aggregated proxy for sentiment; we decomposed sentiment into positive/negative sentiment in model (2) reported in the third column; we considered categorized sentiment (BC/NBC) in model (3), reported in column 4 of Table 36, thereby yielding various interesting results.

First, the HAR model is well suited for modeling bitcoin volatility, explaining more than 77% of the variance. In fact, both lagged daily and weekly components show a positive and significant effect, while monthly RV appears significant only for model (1). This result is in line with Eom et al. (2019) who noted no monthly persistence in Bitcoin volatility.

Second, the aggregated measure of sentiment shows a significant and negative. In fact, when global sentiment increases of 1%, it decreases the realized volatility of 2.6% according to model (1), which is in line with Oad Rajput, I. A. Soomro, and N. A. Soomro 2022. However, this result should be read with precaution as the sentiment proxy is more of a global or aggregated measure. Further, model (1) does not account for asymmetrical effects. In model (2), we instead split sentiment into positive/negative components to allow the model to take further asymmetrical effects of positive/negative sentiment into account. Accordingly, we found that while an increase in the number of positive comments decreases volatility, bitcoin volatility rises following an increase in the number of negative comments. This suggests that investor panic and anxiety increase bitcoin volatility, while positive sentiment attenuates such volatility. Interestingly, in terms of size, the effect of negative sentiment on volatility is almost three times the effect of positive sentiment on realized volatility. This finding is particularly relevant and it indicates further evidence of asymmetry in the bitcoin volatility and sentiment relationship. The high elasticity of bitcoin volatility to negative sentiment with regard to the elasticity of bitcoin volatility to positive sentiment is an important empirical findings.

Third, for model (3), we distinguish NBC investors from the sentiment of BC investors in order to reproduce the effect of Blockchain Competency on bitcoin volatility. Our estimation results of model (3) compare to those of models (1) and

(2) while confirming that sentiment does significantly drives realized volatility. Further, we have improved the analysis of volatility-sentiment relationship while showing that for both NBC and BC investors, both the number of positive comments and the number of negative comments has a positive and significant impact on bitcoin volatility. However, another type of asymmetry appears when considering different types of investors. In fact, when considering positive sentiment, the realized volatility seems significantly four times more sensitive to sentiment of BC users than to Non Blockchain Competent (NBC) users. This suggests that when positive sentiment predominates, BC sentiment and emotions would increase the realized volatility as BC investors might act as market leaders or early adopters and the positive sentiment from this group could signal a genuine belief in Bitcoin's underlying value and potential, leading to increased buying pressure and price volatility.

As for negative sentiment, Non Blockchain Competent (NBC) users show a more marked impact on volatility. Indeed, NBC investors might be more susceptible to emotional reactions to market news and rumors. Negative sentiment from this group could trigger panic selling, also leading to an increase in bitcoin volatility.

Overall, we found that while both groups have a significant positive effect on Bitcoin volatility, their effects differ. Indeed, BC users have a greater impact on Bitcoin volatility through their positive comments, while bitcoin volatility is more significantly sensitive to negative comments of NBC investors. Accordingly, we concluded that the information extracted from positive/negative sentiment and the distinction of NB/NBC investor sentiment is relevant in explaining and reproducing the dynamics of bitcoin volatility. This is an interesting result in line with Bourghelle, Jawadi, and Rozin (2022b) and Bourghelle, Jawadi, and Rozin (2022a).

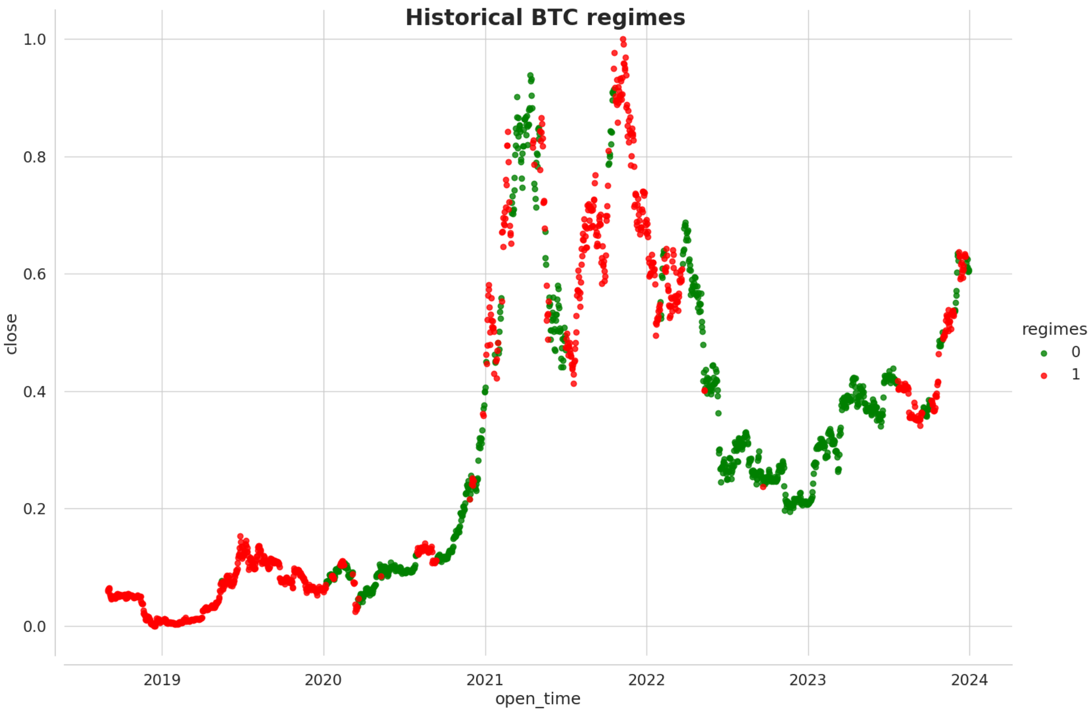
Table 36: Estimation Results of the Linear HAR-X Model

	(1)		(2)		(3)	
$Constant$	0.010***	[11.299]	0.005***	[9.788]	0.005***	[9.441]
RV_d	0.220***	[35.584]	0.186***	[31.397]	0.201***	[33.378]
RV_w	0.019***	[4.203]	0.017***	[4.007]	0.019***	[4.503]
RV_M	0.005**	[2.145]	0.001	[0.557]	0.000	[0.136]
VOL	0.160***	[34.559]	0.135***	[30.463]	0.125***	[27.973]
S	-0.026***	[-5.256]				
S^{+}			-0.057***	[-7.363]		
S^{-}			0.145***	[14.093]		
S^{+}_{BC}					0.040***	[5.464]
S^{+}_{NBC}					0.017**	[2.506]
S^{-}_{BC}					0.014*	[1.742]
S^{-}_{NBC}					0.048***	[4.635]
Number of observations	1949.000		1949.000		1949.000	
Log Likelihood	6022.638		6198.687		6233.929	
BIC	-11999.825		-12344.349		-12399.681	
AIC	-12033.275		-12383.374		-12449.857	
ADJ-R2	0.713		0.760		0.768	

Note: (***), (**) and (*) denote the statistical significance at 1%, 5%, and 10% levels respectively. Values in [.] denote the t-ratios. All explanatory variables enter with one lag.

Next, it is worth to note that while the HAR model or HARX Model is a popular model for volatility and that it has been widely applied to model volatility dynamics, the HAR model only reproduces linear effects of sentiment and it does not capture further nonlinearity or structural breaks in the volatility-sentiment relationship. This may be restrictive, especially since recent papers such as Bourghelle, Jawadi, and Rozin (2022b) showed that sentiment has a nonlinear effect on Bitcoin volatility. In order to take this further nonlinearity into account, we tested the presence of non-linearity in our data, using the structural break test of Bai and Perron (2003). Our main results, reported in the appendix, reject the hypothesis of structural stability and show the presence of different breaks, suggesting further evidence of nonlinearity in the data. In order to reproduce this nonlinearity, we estimated a Markov-switching regime HAR-X model (MS-HAR-X) with two regimes, with the main results reported in Table 37. The results corresponding to these two states (low and high volatility regimes) are plotted in 35. We refer to the low regime in green, while the high regime is indicated in red. This finding is interesting as it shows a significant alignment between the regimes and bitcoin price dynamics. In fact, the bitcoin price reached the low regime during the COVID-19 period and post-period (2020-2021) and in 2023, period during which changes in bitcoin were weak, while the high volatility regimes dominated the volatility dynamics significantly in 2022 and early 2024, period during which bitcoin has jumped being driven by different factors. The alternation between high and low regimes confirms that the MS-HAR-X aligns well with the data and seems outperforming the linear specifications. Interestingly, overall, it also appears that the high regime volatility dominates the low regime volatility, as shown by the state's probability in 37. In fact, with a probability of 95.9%, the volatility remains in the high volatility regime.

Figure 35: Dynamics of Daily Bitcoin Price and Realized Volatility Regimes



From Table 37, our findings show two different interesting results. On the one hand, we point to the relevance of distinguishing between sentiment of BC and NBC users. On the other hand, we show that the effect of sentiment on realized volatility varies with the regime under consideration suggesting the importance of taking nonlinearity and switching regime hypotheses under consideration. In fact, we found that the number of positive posts from BC investors in a low volatility regime shows a positive and significant effect, suggesting that BC users positive comments tend to increase volatility, while positive comments from NBC investors show no significant impact in this state. This highlights that when bitcoin volatility is low, only the action of BC users is expect to ensure investors, stimulate trading and drive volatility. However, negative comments from both users increase volatility in this state. However, in the state of low volatility regime, overall the action of BC dominates that of NBC users. In the high volatility regime, trading volume shows a higher effect than it does in the lower regime, which is in line with the MDH (Mixture Distribution Hypothesis). Interestingly, the number of positive comments from BC investors shows a negative and significant effect on bitcoin volatility, indicating that positive comments from BC investors tend to attenuate bitcoin volatility in this state and address a further control on changes in bitcoin prices. Negative comments from NBC users still show a positive and significant effect, while the effect of BC investors is not significant. Overall, the highest impact still arises from the negative comments of NBC investors and the positive comments of BC investors. This finding confirms the relevance of disentangling sentiment into positive/negative, BC/NBC proxies, and the importance of applying nonlinear models. Finally, when considering the linear and nonlinear volatility regimes as a whole, it appears that realized volatility is more sensitive to BC than to NBC users when considering positive comments, while the inverse is true for negative comments.

Table 37: Estimation Results of MS-HARX model

	low volatility regime coef [z]	high volatility regime coef [z]
$Constant$	0.0050 *** [9.633]	0.0040 *** [6.685]
RV_d	0.2092 *** [24.208]	0.2607 ** [46.360]
RV_w	0.0119 *** [2.647]	0.0100 *** [2.357]
RV_M	0.0016 [0.915]	-0.0013 [-0.497]
VOL	0.0762 *** [21.094]	0.4735 *** [43.450]
S_{BC}^{+}	0.0379 *** [4.490]	-0.0272 *** [-3.412]
S_{NBC}^{+}	-0.0012 [-0.155]	0.0122 * [1.776]
S_{BC}^{-}	0.0186 ** [2.370]	0.0025 [0.256]
S_{NBC}^{-}	0.0212 ** [2.003]	0.0253 ** [2.005]
p[0->0]		0.9591 *** [104.615]
p[1->0]		0.0435 *** [3.734]
Number of observations		1949
Log Likelihood		6846.824
AIC		-13651.648
BIC		-13534.571

Note: The table above shows the results of a Markov switching regime -HAR regression (model (4)). *Constant* is the constant of the regression; RV_d, RV_w, and RV_M are respectively realized volatility of the day, week, and month. VOL is the trading volume, and S_{BC}^{+} and S_{BC}^{-} are respectively the number of positive and negative comments on Reddit from Blockchain Competent Users. S_{NBC}^{+} and S_{NBC}^{-} are respectively the number of positive and negative comments on Reddit from Non Blockchain Competent Users. All explanatory variables enter with one lag.

3.3 Forecasting Bitcoin Volatility

After showing the relevance of sentiment and investor's comments to drive volatility, we propose to test whether relying on this related information might help to improve the forecasting of bitcoin volatility. Accordingly, to forecast bitcoin volatility, we evaluated the forecasting performance of the above models. We reported the main results in Table 38 and Table 39 while comparing the models to each others. Accordingly, we found that when using model (1) and (2) to produce an out-of-sample forecast of Bitcoin realized volatility, splitting sentiment into positive/negative components reduces the Mean Absolute Error (MAE) of the forecast by 9%. The Diebold Mariano test indicates a statistic of -5.111 (p-value 0.000), indicating that the difference between the two forecasts is statistically significant. This result is interesting and suggests that splitting investor sentiment into positive/negative significantly improves bitcoin volatility forecasts. The two models show similar forecasting performance when considering the RMSE loss function.

Table 38: Forecasting Results of Models (1) and (2)

Model	RMSE	MAE
model (1)	0.01	0.0068
model (2)	0.01	0.0062
Ratio	1.000	0.912

Note: The table above shows the estimated values of the two loss functions measured by the RMSE and the MAE. Ratio denotes the ratio of the RMSE and MAE of model (2) to those of the benchmark model (1). In model (1), the HAR model is augmented with volatility and average sentiment of the day. In model (2), the sentiment variable is split into two components: positive and negative sentiment.

Further, when using model (2) and (3) to produce an out-of-sample forecast of Bitcoin realized volatility, the MAE is reduced by 6% and the Diebold Mariano test indicates a statistic of -4.146 (p-value: 0.000), suggesting that the difference in accuracy between the two forecasts is statistically significant, and that splitting the sentiment into BC/NBC investors has significantly improved the bitcoin volatility forecast.

Table 39: Forecasting Results of Models (2) and (3)

Model	RMSE	MAE
model (2)	0.01	0.0062
model (3)	0.01	0.0058
Ratio	1.000	0.935

Note: The table above shows the estimated values of the two loss functions measured by the RMSE and the MAE. Ratio denotes the ratio of the RMSE and MAE of model (3) with regard to those of the benchmark model (2). In Model (2), the sentiment variable is split into two components: positive and negative sentiment. In Model (3), we split the sentiment variable into positive/negative and BC/NBC components.

Overall, these results show the superiority of model (3) and therefore the relevance of information provided by sentiment and investor posts. Further, we provide significant evidence of sentiment split into multiple components when modeling and forecasting Bitcoin volatility. In particular, taking the different types of investors into account by splitting sentiment between BC/NBC users improves significantly the volatility model forecast, which is useful for investors, managers, and market regulators.

3.4 What Drives Sentiment on Reddit?

Our main results reveal the importance of information derived from sentiment to improve the modeling and forecasting of bitcoin volatility. In particular, we highlighted the importance of positive comments from BC users and negative comments from NBC users. To take this further and better characterize the key driver of our sentiment variable, we analyzed the impact of the other sentiment proxy and news article sentiment on the investor sentiment. Interestingly, this analysis of sentiment dynamics is conducted for BC and NBC users. We report the main results in Table 40 and note several interesting results.

First, we found that news articles influence only the number of negative comments posted by Non-Blockchain Competent (NBC) users. Specifically, an increase in negative news articles correlates with a rise in negative comments from NBC users, while an increase in positive news articles corresponds to a decrease in negative comments from this group. BC users appear to be largely unaffected by news sentiment, maintaining a relatively consistent level of positive and negative comments, regardless of news coverage. This can be explained by the fact that NBC users may have a less nuanced understanding of complex systems such as blockchain and cryptocurrencies. Thus, negative news articles can trigger stronger emotional responses, leading to more negative comments from this type of investors. On the other hand, BC users have a deeper understanding of the technology, which can foster a more ”optimistic” outlook, making them less sensitive to negative news articles. This resilience is reflected in the lack of significant correlation between news sentiment and comment negativity. The resilience of BC users to negative news articles could also be explained by confirmation bias, the tendency to seek out information that confirms pre-existing beliefs while ignoring or discounting contradictory evidence. This is in line with work related to cognitive bias (Tversky and Kahneman 1974).

This finding is particularly relevant as it shows that BC and NBC users do not react to news articles in the same way and that news may have a different impact on overall sentiment, especially with less informed investors. It also shows the importance of Blockchain Competency to offset the negative effects of news articles. Cryptocurrency actors might consider implementing measures to mitigate the influence of negative news on user sentiment, such as providing more educational resources.

Second, we found that positive sentiment tends to be short-lived, with the number of positive comments declining after a day of increase yielding weak evidence of long memory in the sentiment dynamics. Conversely, negative sentiment appears more persistent, as the number of negative comments increases following a day of growth. This suggests that negative news among Reddit participants leads to lasting discussion and greater negativity among users, while positive news is more quickly acknowledged and shortly forgotten by users. This finding is also in line with Tversky and Kahneman (1974).

Table 40: Estimation Results of a Sentiment Dynamics

	S^{+}_{NBC}		S^{-}_{NBC}	
Constant	0.003*	[1.735]	-0.017***	[-15.993]
N^{+}	-0.007	[-0.330]	-0.069***	[-5.325]
N^{-}	0.010	[0.850]	0.042***	[5.731]
S^{+}_{NBC}			0.234***	[16.538]
S^{-}_{NBC}	0.362***	[11.341]		
S^{+}_{BC}	0.726***	[48.275]	0.069***	[4.508]
S^{-}_{BC}	-0.216***	[-9.010]	0.484***	[40.987]
$S^{+}_{NBC,t-1}$	-0.154***	[-11.010]		
$S^{-}_{NBC,t-1}$			0.323***	[20.609]
Number of observations	1948		1948	
Log Likelihood	3981.587		4876.664	
BIC	-7910.152		-9700.306	
AIC	-7949.174		-9739.328	
ADJ-R2	0.655		0.897	
	S^{+}_{BC}		S^{-}_{BC}	
Constant	-0.013***	[-7.274]	0.018***	[16.172]
N^{+}	0.039*	[1.930]	0.021	[1.544]
N^{-}	0.009	[0.824]	-0.013*	[-1.762]
S^{+}_{NBC}	0.711***	[50.411]	0.009	[0.599]
S^{-}_{NBC}	-0.073**	[-2.313]	0.595***	[35.797]
S^{+}_{BC}			0.391***	[24.659]
S^{-}_{BC}	0.119***	[5.049]		
$S^{+}_{BC,t-1}$	-0.192***	[-13.967]		
$S^{-}_{BC,t-1}$			0.547***	[44.027]
Number of observations	1948		1948	
Log Likelihood	4040.461		4799.446	
BIC	-8027.900		-9545.870	
AIC	-8066.922		-9584.892	
ADJ-R2	0.657		0.935	

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the *Constant*ant parameter. N refers to the number of news articles while distinguishing positive and

4 Conclusion

This study contributes to the related literature on behavioral finance while investigating the effect of investor sentiment on bitcoin volatility. Accordingly, at least three contributions can be identified. First, using recent developments in Artificial Intelligence, we propose a new measure of investor sentiment while investigating Reddit traders' comments using a deep learning algorithm (FinBERT). Notably, we created a new dictionary to differentiate Blockchain-Competent (BC) and Non-Blockchain-Competent (NBC) users on social media. Second, we propose a nonlinear Markov-Switching HAR-X model to reproduce the dynamics of Reddit sentiment on bitcoin volatility. Our MS-HAR-X model is augmented with positive/negative sentiment and BC/NBC investor sentiment. Our estimation results confirm that this model appropriately reproduces the dynamics of bitcoin volatility. In addition, we identify both low and high volatility regimes, suggesting further evidence of asymmetry and nonlinearity in the sentiment-bitcoin volatility relationship. Interestingly, we show the interest of disentangling positive/negative sentiment and the comments of BC/NBC investors to explain the dynamics of these volatility regimes. In fact, positive comments from BC users increase bitcoin volatility during periods of low volatility and reduce it during periods of high volatility. In a low volatility regime, the number of negative comments from both BC and NBC users increases volatility, while in high volatility periods, only negative comments from NBC users increase volatility. Third, we show that consideration of these different sentiment proxies per investor and per regime help to improve the out-of-sample forecast of bitcoin volatility. Finally, we show that investor sentiment (BC versus NBCs) is not driven in the same way by news articles, highlighting the relevance of Blockchain Competency. These results confirm the importance of sentiment and cognitive biases to improve the modeling and forecasting of volatility and are in line with the related literature on behavioral finance (Tversky and Kahneman 1974). They also show the importance of Artificial Intelligence tools to give reliable measures of investor sentiment. The findings have different policy implications for investors, hedgers, and market regulators.

Bibliography

Araci, D. (2019). “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models”. *ArXiv* abs/1908.10063.

Audrino, F., F. Sigrist, and D. Ballinari (2020). “The impact of sentiment and attention measures on stock market volatility”. *International Journal of Forecasting* 36.2, pp. 334–357.

Auer, R. and S. Claessens (2018). “Regulating cryptocurrencies: assessing market reactions”, p. 15.

Bai, J. and P. Perron (2003). “Computation and analysis of multiple structural change models”. *Journal of Applied Econometrics* 18.1, pp. 1–22.

Baker, M. and J. Wurgler (2006). “Investor Sentiment and the Cross-section of Stock Returns”. *The Journal of Finance* 61.4, pp. 1645–1680.

Bergsli, L. Ø., A. F. Lind, P. Molnár, and M. Polasik (2022). “Forecasting volatility of Bitcoin”. *Research in International Business and Finance* 59, p. 101540.

Bollen, J., H. Mao, and X.-J. Zeng (2011). “Twitter mood predicts the stock market”. *Journal of Computational Science* 2.1, pp. 1–8.

Bourghelle, D., F. Jawadi, and P. Rozin (2022a). “Can Collective Emotions Improve Bitcoin Volatility Forecasts?” *Bankers, Markets & Investors* 171.3, pp. 10–19.

Bourghelle, D., F. Jawadi, and P. Rozin (2022b). “Do collective emotions drive bitcoin volatility? A triple regime-switching vector approach”. *Journal of Economic Behavior & Organization* 196, pp. 294–306.

Bowden, J. and R. Gemayel (2022). “Sentiment and trading decisions in an ambiguous environment: A study on cryptocurrency traders”. *Journal of International Financial Markets, Institutions and Money* 80, p. 101622.

Chowdhury, M. S. R., D. S. Damianov, and A. H. Elsayed (2022). "Bubbles and crashes in cryptocurrencies: Interdependence, contagion, or asset rotation?" *Finance Research Letters* 46.PB.

Corsi, F. (2009). "A Simple Approximate Long-Memory Model of Realized Volatility". *The Journal of Financial Econometrics* 7.2, pp. 174–196.

Daniel, K., D. Hirshleifer, and S. H. Teoh (2002). "Investor psychology in capital markets: evidence and policy implications". *Journal of Monetary Economics* 49.1, pp. 139–209.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". *Association for Computational Linguistics.* Vol. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186.

Dyhrberg, A. H. (2016). "Bitcoin, gold and the dollar – A GARCH volatility analysis". *Finance Research Letters* 16, pp. 85–92.

Eom, C., T. Kaizoji, S. H. Kang, and L. Pichl (2019). "Bitcoin and investor sentiment: Statistical characteristics and predictability". *Physica A: Statistical Mechanics and its Applications* 514, pp. 511–521.

Hamilton, J. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". *Econometrica* 57.2, pp. 357–384.

Huang, A. H., H. Wang, and Y. Yang (2023). "FinBERT : A Large Language Model for Extracting Information from Financial Text". *Contemporary Accting Res* 40.2, pp. 806–841.

Hutto, C. and E. Gilbert (2015). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.*

Jawadi, F., W. Louhichi, H. Ben Ameur, and Z. Ftiti (2019). “Do Jumps and Co-jumps Improve Volatility Forecasting of Oil and Currency Markets?” *The Energy Journal* 0.Special I.

Kahneman, D. and A. Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk”. *Econometrica* 47.2, pp. 263–291.

Katsiampa, P. (2017). “Volatility estimation for Bitcoin: A comparison of GARCH models”. *Economics Letters* 158, pp. 3–6.

Köchling, G., P. Schmidtke, and P. N. Posch (2020). “Volatility forecasting accuracy for Bitcoin”. *Economics Letters* 191, p. 108836.

Kristoufek, L. (2023). “Will Bitcoin ever become less volatile?” *Finance Research Letters* 51.C.

Kumar, A. (2009). “Hard-to-Value Stocks, Behavioral Biases, and Informed Trading”. *The Journal of Financial and Quantitative Analysis* 44.6, pp. 1375–1401.

Lee, A. D., M. Li, and H. Zheng (2020). “Bitcoin: Speculative asset or innovative technology?” *Journal of International Financial Markets, Institutions and Money* 67, p. 101209.

Long, D., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). “Noise Trader Risk in Financial Markets”. *Journal of Political Economy* 98 (No. 4), pp. 703–738.

Long, S. (, B. Lucey, Y. Xie, and L. Yarovaya (2023). ““I just like the stock”: The role of Reddit sentiment in the GameStop share rally”. *The Financial Review* 58.1, pp. 19–37.

Loughran, T. and B. Mcdonald (2011). “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks”. *The Journal of Finance* 66.1, pp. 35–65.

Mao, H., S. Counts, and J. Bollen (2011). *Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data.*

Oad Rajput, S. K., I. A. Soomro, and N. A. Soomro (2022). "Bitcoin Sentiment Index, Bitcoin Performance and US Dollar Exchange Rate". *Journal of Behavioral Finance* 23.2, pp. 150–165.

Ofek, E. and M. Richardson (2003). "DotCom Mania: The Rise and Fall of Internet Stock Prices". *The Journal of Finance* 58.3, pp. 1113–1137.

Ortu, M., S. Vacca, G. Destefanis, and C. Conversano (2022). "Cryptocurrency ecosystems and social media environments: An empirical analysis through Hawkes' models and natural language processing". *Machine Learning with Applications* 7, p. 100229.

Öztürk, S. S. and M. E. Bilgiç (2022). "Twitter & bitcoin: are the most influential accounts really influential?" *Applied Economics Letters* 29.11, pp. 1001–1004.

Pástor, Ľ. and P. Veronesi (2009). "Technological Revolutions and Stock Prices". *American Economic Review* 99.4, pp. 1451–1483.

Perkins, A. B. and M. C. Perkins (1999). *The Internet bubble: inside the overvalued world of high-tech stocks - and what you need to know to avoid the coming shakeout.* 1st ed. New York: HarperBusiness. 283 pp.

Phillips, R. C. and D. Gorse (2017). "Predicting cryptocurrency price bubbles using social media data and epidemic modelling". *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7.

Polasik, M., A. I. Piotrowska, T. P. Wisniewski, R. Kotkowski, and G. Lightfoot (2015). "Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry". *International Journal of Electronic Commerce* 20.1, pp. 9–49.

Sapkota, N. (2022). “News-based sentiment and bitcoin volatility”. *International Review of Financial Analysis* 82, p. 102183.

Shiller, R. J. (2000). *Irrational exuberance.* Princeton, N.J. : Princeton University Press, [2000] ©2000.

Shiller, R. J. (2005). *Irrational exuberance.* 2nd ed. Currency/Doubleday. 304 pp.

Tetlock, P. C. (2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. *The Journal of Finance* 62.3, pp. 1139–1168.

Trimborn, S., M. Li, and W. K. Härdle (2019). “Investing with Cryptocurrencies—a Liquidity Constrained Investment Approach*”. *Journal of Financial Econometrics* 18.2, pp. 280–306.

Tversky, A. and D. Kahneman (1974). “Judgment under Uncertainty: Heuristics and Biases”. *Science* 185.4157, pp. 1124–1131.

Yermack, D. (2013). *Is Bitcoin a Real Currency? An economic appraisal.* NBER Working Papers 19747. National Bureau of Economic Research, Inc.

Yi, Y., M. He, and Y. Zhang (2022). “Out-of-sample prediction of Bitcoin realized volatility: Do other cryptocurrencies help?” *The North American Journal of Economics and Finance* 62, p. 101731.

Yu, M. (2019). “Forecasting Bitcoin volatility: The role of leverage effect and uncertainty”. *Physica A: Statistical Mechanics and its Applications* 533, p. 120707.

5 Appendix

Appendix 1 : Dictionary to detect BC investors

gpu, block, pool, pools, byzantine, cli , consensus,

multisignature, smart contract, dag, double spend,
double spending, etherscan, blockchain explorer,
bitcoin explorer, ewasm, fork, gui, hexadecimal, hyperledger,
mainnet, merkle, mining pool, node, opcode, private key,
public key, ring signature, sha, solidity, trustless,
turing, vitual machine, utxo, full node, segwit, lightning,
algo, algorithm, testnet, interoperability, specification,
P2P, fee, fees, supply, hardware, kyc, layer, addresses,
chain, consumption, bit, multisig, mempool, server

Appendix 2: Results of Bai-Perron Structural Break Tests

supF test statistic: 470.21

supF test p-value: 0.000

Table 41: Number of breakpoints and RSS for Bai-Perron Structural Break Tests

Number of breakpoints	Sum of squared residuals (RSS)	Observation numbers
0	616955950	
1	496942266	1493
2	453781103	797, 1493
3	433835100	796, 1088, 1493
4	430721056	292, 796, 1088, 1493
5	447378990	292, 584, 876, 1200, 1493

Note: Sum of squared residuals (RSS) and Observation numbers for each number of breakpoints.

Chapter 4: Investor Sentiment and Bitcoin Bubble: A Machine Learning Behavioral Approach

Pierre Fay, Fredj Jawadi, David Bourghelle, Philippe Rozin

IAE Lille University School of Management

Abstract

This paper examines whether behavioral factors can help to identify and explain the dynamics of a bitcoin bubble. To this end, we propose a new behavioral approach to test for bitcoin bubbles that we compare with other traditional "bubble detection" methods using a logistic regression approach. Our behavioral approach is based on behavioral and emotional factors (fear, joy, etc.) extracted from YouTube videos. We also propose a new indicator for detecting explosive price patterns. Our findings point to two interesting results. First, behavioral factors, such as fear and joy, as well as appetite for YouTube information, can be useful in identifying bull run and non-bull run bitcoin price regimes. In fact, the results of the Logistic Regression show that our behavioral approach outperforms traditional models in detecting bull runs. Second, when applying our model to a trading strategy, it also beats alternative traditional models.

Keywords: Behavioral factors, Bitcoin bubble, trading strategy, Logistic Regression. JEL: C20, F10, G10.

1 Introduction

In the wake of the digitization of the financial sector, crypto-assets have become increasingly popular, with the Bitcoin being its most iconic example. This decentralized digital currency, created in 2008 and operating on the secure and transparent blockchain technology, was initially viewed as a symbol of an alternative to traditional financial systems. Bitcoin offers a form of digital distributed ledger that ensures secure transactions in a decentralized manner. In fact, its excessive potential profit and considerable volatility has attracted a growing number of investors, transforming the bitcoin into a highly speculative asset. Since its creation, the bitcoin has shown repeated periods of up and down and dramatic swings, sometimes causing bubbles whose dynamics have been compared in the financial literature (Náñez Alonso et al. 2024) with historic bubbles such as the Tulipomania in the Netherlands (1634 to 1637)[1] or the dotcom mania (1995-2000)[2]. These recurring patterns of rapid price appreciation followed by dramatic crashes across vastly different asset classes, from 17th-century tulip bulbs to modern cryptocurrencies, suggest that bubble regimes share common characteristics that transcend time and specific market contexts.

Characteristics, such as excess volatility, investor optimism, and the significant capital flows involved, appear to be key factors in identifying and predicting bull runs or periods of intense speculation in the bitcoin market. Most importantly,

[1]For more details, see:
https://www.bloomberg.com/news/articles/2021-10-21/taleb-calls-bitcoin-a-tulip-bubble-without-the-aesthetics
https://www.theguardian.com/technology/2013/dec/04/bitcoin-bubble-tulip-dutch-banker

[2]https://www.nytimes.com/2019/04/23/technology/bitcoin-tulip-mania-internet.html

they offer invaluable insights into market efficiency, helping investors to maximize gains and protect their portfolios against market corrections and regulators to safeguard financial stability and prevent devastating collapses.

Such intense periods of rapid price appreciation, often called bull runs, are frequently analyzed in the financial literature through the lens of speculative bubbles. In financial markets, a bubble is defined as a period characterized by excessive spread between an asset's market price and its fundamental value (Sornette 2003; Jawadi and Prat 2012). For Bitcoin, however, this characterization is less obvious as it is difficult to identify a formal fundamental value for said cryptocurrency. Accordingly, implicit bitcoin value depends more on expectations of future gains. The gains related to bitcoin trading are time-varying, suggesting that investors are continually revising their estimation of Bitcoin's implicit fundamental value. In fact, Bitcoin is the first cryptocurrency to have experienced a meteoric rise since its creation. In 2025, Bitcoin exceeded \$100,000 for the first time, confirming its status as 'digital gold' and, despite its instability, attracting a growing number of institutional investors. Cryptocurrencies have also attracted the attention of private investors, boosting the value of the sector from \$20 billion to over \$120 billion.

It is also worth noting that the financial literature distinguishes between rational and irrational bubbles. On the one hand, rational bubbles appear when asset prices steadily rise due to investors' belief that they will be able to sell at a higher price in the future, regardless of its fundamental value (Blanchard 1979; Flood and Hodrick 1990). In this case, investors buy the asset not only for its current value but also for the potential capital gains they expect to earn in the future. On the other hand, irrational bubbles are driven by irrational exuberance and a loss of perspective of the asset's true value. These bubbles are often characterized by extreme optimism, fear of missing out, and a disregard for risk. The 2013 Nobel Prize winner in Economics, Robert Shiller, introduced the concept of "irrational exuberance" (R. J. Shiller 2005) to explain the dot bubble in 2000. R. J. Shiller (2005) argued that investors often become overly optimistic about future prospects, leading to a surge in asset prices that are not justified by fundamentals. The overvaluation creates a feedback loop, as rising prices encourage further buying, driving prices even higher. Finally, the bubble eventually bursts, resulting in a sharp decline in asset prices. With regard to bitcoin, the nature of Bitcoin bubbles (rational or irrational) is unclear. While there is evidence of herding behavior and fear of missing out (Bouri, Gupta, and Roubaud 2019; Wang et al. 2023), arguments for its role as a store of value (Baur and Dimpfl 2021) and hedge against inflation (Choi and Shin 2022) also make sense.

Econometrically, the related literature reveals a significant challenge in establishing a universal method to test a bubble. Indeed, multiple methods have been developed to identify bubble regimes, Taipalus (2012), Cheah and Fry (2015), and Chaim and Laurini (2019). Among them, the PSY methodology developed by Phillips, Shi, and Yu (2015a) and Phillips, Shi, and Yu (2015b) is a statistical method used to detect explosive behavior in time series. It has been extensively used to test for bitcoin bubbles (Kyriazis, S. Papadamou, and Corbet 2020; Yao and H.-Y. Li 2021; Y. Li et al. 2021; Cheung, Roca, and and 2015). However, the method primarily focuses on price patterns and does not take behavioral factors into account, even though several related studies have demonstrated the importance of such behavioral drivers (Monschang and Wilfling 2021, Stiglitz 1990; R. J. Shiller 2000; Barber and Odean 2001; Barsade 2002; Taffler et al. 2022). Barber and Odean (2001) showed that investors tend to overestimate their knowledge, leading to excessive trading and overconfidence, particularly during rising markets. R. J. Shiller (2000) showed that excess optimism or overconfidence can make investors believe that asset prices will continue to rise indefinitely, disregarding fundamental valuations, while Barsade (2002) showed that emotions can rapidly spread through investor communities. This "emotional contagion" can lead to widespread market euphoria. Investors tend to follow others' actions, especially during periods of uncertainty, a herding behavior that can help to push prices higher (Christie and Huang 1995). Taffler et al. (2022) noted that investors' actions are frequently driven by underlying psychological needs and fantasies, such as the unconscious pursuit of a "fantastic object" or an investment perceived as offering effortless wealth. The authors highlight the role of emotions such as excitement and anxiety as core drivers of investor behavior. Huber and Sornette (2022) found that the rise in bitcoin price is strongly linked to the capacity to incubate and generate visions, enthusiasm, and excitement. The use of behavioral factors to identify bubbles has not as yet been explored, however, possibly because these behavioral factors have not been specifically observed and remain only approximately proxied.

Relying on such behavioral factors, our study aims to detect and forecast bitcoin bull runs in real time by incorporating sentiment and emotion into a machine learning model. Forecasting bitcoin bull runs is challenging due to the multitude of influencing factors, including behavioral ones. In fact, the speculative nature of bitcoin has been well-documented (Yermack 2013; Cheah and Fry 2015; Bouoiyour and Selmi 2015), and speculation around bitcoin amplifies its price swings: speculators buy and sell substantial amounts of bitcoin following the announcement of an event, such as, for instance—to cite two examples—when China banned initial coin offerings in 2017 (Yue, S. Zhang, and Q. Zhang 2021) or when a cyber-attack or fraud occurs (Rognone, Hyde, and S. S. Zhang 2020). Our study contributes to related studies such as Cheah and Fry 2015; Bouri, Shahzad, and Roubaud 2019; Kyriazis, S. Papadamou, and Corbet 2020 who highlighted the influence of numerous factors on bitcoin price such as market sentiment (Michal Polasik, Kotkowski, and Lightfoot 2015; Bourghelle, Jawadi, and Rozin 2022; Loginova et al. 2024), investor attention (Garcia et al. 2014; Y. Liu and Tsyvinski 2020), regulatory changes (Lyócsa et al. 2020; Chokor and Alfieri 2021), and price manipulation (Griffin and Shams 2020). Kennis (2018) showed that including online discourse sentiment originating from various sources drives daily Bitcoin exchange movements, while Baek and Elbeck (2015) found that bitcoin returns are driven by buyers and sellers and are less influenced by traditional fundamental economic factors[3].

Unlike these studies, our paper is the first to our knowledge to propose a behavioral Machine Learning approach to check for real-time bitcoin bull runs. Our contribution can be particularly useful to identify signals for bitcoin bubbles, allowing investors to be more proactive in detecting potential bull runs and to avoid potential crashes.

[3]the author tested the consumer price index, industrial production, real personal consumption expenditures, S&P 500 index, the 10-year Treasury note, euro exchange rate and the national average unemployment rate

We present two interesting findings. First, we propose a new indicator (BPPI: Bubble Price Pattern Indicator) to detect explosive price patterns. Our indicator significantly improves bull run detection. In fact, the information extracted from behavioral actors such as positive and negative attention, and fear and joy emotions measured from YouTube information is useful as it helps to detect bull runs. Second, our behavioral model outperforms a buy-and-hold strategy and the PSY approach, as we show that it delivers superior returns and yields less related risk.

The remainder of this paper is organized into four sections. Section 2 presents our data and briefly presents the methodology. Section 3 discusses the main empirical results. Finally, Section 4 concludes.

2 Data and Methodology

2.1 The data analysis

Our study uses daily data and covers the period from 31 August 2017 to 1 February 2025. The sample includes 2,712 observations, with 628 trading days showing further evidence of bull runs. Bitcoin price and trading volume data were collected from Binance, the leading cryptocurrency exchange platform[4]).

In order to track changes in bitcoin prices, in line with Sornette 2003; Cheah and Fry 2015, we computed bitcoin volatility (V_t):

$$V_t = |log(P_t) - log(P_{t-1})| \tag{B.1}$$

[4]Binance exchange volume represents more than 6 times the exchange volume of Coinbase, which is the 2nd ranked exchange platform by volume (for more details, see: coinmarketcap.com

where P_t is the close price of the day t.

Momentum, which describes the tendency of prices to continue moving in a specific direction (Jegadeesh and Titman 1993) is relevant in a bull run as sustained upward price movement attracts more investors, creating a positive feedback loop that increases both demand and price. In fact, Y. Liu and Tsyvinski (2020) noted that a 7-day momentum is a significant factor in cryptocurrency returns, suggesting that past positive returns often predict future positive returns.

To take the momentum effect into account, let P_t denote the close price at time t. The momentum (M) over a period of 7 days can be proxied as:

$$M_t = P_t - P_{t-7} \tag{B.2}$$

where:

- M_t represents the momentum at time t.
- P_t is the close price at the current time period.
- P_{t-7} is the close price 7 days ago.

To capture behavioral features, we used YouTube to search for the "Bitcoin" keyword and we extracted the English titles, subtitles, and number of views. Overall, we obtained 172,694 views related to videos published during the sample under consideration. It is worth noting that we only considered videos featuring bitcoin price movements. The video subject was detected using the dictionary of Fay, Bourghelle, and Jawadi (2024).

YouTube's readily available view counts provide a unique and direct measure of investor attention. To analyze this attention, we employed FinBERT, a Large Language Model (LLM), to categorize the sentiment of each video title as positive, negative, or neutral. From these data, we computed daily positive (Y^+) and negative (Y^-) investor attention by adding the number of views of each positive (or negative) video of the day, i.e.:

$$Y_t^+ = \sum_{i=0}^{n} v_{t,i}^+ \tag{B.3}$$

and

$$Y_t^- = \sum_{i=0}^{n} v_{t,i}^- \tag{B.4}$$

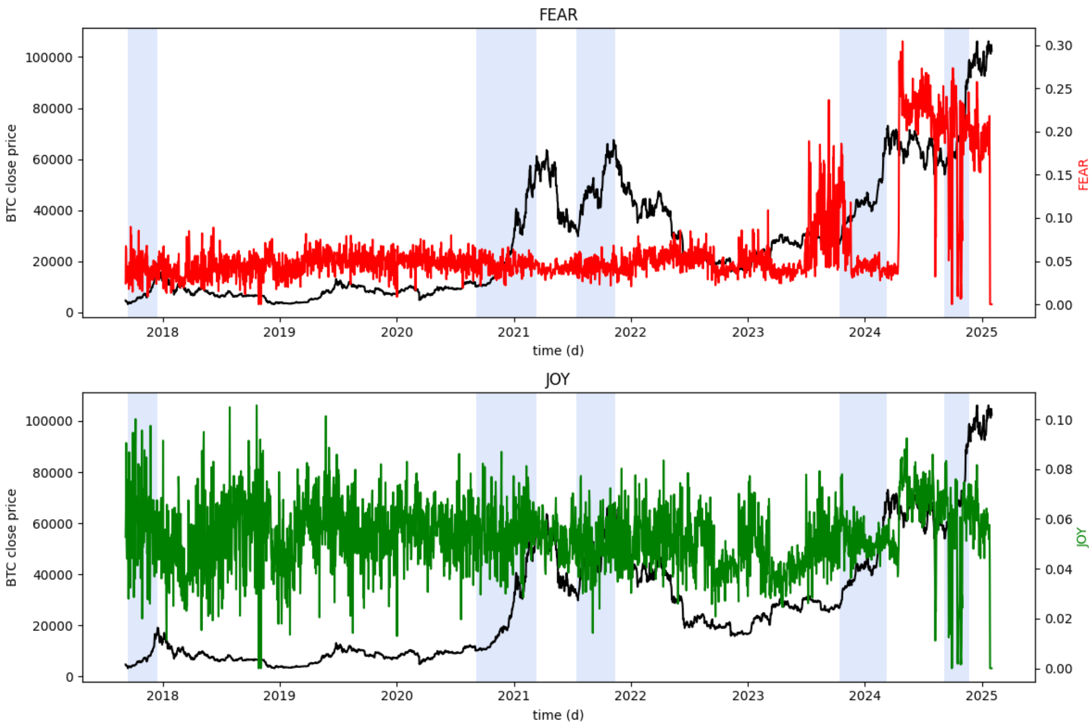
where $v_{t,i}^+$ is the number of views of the positive video i on day t. n is the total number of videos on day t and, $v_{t,i}^-$ is the number of views of the negative video i on day t.

Overall, we extracted the subtitles of our 172,694 YouTube videos related to Bitcoin prices that were published during the study period. The subtitles of each video were processed using FinEmotion proposed by McCarthy and Alaghband (2023). FinEmotion is an open-source project developed for the AI4Finance Foundation. It is a dictionary-based method, based on Text2Emotion, and improved to recognize the eight emotions of Plutchik (1980): anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. FinEmotion works in three steps. First, it pre-processes the text using standard NLP processing methods like tokenization, lemmatization, changing shortcuts [5], and removing stopwords. Second, FinEmotion uses the dictionary to check the emotion category of each word and stores

[5]changing shortcut refers to better spelling symbols and short words. For example, "u" will be translated into "You"

the count for each emotion. Third, the final score is returned by calculating the percentage representation of each emotion in the text. This score is computed for each emotion for each article. To illustrate this, we considered the mean value of this score for the fear emotion ($FEAR$) and the joy emotion (JOY), for the day, plotting the daily fear and joy scores in Figure 36, as well as the bitcoin price.

Figure 36: Dynamics of Fear and Joy Scores



Accordingly, we observed that the Fear emotion is a powerful driving force for investors during a bull run. At the beginning of the growth phase of the bubble, called “rush to possess” by Taffler et al. (2022), prices go up as investors see their peers taking advantage of the trend. This occurs when fear of missing out on a fast-earning opportunity can lead to impulsive and irrational decisions. Following this, investors still appear optimistic in a “manic denial” phase (Taffler et al. 2022), but the price is very far from its usual level and they know deep down that a price correction is coming and that they are to some extent in over their head. They want to benefit as long as possible, but fear will grow until the market correction occurs (Taffler et al. 2022). Joy is also very important during the bubble growth phase. As prices soar, investors become euphoric and may behave irrationally (R. J. Shiller 2005; Taffler et al. 2022). In fact, initial gains boost confidence and encourage further investment, fueling the bull run. As joy spreads among investors, they become overconfident and blind to warning signals (Taffler et al. 2022). They may even reject or dismiss anything that contradicts the idea that the uptrend will continue forever.

2.2 A Bubble Price Pattern Indicator

A bubble can be approximately detected visually, following a huge upward trend and a price that reaches unusually high levels. However, this approach is not concise enough and cannot help investors to protect or hedge their portfolios in a timely manner. To fill this gap, we propose a new indicator which also relies on price trends and variations. In particular, our indicator breaks down the visual perception of a bubble into two distinct, complementary, quantitative components. First, we define the Price Condition (PC) as a variable that equals 1 if the price exhibits an unusually high level, especially when the close price P_t is above the 0.8 quantile of the last 30 days; (PC) is encoded as 0 if not.

By comparing closing price (P_t) at time t to the 0.8 quantile of the past 30 days' closing prices ($Q_{0.8}(P_{t-29:t})$), we establish a threshold for the “unusually high” price. Accordingly, if the current price exceeds this threshold, it indicates that the price is significantly higher than 80% of the prices observed in the preceding 30 days. Interestingly, taking a 30-day window is long enough to smooth out further daily noise but at the same time is short enough to reflect the recent trading market activity, which seems to make sense. For more details of this type of indicator that relies on statistical measures such as quantiles, the reader can see Chandola, Banerjee, and Kumar (2009), and Kamps and Kleinberg (2018). Formally, the PC corresponds to:

$$PC_t = \begin{cases} 1, \text{If } P_t > Q_{0.8}(P_{t-29:t}) \\ 0, \text{Else} \end{cases} \tag{B.5}$$

Second, we compute the Slope Condition (SC) to determine if the price exhibits a huge upward trend. The SC takes the value 1 if the 30-day rolling slope is above its 0.8 quantile of the last 90 days; it is encoded as 0 if not. We take the slope of a 30-day window to be able to smooth out further daily noise and to capture recent market trading. Taking the 0.8 quantile as a threshold enables us to filter the moment when the trend is growing at a significantly faster rate than 80% of the time in the past 3 months. Formally, the slope corresponds to:

$$\text{Slope}(P) = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(i - \bar{i})}{\sum_{i=1}^{n}(i - \bar{i})^2} \tag{B.6}$$

where i is the time index and $(\bar{i})$ its average.

The rolling slope of the day t is defined as:

$$\text{Rolling Slope}_t = \text{Slope}(P_{t-29:t}) \tag{B.7}$$

Next, we specify the slope quantile as:

$$\text{Slope Quantile}_t = Q_{0.8}(\text{Rolling Slope}_{t-89:t}) \tag{B.8}$$

Then, the Slope Condition (SC) is specified as:

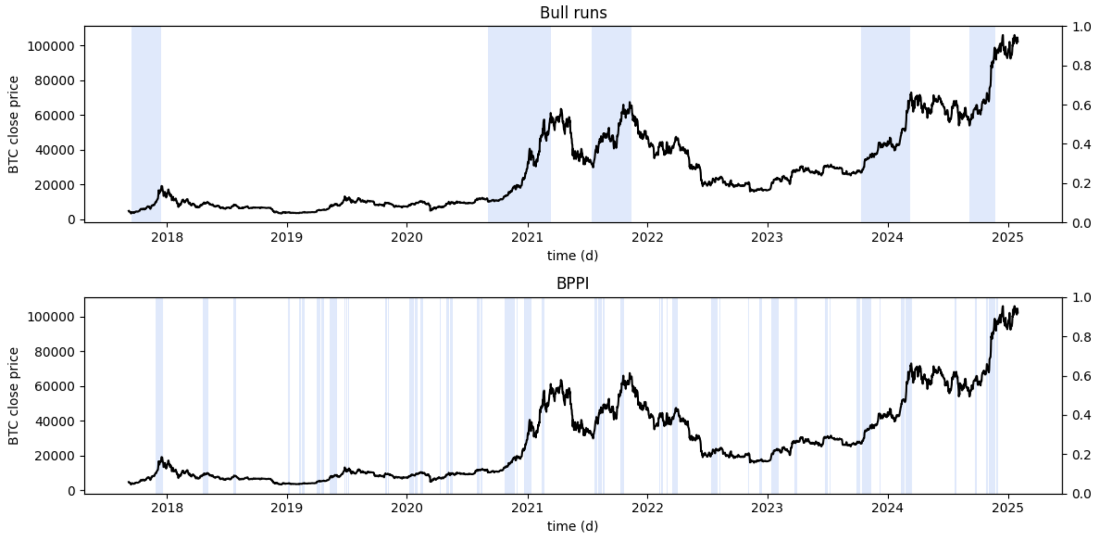
$$SC_t = \begin{cases} 1, \text{If (Rolling Slope}_t > \text{Slope Quantile}_t) \\ 0, \text{If not} \end{cases} \tag{B.9}$$

Finally, we merge our two conditions to create the Bubble Price Pattern Indicator ($BPPI$) as a dummy variable corresponding to

$$BPPI_t = \begin{cases} 1, \text{If } SC_t \wedge PC_t \\ 0, \text{If not} \end{cases} \tag{B.10}$$

We estimated and reported the Bubble Price Pattern Indicator ($BPPI$) in Figure 37. When comparing the dynamics of the BPPI with that of the bitcoin price, we observe that our BPPI indicator captures several explosive price patterns, suggesting its relevance in identifying bitcoin bull runs.

Figure 37: The Bubble Price Pattern Indicator ($BPPI$)



2.3 Historic bull run labeling

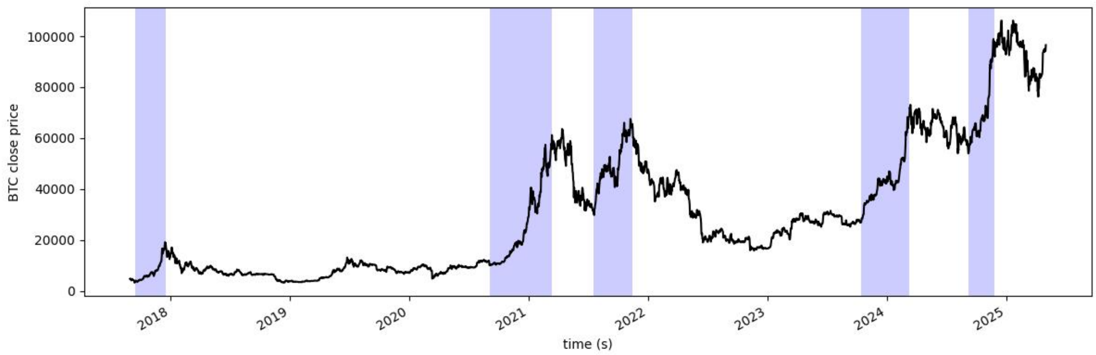
Next, we identified the main historic bull run days and computed their duration. The main results are reported in Table 42 and displayed in Figure 38, where the bull run periods are highlighted in blue.

Table 42: Historic bull runs periods

Start date	End date	Price Increase	Duration
2017-09-16	2017-12-16	414.21%	91 days
2020-09-05	2021-03-13	501.85%	189 days
2021-07-17	2021-11-13	104.25%	119 days
2023-10-14	2024-03-09	154.40%	147 days
2024-09-07	2024-11-23	80.34%	77 days

Note: This table presents a summary of the main identified historic bull run periods for Bitcoin. The periods are defined by their start and end dates, along with the total price increase percentage observed over the duration of the bull run, measured in days.

Figure 38: Bitcoin bull run periods



Overall, while five main bull run periods are identified, the two most important episodes were in September/December 2017 and September 2020/March 2021 (COVID19 period). It is worth noting that most of these periods have been signaled in the bubble-related literature and have been widely discussed as “bull runs” in the media. Interestingly, each period is characterized by specific market narratives. The 2017 bull run, widely recognized as the “Initial Coin Offering

(ICO) Mania"[6], saw Bitcoin's price surge from \$3,714 to \$19,102. This episode was predominantly driven by retail investor speculation in the nascent ICO market, with media coverage playing a significant role in amplifying market sentiment (Schillebeeckx, Tazhibaev, and Gartner 2024). In contrast, the bull runs of 2020 (\$10,166 to \$61,188) and 2021 (\$31,520 to \$64,380) indicate a shift toward greater institutional involvement. The strategic allocation of Bitcoin to corporate balance sheets by some entities such as MicroStrategy[7] and Square[8], alongside its growing consideration as a hedge against the macroeconomic consequences of the COVID-19 pandemic (Y. Zhang, Zhu, and Xu 2021) underpinned the 2020 price increase. As for the 2021 bull run, it was further supported by the introduction of regulated investment vehicles like the Purpose Bitcoin ETF[9] in Canada and the popularity of Non-Fungible Tokens (NFTs) (Nadini et al. 2021). The 2023 price rally[10] [11] (\$26,852 to \$68,313) was largely predicated on the expectation of regulatory approval for spot Bitcoin Exchange-Traded Funds (ETFs) in the US, a development expected to facilitate broader market access. Finally, the most recent bull run in 2024 (\$54,160 to over \$100,000) appears to be a culmination of these trends, with the successful launch of US spot ETFs triggering substantial institutional capital inflows (S. Liu and Yang 2024), compounded by the supply-constricting effect of

[6] See: https://www.ccn.com/ico-mania-strikes-promising-to-transform-capital-infrastructures/

[7] See: https://finance.yahoo.com/news/microstrategy-becomes-first-listed-company-113746283.html

[8] See: https://cnb.cx/30Lcxhb

[9] See: https://copper.co/en-li/insights/market-insights/how-canadas-bitcoin-etf-finally-opened-up-crypto-markets-to-everyone

[10] https://www.cnbc.com/2023/12/27/bitcoin-2023-rally-pumped-up-marathon-coinbase-microstrategy-gbtc.html

[11] See: https://www.nasdaq.com/articles/bitcoin-soared-120-2024-could-it-repeat-performance-2025

bitcoin halving in April 2024[12]. This uptrend could also been seen as the expectation of a further regulation decision promised by US President Donald Trump during his election campaign.

2.4 The PSY methodology

The PSY (Phillips, Shi and Yu) methodology was introduced by (Phillips, Shi, and Yu 2015a; Phillips, Shi, and Yu 2015b) and is widely recognized as a relevant econometric framework to test and date speculative bubbles in asset markets. It has thus frequently been used to identify periods of Bitcoin market bubbles (Cheah and Fry 2015; Bouri, Shahzad, and Roubaud 2019; Kyriazis, S. Papadamou, and Corbet 2020).

Basically, the methodology uses the Backward Super Augmented Dickey-Fuller (BSADF) test statistic, a variation of the Augmented Dickey-Fuller (ADF) test that is computed recursively over an expanding window of observations. When the BSADF statistic exceeds its corresponding critical value, it is considered as evidence of an explosive root, signaling the presence of a bubble and vice versa. The procedure helps to detect a bubble, especially during periods when the test statistic exceeds the critical value.

Formally, the main ADF regression corresponds to:

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + \epsilon_t \tag{B.11}$$

where

[12] https://www.theguardian.com/technology/2024/apr/19/what-is-bitcoin-halving-price

- y_t is the time series to test at time t.
- $\Delta y_t = y_t - y_{t-1}$
- The null hypothesis of a unit root corresponds to: $H_0 : \beta = 0$.

The PSY methodology generalizes this test by computing the ADF statistic recursively. For a given sample period from 1 to T, the BSADF statistic, denoted $ADF_r(m_0)$, is computed at each observation r in an expanding window, starting from a minimum window size m_0. The test is performed for all possible start dates r_0 within a given window $[0, r - m_0]$. For each $r \in [m_0, T]$, the BSADF statistic is defined as:

$$BSADF_r(m_0) = \sup_{r_0 \in [0, r-m_0]} \{ADF_{r_0,r}\} \tag{B.12}$$

where

- $ADF_{r_0,r}$ is the ADF statistic computed over the sample period $[r_0, r]$.
- the m_0 parameter represents the minimum length of the regression window.

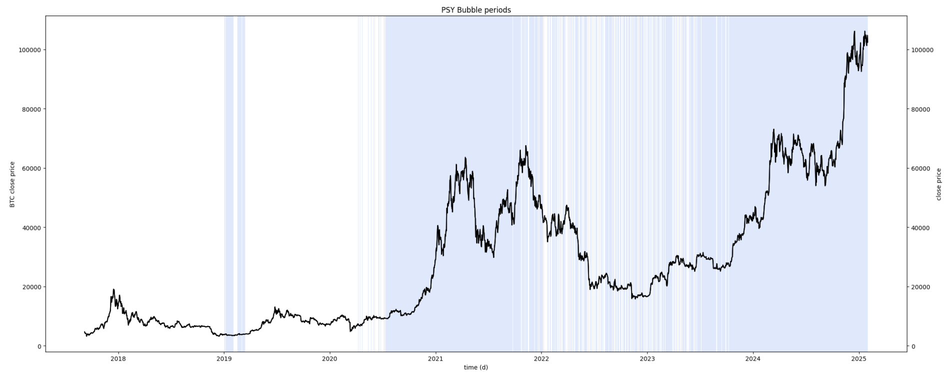
The PSY methodology compares the BSADF statistic with a sequence of critical values that are obtained through Monte Carlo simulations.

In practice, to compute these values, we used the official "psymonitor" R package[13]. For each day, a dummy variable will take the value of 1 if the BSADF is greater than the critical value, suggesting a bull run; otherwise, it takes the value 0. The main result is displayed in Figure 4, where the flagged periods are displayed in blue. It is worth noting that this procedure takes a lot of time. In fact, the related

[13] see: https://itamarcaspi.github.io/psymonitor/

computational procedure took more than four hours and 39 minutes[14].

Figure 39: Periods of bull runs with the PSY methodology



From the results in Figure 39, the PSY test flags the bitcoin as a bubble and prices as unsustainable from mid-2020. However, the PSY test is only based on prices and does not reflect other factors, which seems relatively restrictive. We thus propose extending this analysis and considering behavioral factors such as sentiment and emotions, thereby allowing us to capture additional information about investor choices and expectations. We expect this extension will provide more concise information to identify phases of a bull run.

2.5 The Machine Learning-Logistic Approach

We split the dataset into two parts: 70% of the dataset to train the model [15] and 30% for the test [16]. The bull run period is defined as a dummy variable (1 if we are

[14] with a laptop equipped of a CPU Intel Core Ultra 7 155H×22 and 64GiB of RAM
[15] Training data: from 2017-09-29 to 2022-11-10 (1897 days, with 402 flagged as bubble)
[16] Test data: from 2022-11-11 to 2025-01-31 (813 days, with 226 flagged as bubble)

in a bull run state, 0 otherwise). Accordingly, over the period under consideration, 628 of 2712 days are flagged as a bull run. Training a classification model in these conditions can lead to overfitting by the dominant class as our dataset is imbalanced. Balance is achieved by under-sampling non-bull run periods using the NearMiss algorithm (Mani and I. Zhang 2003) on the training data. In particular, this under-sampling method tackles imbalanced datasets by selectively removing majority class instances. It aims to clean up the class distribution by focusing on majority samples near the minority class. This technique helps to create a more balanced dataset for training machine learning models. The algorithm is frequently used in finance and data science when a classification is needed and the dataset is unbalanced (Mqadi, Naicker, and Adeliyi 2021).

To identify the variables that can be used to discriminate between the two regimes (bull run versus no bull run), we regress our variables using a logistic regression with Lasso regularization (Least Absolute Shrinkage and Selection Operator). This model computes the probability of a given event, and is particularly well-suited for binary classification tasks, where the outcome is binary, as in our study: state 1: a bull run versus state 0: no bull run.

Formally, the logistic regression is a classification algorithm adopted to model the probability of a binary outcome. It uses the sigmoid function to map a linear combination of input features to a probability between 0 and 1. Given the input variables $x_1, x_2, \ldots, x_n$ the probability that the dependent variable Y is equal to 1 (i.e. in a bull run) is indicated by the following sigmoid function:

$$P(Y = 1|x_1, x_2, \ldots, x_n) = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{B.13}$$

where: $\sigma(z)$ represents the predicted probability of the outcome being 1 (bull run state). z is the linear combination of the input features and their corresponding coefficients:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{B.14}$$

In order to control for further collinearity in Equation (14), we applied the Lasso regression for which L1 regularization, also known as Lasso (Tibshirani 1996), is required. This prevents overfitting by adding a penalty term to the model's cost function during training. This penalty is proportional to the absolute value of the coefficients of the above model, corresponding to:

$$\lambda \sum_{i=1}^{n} |\beta_i| \tag{B.15}$$

where:

- λ is the regularization parameter controlling the strength of the penalty.
- $|\beta_i|$ is the absolute value of the i-th coefficient.

This penalty yields the optimization algorithm to drive the coefficients of less important or irrelevant features to zero. It excludes features that have little impact on the model's predictive power. Overall, this method reduces the risk of overfitting and improves our model's specification and estimation.

3 Empirical Analysis

3.1 Preliminary Analysis

First, we checked the stationarity of our variables and reported the main results in Table 2. We found that all the variables are stationary except $FEAR$. Hereafter, we will refer to the first difference of this variable ($\Delta FEAR$).

Table 43: Results of the Unit Root Test

	p-value	ADF Statistic
$BPPI$	0.000	-10.121
$\Delta FEAR$	0.000	-12.353
JOY	0.000	-4.579
Y^+	0.000	-4.501
Y^-	0.000	-18.694
$VOLATILITY$	0.000	-6.041
$MOMENTUM$	0.000	-7.840

Note: This table shows the results of the ADF test. P-value denotes the related probability of the test.

Second, we checked for further multicollinearity in our data using the Variance Inflation Factor (VIF) test and reported the main results in Table 44. We should recall that high values of the VIF indicate a potential multicollinearity effect in our model, while low values do not. Accordingly, all the variables in the regression model show low values (always less than 5), suggesting that multicollinearity is not a concern in our data (Menard 2010).

Table 44: Variance Inflation Factor

Variable	Variance Inflation Factor
$BPPI$	1.182
$\Delta FEAR$	1.273
JOY	1.241
Y^+	1.039
Y^-	1.014
$VOLATILITY$	1.026
$MOMENTUM$	1.209

Note: This table shows the main results of the VIF test.

3.2 Bull Run Detection Test Using Machine Learning Behavioral Logistic Model

Our Machine Learning Behavioral Logistic model essentially aims to assess the probability of a bull run, while identifying the main drivers or signals of a bitcoin bubble. We estimated the above model with L1 Regularization and reported the main results in Table 45.The model is notably trained on data from 2017-09-07 to 2022-11-11, which represents the first 70% of the dataset.

Table 45: Result of a Logistic Regression with L1 Regularization

	Marginal Effect	
Constant	0.452*	[1.882]
BPPI	0.327***	[3.317]
$\Delta FEAR$	-1.516***	[-3.112]
JOY	0.265**	[2.385]
Y^+	1.936***	[5.216]
Y^-	4.362***	[4.461]
VOLATILITY	1.220***	[7.303]
MOMENTUM	0.684***	[5.655]
Number of observations		804.000
Log Likelihood		-401.910
BIC		857.338
AIC		819.821
PSEUDO R2		0.279

Note: (***), (**) and (*) denote the significance at 1%, 5%, and 10% statistical levels respectively. Values in [.] denote the t-ratios. *Constant* denotes the constant parameters. $\sigma(z)$ is the probability to be in a bull run between 0 and 1. The second column shows the marginal effects, while the third column reports their related t-ratios.

First, it is worth recalling that to analyze the results of a logistic regression, we should focus on the values of the marginal effects and their significance. We always compute the exponential value of the marginal effects to simplify their analysis. Accordingly, Table 45 shows different results with important implications for investors and regulators. First, we show that all the factors/drivers under consideration exhibit a significant effect of the probability of a bitcoin bull run. Second, while changes in fear sentiment reduce the probability of a bitcoin bubble by 79%, joy increases the probability of a bubble by 30%. Third, all the other drivers increase the probability of a bitcoin bull run in different proportions: BPPI, volatility, and the momentum effect increase the probability of a bitcoin bubble by 57%, 98%, and 238% respectively. Finally, positive and negative investor attention substantially drives and impacts the probability of a bitcoin bull run.

These results can be documented differently. First, a bull run depends on the bitcoin price acceleration, captured by the Bitcoin Price Pattern Indicator ($BPPI$). In fact, the positive and highly significant coefficient for $BPPI$ indicates that the presence of an explosive price pattern significantly increases the log-odds of being in a bull run state. The result highlights this indicator's ability to discriminate between bull and non-bull runs. Professionals may find it useful to incorporate this new indicator into their trading models.

Second, the momentum ($MOMENTUM$) has a positive and highly significant coefficient, which means that an increase in the 7-day price momentum significantly increases the log-odds of being in a bull run. This result aligns with studies showing that momentum strategies (Jegadeesh and Titman 1993) are able to generate positive payoffs for Bitcoin (Y. Liu and Tsyvinski 2020). The momentum effect can be attributed to several factors, including investor under-reaction to new information, leading to a gradual price adjustment (Hong and Stein 1999), and behavioral biases such as herding, where investors may follow the crowd, further reinforcing existing price trends (R. Shiller 1984; Bouri, Gupta, and Roubaud 2019; Vidal-Tomás, Ibáñez, and Farinós 2019; Ballis and Drakos 2020).

Third, our results confirm the important role of sentiment and emotions in shaping market dynamics during bull runs. In fact, the change in fear $\Delta FEAR$ captured by investor fear on YouTube significantly decreases the log-odds of being in a bull run. This aligns with established findings in behavioral finance, where fear is shown to trigger panic selling, heightening market volatility (Kindleberger and Aliber 2005; Kuhnen and Knutson 2011). Conversely, the positive and significant coefficient for JOY indicates that greater expressions of investor joy on YouTube increase the log-odds of being in a bull run. Consistent with the broader literature, positive sentiment and emotions like joy foster risk-taking and confidence (Kuhnen and Knutson 2011; Choudhary et al. 2024) and are correlated with rising cryptocurrency prices (K. Papadamou et al. 2023). The effect of a combination of the two sentiments on a bitcoin bubble point to the alternation of bitcoin between up and down with regard to the intensity of each sentiment.

Fourth, the significant positive coefficients for both positive and negative investor attention (respectively Y^+ et Y^-) indicate that rising attention significantly increases the log-odds of being in a bull run. These results align with the theory of Irrational Exuberance (R. J. Shiller 2005) as well as the theory of emotional finance and the work of Richard Taffler (Taffler et al. 2022).

Recognizing the predictive power of investor attention on bull run probabilities is extremely valuable for tactical asset allocation. Portfolio managers can include such investor attention measures in their analysis frameworks to identify potential periods of heightened risk or opportunity. The finding is also important for regulators as it underscores the importance of monitoring online platforms and investor sentiment. Regulators might need to develop tools and techniques to track surges in investor attention and assess their potential impact on market stability. This is particularly relevant in the context of social media and the rapid dissemination of information (and misinformation). In fact, positive videos can act as a form of social proof, reinforcing a bullish narrative and making investors feel more confident in their positions or in buying while negative videos (Fear, Uncertainty, Doubt) can be seen by speculators as an opportunity to "buy the dip". Additionally, negative videos might be viewed with skepticism or simply as noise, especially if the price continues its upward trend. Investors may exhibit confirmation bias, selectively interpreting information that reinforces their optimistic outlook (Taffler et al. 2022; Nickerson 1998).

In other words, increasing attention on both positive and negative videos also aligns with the work of Taffler et al. 2022, showing that investment decisions are driven by powerful feelings, whether conscious or unconscious, that combine both excitement (pleasure of imagined future gains) and anxiety (pain of potential losses). The fear of these gains evaporating is also described in the paper by Kahneman and Tversky 1979 who show that investors are highly sensitive to anything that threatens their portfolio.

Additionally, our results suggest that even negative attention can paradoxically fuel a bull run. In fact, for bitcoin, a controversial video or a critical review between influencers might generate attention, attracting new investors or encouraging existing ones to trade, and thereby contributing to the upward trend. This highlights the importance of having a clear legal framework to make YouTubers aware of their responsibilities and duties when they talk about crypto, especially when they are followed by a large audience.

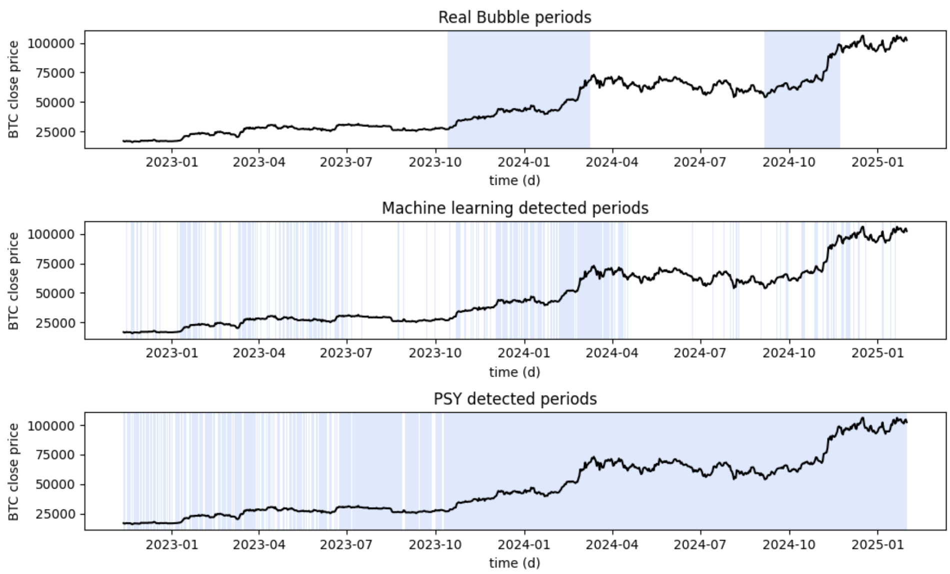
Finally, the positive and highly significant coefficient for $VOLATILITY$ suggests that higher price volatility significantly increases the log-odds of being in a bull run. The emergence of bubbles, which can be part of a bull run, is often associated with heightened volatility as speculative trading increases (Kindleberger and Aliber 2009). It is also consistent with French, Schwert, and Stambaugh (1987) who suggest a positive relationship between expected returns and risk (volatility), implying that higher expected returns in a bull market might come with higher volatility.

Overall, while our model points to the relevance of behavioral factors to better identify bitcoin bubbles, there is still room for improvement through the introduction of additional features to enhance bubble prediction, potentially including regulatory changes and information about security issues in the crypto market. This is crucial for both professionals, who can use these insights to improve risk management and trading strategies, and for academics, who can hone our understanding of bubble formation and help define more effective regulatory frameworks, checking, for instance, whether the information captured by behavioral factors can improve investors' trading strategies.

3.3 Can a Behavioral Machine Learning Model Beat Traditional Trading Strategies?

This section compares our behavioral Machine Learning model with a buy-and-hold strategy and a PSY methodology strategy. To this end, we use data from 11-12-2022 to 01-31-2025 (812 days), representing the last 30% of our dataset for this test, which covers 226 days manually labeled as bull runs. For the benchmark buy-and-hold strategy, we buy Bitcoin on the first day of the sample and we sell it on the last day of the sample. For the PSY strategy, we used the value returned by the PSY test. If the PSY test statistic for the day is above the critical value, the portfolio should contain Bitcoin for the day; if it is not above the critical value, it should not contain Bitcoin. For our machine learning-related trading strategy, we used our model trained on the first 70% of our data to provide a real-time prediction of the state of the market for each day (bull run or not). If the state predicted by our model is 1 (bull run), the portfolio should contain Bitcoin for the day; if not, it should not. The main results of the out-of-sample classification are displayed in Figure 40, where periods in blue refer to periods identified as bubbles. Accordingly, we can note that the PSY considers a large part of the periods as bubbles.

Figure 40: Classification results of the test period



Furthermore, we investigated the classifiers using 4 main indicators: True Positive (TP), the number of correctly predicted positive classes; True Negative (TN), the number of correctly predicted negative classes; False Positive (FP), the number of incorrectly predicted positive classes; and False Negative (FN), the number of incorrectly predicted negative classes.

Accordingly, we specify the accuracy as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{B.16}$$

It is worth recalling that accuracy is a key factor when detecting bull runs, as entering a market based on inaccurate bull run signals can make investors vulnerable to sharp corrections. Additionally, not being able to extract a profit before the market reverses can lead to a canceling out part of the gains made during the uptrend.

In practice, our behavioral machine learning model achieved an accuracy factor of 65.76%. With two possible outcomes (bull run or not), a purely random guess would yield an accuracy factor of around 50%. Thus, our model significantly outperforms the baseline, indicating that it learned some meaningful patterns in the data that allow it to distinguish between bull and non-bull run periods. However, with 34.24% of incorrect predictions, there remains room for improvement. Further research could focus on other variables to improve the accuracy of the model. Otherwise, the PSY methodology achieved an accuracy rating of only 37.81%, while correctly detecting most of the bull runs; it flagged the majority of the sample period, but failed to recognize when the bull run was finished.

For the sake of robustness, we took fees related to the test into account; we added 0.1% of fees[17] for each action (buying or selling). Taking these fees into consideration is important as it can significantly reduce the results of the strategy.

The main results of the strategy are reported in Table 46. In particular, we compared the three strategies across five key financial metrics: Returns, Volatility, Sharpe Ratio, Maximum Draw-down, and Value at Risk (VaR)[18]. Accordingly, we show that the returns achieved over the test period are important for all three strategies, although our strategy performed slightly better than the buy-and-hold and much better than the PSY strategy. The model offers useful guidelines and insights to enter the market in bullish times and to exit the market when it reverses. It thus improves volatility and, as a consequence, the Sharpe Ratio. Maximum draw-down and Value at Risk are also reduced for our portfolio compared to the other strategies, indicating a less risky portfolio for the strategy when our model is used.

[17] 1% is the current fees applied in Binance: https://www.binance.com/en/fee/schedule

[18] The volatility is calculated as the standard deviation of returns over the period, the Sharpe ratio is calculated as the ration of bitcoin returns on average over the volatility, the max draw-down is the maximum potential loss over the test period at a 5% confidence level.

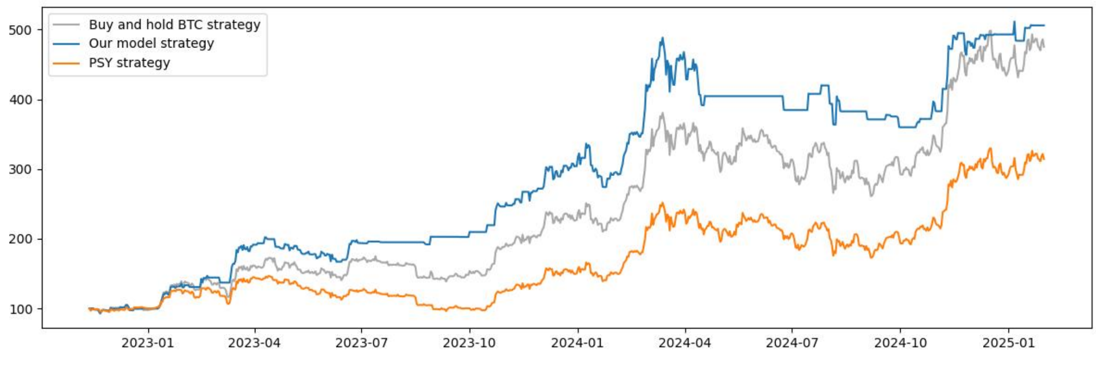
Table 46: Comparative performance of the strategies

	Buy and Hold Strategy	PSY Strategy	Our Strategy
Returns	3.75	2.14	4.06
Volatility	0.025	0.023	0.020
Sharpe ratio	2.14	1.47	2.82
Max Draw-down	-0.31	-0.34	-0.26
VaR	0.036	0.035	0.026

Note: The table presents a comparative performance analysis between a "Buy-and-Hold" investment strategy, "PSY" Strategy, and our Strategy across five key financial metrics: Returns, Volatility, Sharpe Ratio, Maximum Drawdown, and Value at Risk (VaR).

We also reported the cumulative returns of our portfolio for the three strategies in Figure 41.

Figure 41: Backtest cumulative portfolio



Overall, even if the PSY framework can be a powerful tool to detect statistical explosiveness in a time series, our results show that this criterion is not sufficient to identify a bull run. The use of sentiment and emotion data in our classifier results in a more accurate classification, underscoring their importance of their inclusion for real-time classification.

4 Conclusion

This paper proposes a new approach to detecting Bitcoin bull runs, offering valuable insights for both market participants and regulators. Our approach relies on the financial literature on behavioral finance as well as on recent developments in Machine Learning. We investigate the main behavioral characteristics of investors with regard to their sentiment and emotions: overconfidence, appetite for information, fear and joy, and attention. Using these factors, we propose a new machine learning model that collects such behavioral features from YouTube videos to detect a Bitcoin bull run regime. We also propose a new indicator for detecting explosive bitcoin price patterns. We incorporate our Machine Learning Behavioral Model into a trading strategy and compare it with traditional trading strategies to assess its practical accuracy and financial performance. Our findings offer three interesting results. First, our Bitcoin Pattern Price Indicator (BPPI) is relatively successful in discriminating bull run and non-bull run periods. The indicator offers a new quantitative tool for identifying periods of accelerated price growth, potentially providing signals of emerging bull markets. Second, we show that behavioral factors are significant factors in identifying bubbles. In fact, sentiment and emotions are key elements when discriminating between bull run and non-bull run periods. This indicates that understanding investors' emotions and sentiments by exploring information shared on social media like YouTube can provide investors with a significant edge. Third, the strategy using our Machine Learning behavioral model outperforms the traditional trading strategy, offering a better return and performance and less risk than the other strategies. Overall, these results are interesting for investors and regulators as they contribute to further insights into the mechanisms behind frequent bull runs in the Bitcoin market. Our model can help investors develop more sophisticated trading strategies, enabling them to take advantage of the Bitcoin price more effectively and to achieve better financial outcomes. For regulators, the ability to detect excessive speculative behavior, often a precursor to market instability, is crucial to protect retail investors. In fact, our Machine Learning behavioral model can help to detect periods of excessive speculative behavior in the Bitcoin market, allowing regulators to consider proactive measures. Furthermore, unlike conventional bubble detection techniques that primarily rely on price, our results highlight the importance of using sentiment and emotion when trying to detect bull runs. Future research could focus on identifying other features that could improve the performance and robustness of the model or explore its application in alternative markets that exhibit similar speculative characteristics.

Bibliography

Baek, C. and M. Elbeck (2015). "Bitcoins as an investment or speculative vehicle? A first look". *Applied Economics Letters* 22.1, pp. 30–34.

Ballis, A. and K. Drakos (2020). "Testing for herding in the cryptocurrency market". *Finance Research Letters* 33, p. 101210.

Barber, B. M. and T. Odean (2001). "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment*". *The Quarterly Journal of Economics* 116.1, pp. 261–292.

Barsade, S. G. (2002). "The Ripple Effect: Emotional Contagion and its Influence on Group Behavior". *Administrative Science Quarterly* 47.4, pp. 644–675.

Baur, D. G. and T. Dimpfl (2021). "The volatility of Bitcoin and its role as a medium of exchange and a store of value". *Empirical Economics* 61.5, pp. 2663–2683.

Blanchard, O. J. (1979). "Speculative bubbles, crashes and rational expectations". *Economics Letters* 3.4, pp. 387–389.

Bouoiyour, J. and R. Selmi (2015). "What Does Bitcoin Look Like?" *Annals of Economics and Finance* 16.2, pp. 449–492.

Bourghelle, D., F. Jawadi, and P. Rozin (2022). "Do collective emotions drive bitcoin volatility? A triple regime-switching vector approach". *Journal of Economic Behavior & Organization* 196, pp. 294–306.

Bouri, E., R. Gupta, and D. Roubaud (2019). "Herding behaviour in cryptocurrencies". *Finance Research Letters* 29, pp. 216–221.

Bouri, E., S. J. H. Shahzad, and D. Roubaud (2019). "Co-explosivity in the cryptocurrency market". *Finance Research Letters* 29, pp. 178–183.

Chaim, P. and M. P. Laurini (2019). “Is Bitcoin a bubble?” *Physica A: Statistical Mechanics and its Applications* 517, pp. 222–232.

Chandola, V., A. Banerjee, and V. Kumar (2009). “Anomaly detection: A survey”. *ACM Computing Surveys* 41.3, pp. 1–58.

Cheah, E.-T. and J. Fry (2015). “Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin”. *Economics Letters* 130, pp. 32–36.

Cheung, A. (.-K., E. Roca, and J.-J. S. and (2015). “Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices”. *Applied Economics* 47.23, pp. 2348–2358.

Choi, S. and J. Shin (2022). “Bitcoin: An inflation hedge but not a safe haven”. *Finance Research Letters* 46, p. 102379.

Chokor, A. and E. Alfieri (2021). “Long and short-term impacts of regulation in the cryptocurrency market”. *The Quarterly Review of Economics and Finance* 81, pp. 157–173.

Choudhary, S., R. Bondia, V. Srivastava, and P. Chandra Biswal (2024). “Uncovering the Bitcoin investment behavior: An emerging market study”. *Investment Management and Financial Innovations* 21.4, pp. 35–48.

Christie, W. G. and R. D. Huang (1995). “Following the Pied Piper: Do Individual Returns Herd around the Market?” *Financial Analysts Journal* 51.4, pp. 31–37.

Fay, P., D. Bourghelle, and F. Jawadi (2024). “Bitcoin returns and YouTube news: a behavioural time series analysis”. *Applied Economics* 0.0, pp. 1–19.

Flood, R. P. and R. J. Hodrick (1990). “On Testing for Speculative Bubbles”. *Journal of Economic Perspectives* 4.2, pp. 85–101.

French, K. R., G. Schwert, and R. F. Stambaugh (1987). “Expected stock returns and volatility”. *Journal of Financial Economics* 19.1, pp. 3–29.

Garcia, D., C. J. Tessone, P. Mavrodiev, and N. Perony (2014). “The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy”. *Journal of The Royal Society Interface* 11.99, p. 20140623.

Griffin, J. M. and A. Shams (2020). “Is Bitcoin Really Untethered?” *The Journal of Finance* 75.4, pp. 1913–1964.

Hong, H. and J. C. Stein (1999). “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets”. *Journal of Finance* 54.6, pp. 2143–2184.

Huber, T. A. and D. Sornette (2022). “Boom, Bust, and Bitcoin: Bitcoin-Bubbles as Innovation Accelerators”. *Journal of Economic Issues* 56.1, pp. 113–136.

Jawadi, F. and G. Prat (2012). “Arbitrage costs and nonlinear adjustment in the G7 stock markets”. *Applied Economics* 44.12, pp. 1561–1582.

Jegadeesh, N. and S. Titman (1993). “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”. *The Journal of Finance* 48.1, pp. 65–91.

Kahneman, D. and A. Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk”. *Econometrica* 47.2, p. 263.

Kamps, J. and B. Kleinberg (2018). “To the moon: defining and detecting cryptocurrency pump-and-dumps”. *Crime Science* 7.1, p. 18.

Kennis, M. A. (2018). *Multi-channel discourse as an indicator for Bitcoin price and volume movements.*

Kindleberger, C. P. and R. Z. Aliber (2005). *Manias, panics, and crashes: a history of financial crises*. 5th ed. Wiley investment classics. Hoboken, N.J: John Wiley & Sons. 309 pp.

Kindleberger, C. P. and R. Z. Aliber (2009). *Manias, panics, and crashes: a history of financial crises*. 5. ed., [repr.] Wiley investment classics. Hoboken, NJ: Wiley. 355 pp.

Kuhnen, C. M. and B. Knutson (2011). “The Influence of Affect on Beliefs, Preferences, and Financial Decisions”. *Journal of Financial and Quantitative Analysis* 46.3, pp. 605–626.

Kyriazis, N., S. Papadamou, and S. Corbet (2020). “A systematic review of the bubble dynamics of cryptocurrency prices”. *Research in International Business and Finance* 54, p. 101254.

Li, Y., Z. Wang, H. Wang, M. Wu, and L. Xie (2021). “Identifying price bubble periods in the Bitcoin market-based on GSADF model”. *Quality & Quantity* 55.5, pp. 1829–1844.

Liu, S. and C. Yang (2024). “Spot cryptocurrency ETFs: Crypto investment products or stepping stones toward tokenization”. *Finance Research Letters* 69, p. 106150.

Liu, Y. and A. Tsyvinski (2020). “Risks and Returns of Cryptocurrency”. *The Review of Financial Studies* 34.6, pp. 2689–2727.

Loginova, E., W. K. Tsang, G. Van Heijningen, L.-P. Kerkhove, and D. F. Benoit (2024). “Forecasting directional bitcoin price returns using aspect-based sentiment analysis on online text data”. *Machine Learning* 113.7, pp. 4761–4784.

Lyócsa, Š., P. Molnár, T. Plíhal, and M. Širaňová (2020). “Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin”. *Journal of Economic Dynamics and Control* 119, p. 103980.

Mani, I. and I. Zhang (2003). “kNN approach to unbalanced data distributions: a case study involving information extraction”. *Proceedings of workshop on learning from imbalanced datasets*. Vol. 126. 1. ICML United States, pp. 1–7.

McCarthy, S. and G. Alaghband (2023). “Enhancing Financial Market Analysis and Prediction with Emotion Corpora and News Co-Occurrence Network”. *Journal of Risk and Financial Management* 16.4.

Menard, S. (2010). *Logistic Regression: From Introductory to Advanced Concepts and Applications*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.

Michal Polasik Anna Iwona Piotrowska, T. P. W., R. Kotkowski, and G. Lightfoot (2015). “Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry”. *International Journal of Electronic Commerce* 20.1, pp. 9–49.

Monschang, V. and B. Wilfling (2021). “Sup-ADF-style bubble-detection methods under test”. *Empirical Economics* 61.1, pp. 145–172.

Mqadi, N. M., N. Naicker, and T. Adeliyi (2021). “Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss”. *Mathematical Problems in Engineering* 2021.1, p. 7194728.

Nadini, M., L. Alessandretti, F. Di Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli (2021). “Mapping the NFT revolution: market trends, trade networks, and visual features”. *Scientific Reports* 11.1.

Náñez Alonso, S. L., J. Jorge-Vázquez, M. Á. Echarte Fernández, and D. Sanz-Bas (2024). “Bitcoin’s bubbly behaviors: does it resemble other financial bubbles of the past?” *Humanities and Social Sciences Communications* 11.1, p. 715.

Nickerson, R. S. (1998). “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises”. *Review of General Psychology* 2.2, pp. 175–220.

Papadamou, K., J. Patel, J. Blackburn, P. Jovanovic, and E. De Cristofaro (2023). *From HODL to MOON: Understanding Community Evolution, Emotional Dynamics, and Price Interplay in the Cryptocurrency Ecosystem.*

Phillips, P. C. B., S. Shi, and J. Yu (2015a). "TESTING FOR MULTIPLE BUBBLES: HISTORICAL EPISODES OF EXUBERANCE AND COLLAPSE IN THE S&P 500". *International Economic Review* 56.4, pp. 1043–1078.

Phillips, P. C. B., S. Shi, and J. Yu (2015b). "Testing for multiple bubbles: Limit Theory for Real-Time Detectors". *International Economic Review* 56.4, pp. 1079–1134.

Plutchik, R. (1980). "A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION". *Theories of Emotion.* Elsevier, pp. 3–33.

Rognone, L., S. Hyde, and S. S. Zhang (2020). "News sentiment in the cryptocurrency market: An empirical comparison with Forex". *International Review of Financial Analysis* 69, p. 101462.

Schillebeeckx, S. J., S. Tazhibaev, and J. Gartner (2024). "FOMO and the ICO: The changing salience of quality signals". *Digital Business* 4.2, p. 100087.

Shiller, R. (1984). "Stock Prices and Social Dynamics". *Brookings Papers on Economic Activity* 15.2, pp. 457–510.

Shiller, R. J. (2000). "Measuring Bubble Expectations and Investor Confidence". *Journal of Psychology and Financial Markets* 1.1, pp. 49–60.

Shiller, R. J. (2005). *Irrational exuberance.* 2nd ed. Currency/Doubleday. 304 pp.

Sornette, D. (2003). *Critical Events in Complex Financial Systems.* Princeton: Princeton University Press.

Stiglitz, J. E. (1990). "Symposium on Bubbles". *Journal of Economic Perspectives* 4.2, pp. 13–18.

Taffler, R., X. Bellotti, V. Agarwal, and L. Li (2022). “Investor Emotions and the Psychodynamics of Asset Pricing Bubbles: A Chinese Perspective”. *Journal of Behavioral Finance* 25.3, pp. 309–333.

Taipalus, K. (2012). *Detecting asset price bubbles with time-series methods.* Scientific monographs E:47. Helsinki: Finlands Bank. 208 pp.

Tibshirani, R. (1996). “Regression Shrinkage and Selection Via the Lasso”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288.

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

Wang, Q., Q. Huang, X. Wu, J. Tan, and P. Sun (2023). “Categorical uncertainty in policy and bitcoin volatility”. *Finance Research Letters* 58, p. 104664.

Yao, C.-Z. and H.-Y. Li (2021). “A study on the bursting point of Bitcoin based on the BSADF and LPPLS methods”. *The North American Journal of Economics and Finance* 55, p. 101280.

Yermack, D. (2013). *Is Bitcoin a Real Currency? An economic appraisal.* NBER Working Papers 19747. National Bureau of Economic Research, Inc.

Yue, W., S. Zhang, and Q. Zhang (2021). “Asymmetric News Effects on Cryptocurrency Liquidity: an Event Study Perspective”. *Finance Research Letters* 41, p. 101799.

Zhang, Y., P. Zhu, and Y. Xu (2021). “Has COVID-19 Changed the Hedge Effectiveness of Bitcoin?” *Frontiers in Public Health* 9.

General conclusion

Two main assumptions in the financial literature are linked to market efficiency and investor rationality, which dominated the debate in the 60s and 70s, suggesting that the market is able to ensure an optimal allocation of capital and that investors should behave rationally and form rational expectations. Dealing with these two assumptions would naturally produce an efficient market and eliminate any mispricing. In the aftermath of the 1980s, and especially after the large stock market crash in October 1987 and the collapse of several international capital markets, both rationality and market efficiency have received serious criticism. This critical analysis opened the door to an alternative approach that incorporated a new paradigm which moved away from the representative agent hypothesis, allowing for investor heterogeneity and laying the foundations for behavioral finance. It reinstates the role of cognitive biases and investors' feelings and emotions which are at the heart of the analysis. Behavioral finance theory has been particularly successful in that it helped to improve analysis of financial asset prices and better explained the volatility excesses that characterized most financial assets in the 1990s. The main debate was about the substitution and/or complementarity between fundamental factors and behavioral factors to explain the dynamics of traditional asset pricing (stock prices, bonds, exchange rates). Recently, the debate experienced an accelerated challenge with consideration to cryptocurrencies. In fact, cryptos have shown substantial volatility excess over the last decade. Furthermore, unlike traditional assets, it is not possible to associate formal fundamental factors with cryptos, making consideration of behavioral factors to explain and forecast the dynamics of cryptos a relevant issue. However, such behavioral factors are difficult to observe, leading us to the second question/problem, namely, the identification of indicators likely to measure them (attention, sentiment, emotions, etc.). These two main questions are at the heart of this PhD, and my main task and contribution is to provide reasoned answers to them. . Along with other cryptocurrencies, Bitcoin has earned considerable interest since its launch in 2008, and following the advent of blockchain technology. As the first peer-to-peer (P2P) digital currency introduced, Bitcoin is characterized by lower transaction fees and faster settlement times compared to traditional banking systems, especially for international payments. Additionally, its decentralized nature offers enhanced security, transparency, and predictable issuance. However, its drawbacks include high volatility, making it a risky store of value; regulatory uncertainty persists globally, hampering widespread adoption; and its energy-intensive mining process raises significant environmental concerns due to its carbon footprint and electronic waste.

Unlike conventional currency, Bitcoin is not controlled by an organization but by multiple decentralized entities (individuals or enterprises) operating on the blockchain, a digital and distributed ledger. The process of producing Bitcoin and validating transactions is known as Bitcoin mining. In essence, Bitcoin was intended to provide security and transparency, making it independent of a central organization. However, it has gained exceptional global popularity as a means to escape the original monetary system, for illegal transactions, money laundering, and tax evasion, and is now mainly used for speculation.

Furthermore, the involvement of speculators considering Bitcoin as an investment vehicle has fueled the rapid growth of the cryptocurrency market. Since its inception in 2008, Bitcoin has experienced considerable price appreciation and multiple episodes of high volatility. Its price increased from around 0.10$ in 2010 to almost 110,000$ in July 2025, with multiple episodes of huge price increases followed by significant crashes, a phenomenon that has been extensively analyzed in the literature. The surge in 2017, culminating at almost $20,000, was largely attributed to retail investor interest and the ICO boom, before a significant crash in 2018, called by the press the ”crypto winter”. Another surge in 2020-2021, attributed to institutional adoption pushing prices to new all-time highs above $60,000, was followed by a substantial correction in 2022 following the Terra/Luna and FTX collapses.

Since it has no underlying cash flow, it seems challenging to analyze Bitcoin and its extreme volatility regimes like other assets. Due to the lack of consensus regarding its intrinsic value, investor behaviors are determined by expected profits. In fact, Bitcoin is primarily viewed as a speculative asset, and its appeal rests on the belief that its price will appreciate significantly in the future. This forward-looking, expectation-driven dynamic makes the market highly susceptible to shifts in sentiment and narrative rather than concrete financial performance. As a result, investor psychology is particularly important in the Bitcoin market. When valuation is subjective and driven by future expectations, the collective emotional and cognitive biases of market participants can exert an outsized influence on price movements. The financial literature highlights factors such as overconfidence bias, loss aversion, herding behavior, confirmation bias, or availability heuristics, among others, which have a significant effect on asset prices in general.

These factors, operating within a market that lacks traditional valuation anchors, create a feedback loop where sentiment (whether individual or collective, sometimes leading to mechanisms of hubris) drives price, and price movements further fuel sentiment. This inherent characteristic is a significant contributor to Bitcoin's extreme volatility regimes, making it a fascinating and challenging asset to analyze. The measurement of sentiments is considered a qualitative approach, but can also be quantified. Investor sentiments are quantified in various ways, such as using proxies like the investor volatility index (VIX) by the Chicago Board of Options Exchange (CBOE). Many researchers also construct sentiment indices according to their needs by extracting search data from Google Trends using various methods or news article titles with dictionary-based analysis methods.

Despite these advances in sentiment quantification, a persistent challenge in cryptocurrency research lies in fully capturing the complex, dynamic interplay between these quantifiable behavioral indicators and Bitcoin price patterns. The lack of a universally accepted fundamental valuation model for cryptocurrencies exacerbates this problem, as it leaves a significant void in our ability to rationally explain and predict its price movements, leading to mispricing, poor risk management, or inefficient capital allocation.

This research aims to deepen our understanding of the mechanisms governing cryptocurrency prices. Our results demonstrate that, contrary to the hypothesis of purely random Bitcoin price dynamics, behavioral factors such as investor sentiment and attention exert a significant influence on the returns and volatility of this asset and, in particular, allow phases of excessive speculation to be identified.

In the first chapter, we explored the key concepts for understanding the world of crypto-currencies. Indeed, it is rich in technical terms (blockchain, consensus, smart contracts, etc.) and is rapidly evolving. Understanding the context, concepts and terminology specific to crypto-currencies makes it easier for the reader to navigate the rest of this thesis. Our aim here is to explain how digital currencies, and bitcoin in particular, work.

In the second chapter, our research explores how investors' attention and sentiment on YouTube help account for Bitcoin returns and can improve predictions. Although YouTube is a popular platform for crypto-related content, to our knowledge, its influence on Bitcoin prices has not been investigated to date. Using artificial intelligence, we extracted sentiment and thematic content from videos, then applied Granger causality tests and a vector autoregression (VAR) model to assess lead-lag relationships between these behavioral variables and Bitcoin returns over the period 2018-2023.

This research yielded three key results. First, investor attention significantly improves bitcoin return predictions, with its impact varying depending on the video topic. Second, behavioral biases on YouTube—such as overreaction to positive news involving institutional investors or influential personalities—can affect bitcoin returns. Third, shocks to investor attention and the number of negative videos can create feedback loops that lead to sharp price fluctuations. These findings strongly suggest that taking the attention and sentiment conveyed on YouTube videos into account can be useful in improving models for forecasting bitcoin returns.

In the third chapter, our study explored how investors' blockchain competency influences Bitcoin volatility. Recognizing that while popular, cryptocurrencies' underlying technology is often misunderstood, we investigated the impact of this knowledge on Bitcoin volatility. Using a large Reddit dataset, we developed a new dictionary-based NLP method to identify technical discussions. This allowed us to categorize messages by sentiment (positive/negative) and by the sender's apparent blockchain competency. We then analyzed whether messages from blockchain-competent users affected Bitcoin volatility differently from those from less informed investors.

This chapter yielded three key results. First, differentiating sentiment according to its polarity (positive/negative) and the investor's blockchain expertise significantly improves bitcoin volatility forecasts. Second, the sentiment expressed by blockchain-savvy investors has a significant and non-linear effect on bitcoin volatility. Third, the impact of online news articles on investor sentiment varies significantly depending on the investor's blockchain expertise. This research highlights the complex factors behind bitcoin's volatility due to its technological dimension, and highlights the role of investor knowledge and online communities.

In the fourth chapter, we developed an improved model for detecting bullish phases in bitcoin by incorporating behavioral and emotional finance variables. Bitcoin is renowned for its price rises followed by crashes, and while the PSY (Phillips, Shi, and Yu 2015) method is commonly used to identify these speculative phases, we sought a more robust approach. We proposed a new method based on logistic regression to detect such speculative episodes. Our research focused on whether behavioral and emotional indicators such as fear, joy or appetite for information can help differentiate periods of excessive speculation. We also put forward an econometric model using these indicators to detect such phases.

This chapter led to three main findings. First, behavioral and emotional variables are highly effective in detecting excessive speculation phases. Second, our model outperforms the PSY method in terms of accuracy for identifying Bitcoin bubbles. Third, implementing our model in a trading strategy resulted in superior financial performance compared to a traditional "buy and hold" approach during the test period. These results show that incorporating behavioral and emotional indicators significantly enhances the ability to identify and profit from Bitcoin's speculative phases, offering a more effective alternative to traditional bubble detection methods.

It should be noted that a community medium such as YouTube or Reddit does not necessarily allow us to capture all the judgments or emotions experienced collectively. Not all sentiments and emotions expressed continuously at high speed (high frequency) are necessarily identified. Similarly, while a custom dictionary has been developed to detect blockchain skill levels, it cannot capture all the relevant technical terms in the blockchain space. This space, and the semantics of the content that develops within it, change rapidly as the technology evolves and new technical terms constantly emerge. Finally, from an empirical point of view, a logistic regression assumes a linear relationship between the independent variables (behavioral/emotional indicators) and the logarithm of the dependent variable (bull detection). However, the relationship between human emotions, market behavior and bitcoin price spikes may well be non-linear in nature, which could limit the ability of the proposed model to fully capture these dynamics. We therefore need to supplement this work with a more complex, non-linear model.

Crypto-currencies continue to raise numerous questions covering the entire economic spectrum, from monetary policy to market microstructures. The question of their ability to replace traditional monetary assets is becoming increasingly acute. The study we propose in this doctoral thesis provides an overview of the analysis of crypto-currencies, and Bitcoin in particular. It highlights market structures, volatility of digital assets, and analysis of investor behavior. We show that crypto-currency price dynamics are largely driven by behavioral variables and investor emotions. Our research also demonstrates the non-linear effects of sentiment on Bitcoin price dynamics. Specific to certain investor communities, sentiment appears in various forums and social networks, where debates are launched around issues such as investment opportunities, price volatility, or the structural instability of digital assets. It thus seems essential to delve deeper into this sentiment-price relationship in the field of crypto-currencies.

Our work opens the way to future research.

A first line of future research could focus on the influence of images conveyed on YouTube and their impact on Bitcoin prices. Chapters 2 and 4 of this thesis highlight the significant influence of YouTube on Bitcoin prices, in particular the role of sentiments and emotions extracted from the title and subtitles. A new way of measuring these factors could be to analyze non-verbal aspects, using images from the videos. Furthermore, given that a large proportion of these videos focus on technical price analysis, it is essential to study the extent to which they can contribute to the emergence of self-fulfilling prophecies. Do these predictions influence investor behavior and lead to their self-validation in the market? This study could reveal a fascinating feedback loop between content creation, market psychology, and price evolution.

A second line of research could focus on the implications of investor competence. We investigated the role of investors' blockchain knowledge on bitcoin volatility. Future research could considerably deepen our understanding of its importance by quantifying the direct financial implications of heterogeneity in blockchain competence. Further work could investigate the extent to which different levels of blockchain knowledge might translate into risk-adjusted returns, actual performance, and sensitivity to specific market pitfalls. It could also be interesting to investigate the evolution of investor skills over time. How do investors acquire knowledge about blockchain? Which channels are the most effective? Does increased competence lead to changes in investment strategies, risk perception, and greater financial performance?